Genetic landscape and novel disease mechanisms from a large LGMD cohort of 4656 patients

Babi Ramesh Reddy Nallamilli, Emory University
Samya Chakravorty, Emory University
Akanchha Kesari, Emory University
Alice Tanner, Emory University
Arunkanth Ankala, Emory University
Thomas Schneider, Emory University
Cristina da Silva, EGL Genetics-Eurofins Tucker
Randall Beadling, EGL Genetics-Eurofins Tucker
John Alexander, Emory University
Syed Askree, Emory University

Only first 10 authors above; see publication for full author list.
Genetic landscape and novel disease mechanisms from a large LGMD cohort of 4656 patients

Babi Ramesh Reddy Nallamilli¹, Samya Chakravorty¹, Akanchha Kesari¹,², Alice Tanner¹,², Arunkanth Ankala¹,² – Thomas Schneider², Cristina da Silva², Randall Beadling², John J. Alexander¹,², Syed Hussain Askree¹,², Zachary Whitt¹,³, Lora Bean¹,², Christin Collins³, Satish Khadilkar⁴,⁵, Pradnya Gaitonde⁶, Rashna Dastur⁶, Matthew Wicklund⁷, Tahseen Mozaffar⁶, Matthew Harms⁹, Laura Rufibach¹⁰, Plavi Mittal¹¹ & Madhuri Hegde¹

¹Emory University Department of Human Genetics, Atlanta, Georgia 30322
²EGL Genetics-Eurofins Tucker, Atlanta, Georgia 30084
³Augusta University, Augusta, Georgia 30912
⁴Department of Neurology, Bombay Hospital, Mumbai, Maharashtra, India
⁵Department of Neurology, Sir J J Group of Hospitals, Grant Medical College, Mumbai, Maharashtra, India
⁶Centre for Advanced Molecular Diagnostics in Neuromuscular Disorders (CAMND), 400022 Mumbai, India
⁷Neurology, The University of Colorado at Denver - Anschutz Medical Campus, Aurora, Colorado 80045
⁸Neurology, University of California, Irvine, Orange, California 92868
⁹Department of Neurology, Columbia University, New York, New York 10032
¹⁰Jain Foundation, Seattle, Wisconsin 98115
¹¹In-Depth Genomics, Bellevue, Washington

Abstract

Objective: Limb-girdle muscular dystrophies (LGMDs), one of the most heterogeneous neuromuscular disorders (NMDs), involves predominantly proximal-muscle weakness with >30 genes associated with different subtypes. The clinical-genetic overlap among subtypes and with other NMDs complicate disease-subtype identification lengthening diagnostic process, increases overall costs hindering treatment clinical-trial recruitment. Currently seven LGMD clinical trials are active but still no gene-therapy-related treatment is available. Till-date no nation-wide large-scale LGMD sequencing program was performed. Our objectives were to understand LGMD genetic basis, different subtypes’ relative prevalence across US and investigate underlying disease mechanisms. Methods: A total of 4656 patients with clinically suspected-LGMD across US were recruited to conduct next-generation sequencing (NGS)-based gene-panel testing during June-2015 to June-2017 in CLIA-CAP-certified Emory-Genetics-Laboratory. Thirty-five LGMD-subtypes-associated or LGMD-like other NMD-associated genes were investigated. Main outcomes were diagnostic yield, gene-variant spectrum, and LGMD subtypes’ prevalence in a large US LGMD-suspected population. Results: Molecular diagnosis was established in 27% (1259 cases; 95% CI, 26–29%) of the patients with major contributing genes to LGMD phenotypes being: CAPN3(17%), DYSF(16%), FKRP(9%) and ANOS1(7%). We observed an increased prevalence of genetically confirmed late-onset Pompe disease, DNAJB6-associated LGMD subtype1E and CAPN3-associated autosomal-dominant LGMDs. Interestingly, we identified a high prevalence of patients with pathogenic variants in more than one LGMD gene suggesting possible synergistic heterozygosity/digenic/multigenic contribution to disease presentation/progression that needs consideration as a part of diagnostic modality. Interpretation: Overall, this study has improved our understanding of the relative prevalence of different LGMD subtypes, their respective genetic etiology, and the changing paradigm of their inheritance modes and novel mechanisms that will allow for improved timely treatment, management, and enrolment of molecularly diagnosed individuals in clinical trials.
Introduction

Limb-girdle muscular dystrophies (LGMD) are a group of heterogeneous genetic disorders involving proximal muscle weakness with autosomal-recessive or-dominant inheritance. LGMD have more than 30 different subtypes linked to specific gene loci, which manifest in very overlapping and heterogeneous phenotypes.\(^1\)\(^-\)\(^3\) This is likely the result of a tight link between associated muscle proteins within the sarcomere-sarcoplemma-sarcoplasm-extra-cellular-matrix network (Fig. S1). The dominant forms (LGMD1A to LGMD1G) are less common than the recessive forms, representing fewer than 10\% of LGMD cases.\(^1\)\(^,\)\(^4\) Defects in a wide variety of muscle-related genes, including ones associated with severe congenital muscular dystrophy, are responsible for the 23 recessive forms (LGMD2A to LGMD2W).\(^1\)\(^,\)\(^4\)\(^-\)\(^6\) Even though each of the individual LGMD subtypes is relatively rare, they are estimated to affect 1 in 14,500 to 1 in 123,000 individuals causing them to collectively affect many people (60,000 to 500,000 individuals) worldwide and is one of the most common muscular dystrophies.\(^5\)

The substantial overlap and variability among each LGMD subtype in the age of onset, severity, and affected muscle groups, and the fact that cost of gene testing was previously unaffordable for patients, make definitive diagnosis highly elusive.\(^7\) The lack of establishing a genotype prior to targeted therapeutic approaches obstructs recruitment of LGMD patients to available clinical trials.\(^8\)\(^-\)\(^11\) Only five of the 12 LGMD trials are LGMD-subtype-specific.\(^8\) Recently, the FDA approved Phases 1b and 2 of clinical trial NCT02579239 for (Resolaris™) an orphan drug developed by aTyR Pharma for treatment of LGMD2B based on targeting the immune component of the disease subtype. No global or other subtype-specific approved therapy is available yet. In this genomic era, establishment of such therapies is greatly facilitated by a comprehensive diagnostic approach including clinical-pathological evaluation, and cost-effective parallel testing of several genes using next-generation sequencing (NGS) so that faster diagnosis can be achieved to help patients to avail such personalized therapies, management, or to be recruited for specific clinical trials.

Several studies reported the use of exomes or gene panels in small LGMD cohorts elsewhere.\(^12\)\(^-\)\(^16\) We previously showed that for NMDS such as LGMD, comprehensive gene-panel testing has a higher clinical yield (46\%) than single-gene testing (15\%) or has about 18\% higher detection rate of causative pathogenic variants than exome sequencing.\(^17\)

Previously, multiple targeted disease-specific panel NGS efforts including that of LGMD for molecular diagnostics were performed.\(^12\)\(^,\)\(^18\)\(^-\)\(^24\) But to our knowledge, this study uniquely surpasses them by recruiting a very large number (4656) of patients clinically suspected of a specific disorder (LGMDs) in the USA irrespective of ethnicities. We aimed to provide complete molecular diagnosis to this large group of clinically characterized LGMD patients. We also aimed to understand the respective prevalence, the gene-variant spectrum of LGMD subtypes across the USA, and novel disease mechanisms leading to unusual clinical presentation and progression of LGMD that can broaden understanding of disease inheritance and apply to other heterogeneous inherited diseases.

Methods

Patient enrollment

The Muscular Dystrophy Association (MDA) and Jain Foundation jointly launched large-scale LGMD molecular diagnostic testing through Emory Genetics Laboratory (EGL). The study included 4656 individuals not reported previously who underwent molecular testing from June 2015 to June 2017. The study protocols have received prior approval by the Emory University Institutional Review Board, and that informed consent was obtained from each subject for genetic testing using the targeted LGMD gene-panel NGS and for collecting and using their clinical data in this study.

The inclusion criteria for this study were clinical suspicion of LGMD based on clinical data, questionnaire answers, and prior imaging studies on muscle biopsies by respective physicians and clinics, and no prior confirmed molecular testing of disease. Since most of the 4656 patients are adults, in most cases parental sample was not available for segregation analysis, and hence this study is mostly proband related, albeit some segregation studies on some families was performed. For the Jain Foundation program patients clinically suspected of having LGMD were enrolled in the study with informed consent through an institutional review board protocol (IRB00075815) approved by Emory University with the help of Jain Foundation. The Jain Foundation developed the “Automated limb-girdle muscular dystrophy diagnostic assistant” tool, popularly referred to as ALDA (https://jain-foundation.org/alda/content/login-tool) where physicians fill up their respective patient clinical parameters which is helpful in clinical prediction of the LGMD subtype. ALDA includes prediction of all currently identified LGMD subtypes as well as other muscular dystrophies having similar clinical presentations.\(^25\) The ALDA tool results report out two main parameters, probability and concordance. Based on the provided patient clinical information by the physician, the probability percentage indicates up to the top three most likely clinical diagnoses.
for each patient. The concordance score represents how well the patient phenotype corresponds to the predicted LGMD subtypes, based on the available information in the published scientific literature. A subset of the patients with muscle phenotypes were initially screened using the online ALDA tool questionnaires and all positively screened LGMD patients by respective physicians were enrolled in the current large-scale study. Additionally, patients, only in consultation with their respective clinics and physicians, could enter the study by answering the sequencing program website (http://lgmd-diagnosis.org/) (https://www.jain-foundation.org/patient-physician-resources/free-genetic-sequencing) questions that were stored as part of clinical data (See Questionnaires in Supporting Information Methods). For the MDA program, neurologists from different MDA clinics across US also referred LGMD-suspected patients for this large-scale molecular testing program and also used the ALDA tool for providing us quantifiable clinical data.

During recruitment, ethnicity of patients was not one of the questions that we asked in our intake questionnaire. Although, in ALDA, we ask whether the patient was Finnish, Other Northern European, Japanese, other not listed, or unknown (see questionnaires in Supporting Information Methods). The reason for these specific choices is because there are specific types of LGMDs that are more common in these ethnicities so that is important to take these into account when predicting the possible type of LGMDs.

For details of the patient questionnaires for recruitment to the sequencing program, see (Supporting Information Methods).

**Targeted gene panel testing**

Clinical NGS of targeted gene-panel was performed at the Emory Genetics Laboratory (EGL), a facility certified by the Clinical Laboratory Improvement Amendments and College of American Pathologists (CLIA-CAP) as previously described.\(^1^7\) Genomic DNA was isolated from peripheral blood samples or saliva samples for each patient referred for molecular testing. The list of potential LGMD genes to screen is too large for a single gene testing approach, but is very well suited for a targeted NGS panel. Genes selected for this panel were based on their evidence for disease causality, their relative contribution to disease as described in the literature\(^2^6\) and relative cost analysis. A total of 35 genes related to specific LGMD subtypes or other muscular dystrophies with clinical overlap with LGMD features were included in this NGS panel (Table S2). Targeted NGS was performed as described by us previously\(^1^7\) with 100% coverage for each gene in the panel using Agilent SureSelect target enrichment. Patient genomic DNA was sheared to a 200 bp average, followed by end-repair, A-tailing, adapter ligation, and library amplification. Samples were then hybridized with the capture library. The probes were biotin tagged to allow for capture by streptavidin beads following hybridization. Prepared libraries were then indexed and amplified. Direct sequencing of the amplified captured regions was performed using next generation short base pair read sequencing. Although this assay was not designed to detect intragenic deletions and duplications, breakpoint bait was included in NGS library for common adult-onset Pompe deletion of exon 18 and two intronic dysferlin pathogenic variants.

**NGS data analysis and variant interpretation**

We analyzed targeted NGS variant data, and identified variants classified according to board standards and guidelines\(^2^7,2^8\) (http://www.egl-eurofins.com/emvclass/emvclass.php). Alignment to the human reference genome (hg19) and variant calling was performed using NextGENe\(^\text{®}\). For all 35 genes included in the panel (Table S2), the analyzed region includes the coding exons and 10 bp of flanking intronic region on both sides of each exon. Certain known deep intronic variants were also targeted. We analyzed targeted NGS variant data carefully considering other important parameters such as available clinical phenotype information, family history and sequence result (if available) and previous undiagnosed diagnostic genetic test results. Identified variants were classified according to standards and guidelines of the American College of Medical Genetics and Genomics and classification were made available\(^2^7,2^8\) (http://www.egl-eurofins.com/emvclass/emvclass.php). Diagnostic yield was calculated based on definitive diagnosis of patients harboring pathogenic variants in autosomal-recessive, -dominant, and X-linked genes and patients with one pathogenic and one variant of uncertain significance (VUS) in the same LGMD gene. Prevalence of each identified LGMD subtype identified by our clinical NGS panel diagnostic program was established and compared. Major genes contributing to LGMD genes were identified from the patients with definitive molecular diagnosis having pathogenic variants in autosomal-recessive, -dominant, and X-linked genes. Homozygosity was compared for different subtypes of LGMD and relatively common variants were identified in different LGMD-associated genes.

**Peripheral blood mononuclear cells (PBMC) assay for Dysferlin (DYSF) protein estimate**

Cases suspected of dysferlinopathy, as in LGMD2B or Miyoshi myopathy with DYSF (OMIM# 603009) variant(s), were analyzed for DYSF protein expression using our established blood PBMC assay.\(^2^5,2^9,3^0\)
Results

Definitive molecular diagnosis and demographics

Seventy percent of the patients enrolled in the study were over 18 years of age, and 30 percent were less than 18 years with patient mean age during participation was about 40 years equally distributed among the two genders (male: 50.6%, female: 49.4%). About half of the patients (2359/4656; 50.66%) were recruited through more than 150 MDA clinics across the US (https://www.mda.org/) and the remaining half (2297/4656; 49.33%) recruited through the Jain Foundation. Overall diagnostic yield was based on definitive diagnosis of 1003 patients harboring pathogenic variants in autosome-recessive, -dominant, and X-linked genes and 256 patients with one pathogenic and one VUS in the same recessive LGMD gene, totaling to 1259 diagnosed cases. Gene contribution percentages, that is, yield of each LGMD subtypes were based on the 1003 patients harboring pathogenic variants with definitive diagnosis. The diagnostic yield does not include the possible digenic/multigenic cases that harbor a single pathogenic variant in ≥2 genes. Critical analysis of this extensive molecular data confirmed molecular diagnosis for 27% (1259/4656 cases) of all the patients belonging to eleven different autosomal-recessive LGMD subtypes associated with DYSF(OMIM# 603009), FKRP(OMIM# 606596), GAA(OMIM# 606800), SGCA(OMIM# 600119), SGCB(OMIM# 600900), SGCG(OMIM# 608896), TRIM32 (OMIM# 602290), FKTN(OMIM# 607440), ANO5 (OMIM# 608662), POMT2(OMIM# 607439), and ISPD (OMIM# 614631). However, no cases with a definitive diagnosis were identified for the four autosomal-recessive LGMD subtypes associated with POMT1(OMIM# 607423), SGCD(OMIM# 601411), DAG1(OMIM# 128239), and POMGNT1(OMIM# 606822). Majority of the patients have pathogenic variants in one of the following genes: CAPN3 (OMIM# 114240) 17% (175/1003 cases), DYSF (OMIM# 603009) 16% (167/1003 cases), FKRP (OMIM# 606596) 9% (87/1003 cases), and ANO5 (OMIM# 608662) 7% (72/1003 cases); indicating that these genes are likely the major contributors to LGMD phenotypes in the US population (Fig. 1A and Fig. S2). Pathogenic variants were identified in dominant LGMD genes DNAJB6(OMIM# 611332), MYOT(OMIM# 604103), and CAV3(OMIM# 601253), but not in the TNPO3(OMIM# 610032) gene. Pathogenic variants also occurred in 10 genes with both autosomal-recessive and -dominant inheritance including CAPN3, COL6A1, COL6A2(OMIM# 120240), COL6A3 (OMIM# 120240), DES(OMIM# 125660), GNE(OMIM# 603824), LMNA(OMIM# 150330), TCAP OMIM# 604488), and TTN(OMIM# 188840). In 24% (1125/4656 cases) of the patients, we found no reportable variants which are true negative group in our targeted LGMD gene-panel NGS. In the rest (49%; 2272/4656 cases) patients, we identified

Figure 1. Major contributing LGMD genes. (A) Molecular diagnosis has been established in 27% of the patients. A majority of these patients had a pathogenic variant in one of the following genes CAPN3 17%(175/1003), DYSF 16%(167/1003), FKRP 9%(87/1003), and ANO5 7%(72/1003) indicating that these genes are likely the major contributors to LGMD phenotype. (B) Number of unique pathogenic variants identified. Numbers of identified pathogenic variants were compared among the major contributing LGMD genes to understand the allelic heterogeneity of these genes. DYSF, CAPN3, and COL6A1 were identified with the most pathogenic variants including 133, 95 and 40, respectively, in each gene, indicating more allelic heterogeneity in these genes.
reportable variants but did not confirm molecular diagnosis such as the cases with one pathogenic variant in a recessive genes or cases with one or two VUSs in a gene, or cases with a single pathogenic variant or VUS in multiple genes. Seventy-two percent of identified reportable variants are classified as VUS, and 23% of the variants are classified as pathogenic variants (Fig. 2). Patients had at least one VUS either in DYSF (90 patients), CAPN3 (44 patients), or in any of the three collagen genes (246 patients) (Table S7) without any reportable variant in other LGMD genes. Overall, 45% of patients were of “Northern European” origin except Finnish, 29% were “Other Not Listed”, 20% were “Unknown”, and the remaining 5% comprised of ethnic origins of Finnish, North Asia, South Asia, Hispanic, Native American, and African American. These data originated from self-reported responses by the patients on the testing form (see Supporting Information Methods).

Allelic heterogeneity of major LGMD contributing genes

We identified 133 unique pathogenic variants in the DYSF gene followed by 95 pathogenic variants in CAPN3 and 40 pathogenic variants in COL6A1 (OMIM# 120220) thus indicating a high allelic heterogeneity in these genes (Figs. 1B and 2). This large-scale study confirmed calpainopathy as the most common LGMD subtype (CAPN3-associated LGMD2A: 175/1003 cases; 17%) in the USA. Sixteen percent of patients (167/1003 cases) had a definitive diagnosis of LGMD2B by identifying two pathogenic variants in the DYSF gene indicating LGMD2B is the second major contributor among the LGMD subtypes (Fig. 1A). Splice site pathogenic variant c.2643 + 1G>A was identified in 20 LGMD2B patients making it as the most common DYSF pathogenic variant followed by c.5713C>T (p.R1905X) and c.5979dupA, as the second most common variants each found 12 patients. Extensive homozygosity of variants was observed in the DYSF (10%), FKRP (7%), and CAPN3 (6%) genes (Fig. 3).

Mode of inheritance in LGMD subtypes

In addition to being the most common LGMD subtype (CAPN3-associated LGMD2A: 17%), we report that 17 unrelated patients with a 21 bp in-frame deletion (c.643_663del21) in CAPN3 gene who did not possess a second reportable variant (Table 2). This variant occurred in multiple different families with muscle disease, segregating with dominant inheritance. A second 15 bp in-frame deletion (c.598_612del15) was also identified in different 16 unrelated patients. Most of these 16 patients had no identified second variant, indicating an increased prevalence of autosomal-dominant CAPN3-associated LGMD subtype (Table 2).

Prevalence of late-onset Pompe in large LGMD-suspected cohort

We identified 28 LGMD-suspected patients, all with two GAA pathogenic variants (Table 1). We also identified 10 patients with one pathogenic variant and one VUS in
GAA which is likely to be pathogenic given no other mutation in GAA or other genes and that the patients show clinical symptoms (Table S3), totaling to 38 identified late-onset Pompe cases (38/4656; 0.8% prevalence). It is interesting to note that none of the patients had homozygous variants in GAA (Table 1) with the c.-32-13T>G leaky splice variant identified in 31 of 38 LOPD patients in the compound-heterozygous state. The common exon-18 deletion (c.2481 + 102_2646 + 31del) was found in five patients along with the leaky-splice variant. The third common variant, c.525delT, was identified in five LOPD patients. Ninety-one patients had at least one pathogenic variant in GAA indicating further deletion, duplication, and enzymatic analysis are warranted in these individuals.

**DNAJB6-associated LGMD subtype 1E molecular diagnosis and prevalence**

DNAJB6 (OMIM# 611332) is an autosomal-dominant gene associated with limb-girdle muscular dystrophy 1E (LGMD1E). We identified 29 LGMD1E cases with pathogenic variants and 13 additional cases with novel missense variants in DNAJB6 (OMIM# 611332) gene. MYOT (OMIM# 604103) was the next major contributor to the autosomal-dominant LGMD group with pathogenic variants identified in 30 patients.

**Multigenic inheritance in muscular dystrophy**

Instead of one locus, alterations in two genes are required for phenotypic expression in digenic inheritance as in retinitis pigmentosa (ROM1 (OMIM# 180721) and PRPH2 (OMIM# 179605)), as well as in different muscular dystrophies such as Emery-Dreifuss, limb-girdle, Ullrich congenital and facioscapulohumeral, indicating more complex nature of disease presentation and progression in muscular disorders. Our data identified pathogenic variants in more than one LGMD genes in at least 31 individuals suggesting potential digenic or multiple genes contribution to LGMD disease presentation and progression (Fig. 4 and Table S1). Majority of these patients showed atypical clinical presentation and in some
Table 1. Summary of GAA variants identified in late-onset Pompe patients. Identification of 28 patients with two GAA pathogenic variants indicates the increased prevalence of late-onset Pompe disease in this study. (AOP: Adult-onset Pompe).

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Gender</th>
<th>Age</th>
<th>Gene</th>
<th>Variant 1</th>
<th>Variant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOP1</td>
<td>Female</td>
<td>61</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.1124G&gt;T (p.R375L)</td>
</tr>
<tr>
<td>AOP2</td>
<td>Female</td>
<td>79</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2140delC</td>
</tr>
<tr>
<td>AOP3</td>
<td>Female</td>
<td>33</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.525delT</td>
</tr>
<tr>
<td>AOP4</td>
<td>Male</td>
<td>71</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.1912G&gt;T (p.G638W)</td>
</tr>
<tr>
<td>AOP5</td>
<td>Unknown</td>
<td>54</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2512C&gt;T(p.Q838X)</td>
</tr>
<tr>
<td>AOP6</td>
<td>Male</td>
<td>66</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2481 + 102,2646 + 31del (Exon 18 deletion)</td>
</tr>
<tr>
<td>AOP7</td>
<td>Male</td>
<td>70</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2481 + 102,2646 + 31del (Exon 18 deletion)</td>
</tr>
<tr>
<td>AOP8</td>
<td>Female</td>
<td>44</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2481 + 102,2646 + 31del (Exon 18 deletion)</td>
</tr>
<tr>
<td>AOP9</td>
<td>Male</td>
<td>18</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2481 + 102,2646 + 31del (Exon 18 deletion)</td>
</tr>
<tr>
<td>AOP10</td>
<td>Male</td>
<td>40</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2238G&gt;Alp.W746X</td>
</tr>
<tr>
<td>AOP11</td>
<td>Male</td>
<td>59</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.1655ST&gt;C(p.L552P)</td>
</tr>
<tr>
<td>AOP12</td>
<td>Male</td>
<td>70</td>
<td>GAA</td>
<td>c.736delC</td>
<td>c.546G&gt;Alp.T183T</td>
</tr>
<tr>
<td>AOP13</td>
<td>Female</td>
<td>53</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.1841C&gt;Alp.T614K</td>
</tr>
<tr>
<td>AOP14</td>
<td>Male</td>
<td>68</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.1143delC</td>
</tr>
<tr>
<td>AOP15</td>
<td>Female</td>
<td>40</td>
<td>GAA</td>
<td>c.853C&gt;T</td>
<td>c.2560C&gt;T(p.R854X)</td>
</tr>
<tr>
<td>AOP16</td>
<td>Male</td>
<td>41</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2560C&gt;T(p.R854X)</td>
</tr>
<tr>
<td>AOP18</td>
<td>Male</td>
<td>70</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.1064T&gt;C(p.L355P)</td>
</tr>
<tr>
<td>AOP19</td>
<td>Female</td>
<td>49</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.1655T&gt;C(p.L552P)</td>
</tr>
<tr>
<td>AOP20</td>
<td>Female</td>
<td>56</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.525delT</td>
</tr>
<tr>
<td>AOP21</td>
<td>Female</td>
<td>36</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.1827delC</td>
</tr>
<tr>
<td>AOP22</td>
<td>Male</td>
<td>80</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.525delT</td>
</tr>
<tr>
<td>AOP23</td>
<td>Male</td>
<td>33</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.258dupC</td>
</tr>
<tr>
<td>AOP24</td>
<td>Female</td>
<td>46</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.766_785delTATATCACAGGCCTCGCCGAinsC</td>
</tr>
<tr>
<td>AOP25</td>
<td>Male</td>
<td>61</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2481 + 102,2646 + 31del (Exon 18 deletion)</td>
</tr>
<tr>
<td>AOP26</td>
<td>Male</td>
<td>18</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2525delT</td>
</tr>
<tr>
<td>AOP27</td>
<td>Female</td>
<td>56</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2525delT</td>
</tr>
<tr>
<td>AOP28</td>
<td>Female</td>
<td>8</td>
<td>GAA</td>
<td>c.-32-13T&gt;G</td>
<td>c.2242dupG</td>
</tr>
</tbody>
</table>

Cases unusually rapid or very slow progression (Table S1). For example, a 32-year-old Caucasian male patient had initial symptoms in his late teens including onset of muscle weakness in arms and legs as well as elevated CK levels was diagnosed with LGMD, but unusual, rapid progression of proximal leg and girdle weakness occurred. Two homozygous variants in two muscle genes ANO5 and SGCA (Fig. 4; Table S1) could be potentially contributing to the unusual presentation in this individual. A 16-year-old male clinically diagnosed with LGMD, but with an unexpected severe progression had heterozygous loss of COL6A2 coupled with a homozygous loss of ANO5 possibly contributing to the rapid disease progression and severe phenotype (Fig. 4; Table S1). In this family, the 40-year-old mother with heterozygous loss of both ANO5 and COL6A2 showed mild myopathy, whereas her 8-year-old daughter with the same genotype had no muscle weakness indicating variable presentation and the possibility of later-onset clinical features. There are multiple other patient cases identified with single pathogenic variant in one recessive gene and another single VUS in a different recessive gene with no other reportable variant.

**Deep intronic variant in DYSF gene**

Recently, we characterized a deep intronic mRNA splice-altering pathogenic variant c.4886+1249G>T in DYSF gene that inserts a novel 177 bp pseudoexon at the exon 44–45 junction. This splicing causes an in-frame insertion of 59 additional amino acids in the dysferlin protein. This intronic variant segregated in a compound heterozygous state with the other pathogenic variant in two siblings (Fig. 5). Dysferlin protein expression levels in PBMC showed an evident pathogenic effect of the deep intronic splice variant in significantly reducing DYSF expression in the compound heterozygous states (Fig. 5). We targeted this deep intronic region in the current LGMD NGS panel and successfully identified it in one family. Overall we have been able to identify 45 different deep intronic variants including pathogenic or VUS across nine genes in our panel (DMD, CAPN3, DNAJB6, and...
GAA, DYSF, FKTN, COL6A1, COL6A2, COL6A3) in all different unrelated patients in our cohort (additionally see Tables S3–S7).

**Collagen family genes**

Collagen family genes COL6A1/A2/A3 are associated with both autosomal-dominant and recessive forms of muscular dystrophies. Variants located in the Gly-X-Y domain-encoding region of these genes are associated with dominant forms of myopathy. Here we report 55 patients with pathogenic variants in the Gly-X-Y domain of these three genes; COL6A1 is the major contributor with 35 patients. Eleven patients harbored two pathogenic variants in one of these genes, with eight patients having two pathogenic variants in COL6A2 indicating a role in recessive myopathy. Seventy-nine patients were identified with one pathogenic variant in any of these three genes (Table S5).

**Discussion**

The severity of the different LGMD subtypes varies with most of the dominant forms having later-onset ages and milder-clinical presentations than the recessive subtypes. A large number of affected patients remain without a definitive molecular diagnosis regardless of >30 LGMD genes known, because of the current challenges of interpreting variants such as deep-intronic or regulatory variants in known genes, or novel yet-to-be implicated genes. Most of the LGMD subtypes are genetically poorly studied, and limited knowledge is available regarding the mutation spectrum of different LGMD genes. This clinical-diagnostic program provided a valuable opportunity to gain a deeper understanding of rare LGMD subtypes by studying 4656 patients throughout the USA. Moreover, since majority of these patients had the adult-onset disease, our study adds to the knowledge base for use in newborn-genetic-screening programs that will enable earlier intervention if available. Major contributors to LGMD phenotypes were found to be CAPN3 (17%), DYSF (16%), FKRP (9%), and ANOS (7%) (Fig. 1A). No reportable variants were identified in 24% patients indicating a possible role of deep-intronic/regulatory-variants/novel genes.

Our LGMD panel diagnostic yield (27%) is smaller than that of our previous NMD panel (46%) since previously yield was for all NMDs, not just for LGMD, and the cohort was much smaller (81 patients). Identification of pathogenic variants in dominant LGMD genes DNAJB6, MYOT, and CAV3, but not in the TNPO3 gene indicates a higher prevalence of dominant LGMD-subtypes 1E,1A,1C, respectively, and the rarity of LGMD-1F. The 27% diagnostic yield included the 256 patients with one pathogenic and one VUS in the same LGMD gene since it is most likely these VUSs are likely pathogenic.
the patients show clinical symptoms and no other variant was identified in the same or any other gene. The diagnostic yield of our study is slightly higher than some previous exome sequencing studies in neurological disorders, even with a substantial patient cohort recruited based on specific disease phenotypes.41,42 This suggests the importance of using screening criteria of disease-specific clinical information carefully to judge the type of NGS-diagnostic testing for patient recruitment (fig. 3 in 43) and justifying subsequent advantage of targeted NGS panel over exome sequencing.

Due to overlapping phenotypes and complex presentations, it is difficult for clinicians to provide an accurate clinical diagnosis especially for Duchenne/Becker muscular dystrophy (DMD/BMD) patients with milder presentations, which often mimics limb-girdle phenotypes. Unlike some well-known LGMD gene panels from popular commercial diagnostic clinics, we included common X-linked muscular dystrophy genes DMD (OMIM# 300377), EMD (OMIM# 300384), FHL1 (OMIM# 300163) in the NGS LGMD-panel, and 43 patients were identified with pathogenic variants in DMD. Also, the complete lack of coverage for consecutive exons in the DMD gene was also observed in 10 patients indicating possible pathogenic deletions. These pathogenic deletions stress the importance of comprehensive clinical evaluation for DMD/BMD patients with a milder-presentation before referring to NGS-panel testing for other muscular dystrophies.43

Dysferlinopathies (LGMD2B and Miyoshi myopathy) include autosomal-recessive muscle diseases caused by pathogenic variants in dysferlin gene (DYSF), characterized by proximal muscle weakness, difficulty in running, climbing, increased fatigue.44,45 DYSF, CAPN3, and COL6A1 have the highest number of pathogenic variants indicating more allelic heterogeneity in these genes compared to others in the panel (Fig. 1A). Comparing Figure 1A and B, it is seen that FKRP, the third most contributing gene (Fig. 1A), has low allelic heterogeneity (lesser number of unique pathogenic variants), whereas TTN harbors high number of unique pathogenic variants (higher allelic heterogeneity) (Fig. 1B) even though being one of the least-contributing genes. This indicates that even greater number of unique pathogenic variants such as deep-intronic/deletion/duplications could potentially be identified in TTN with confirmed LGMD2J diagnosis, if we use whole genome sequencing instead of panel testing. These indicate therefore, LGMD overall might still remain clinically under diagnosed. Moreover, extensive homozygosity of variants in DYSF, FKRP, and CAPN3 genes highlights the immediate need for carrier testing for these LGMD subtypes (Fig. 3). The 79 patients identified with one pathogenic variant in any of the three collagen genes (Table S5) may not be of diagnostic value since COL6A1/A2/A3 variants can be either recessive/dominant, depending on the individual’s family history. Hence further functional studies using target muscle biopsy tissue, segregation analysis and nature history studies are needed.

Figure 5. Segregation analysis of a deep intronic DYSF variant. Variant c.4886 + 1249G>T in DYSF gene was identified in a large family with LGMD2B. A deep intronic variant in DYSF gene alters mRNA splicing and ultimately results in inframe insertion of a new pseudoexon in dysferlin. Percentages indicate levels of expression of dysferlin protein in peripheral blood mononuclear cells (PBMCs) compared to the control.
for those patients with one pathogenic variant in a collagen gene.

Late-onset Pompe disease (LOPD) is a rare, but potentially treatable lysosomal storage disorder characterized by progressive glycogen accumulation and muscle weakness, often with a limb-girdle pattern of weakness. It is essential that such patients should not be overlooked. Increased prevalence (2.4–3.63%) in LOPD was observed previously in LGMD-suspected or hyperCKemia patients using targeted sequencing approach.46–48 But these studies are different with regard to much lower number of patient cohort that unlike our study also included non-LGMD other NMD patients such as other congenital myopathies, distal myopathies, and isolated hyperCKemia in European population. Our results of prevalence of LOPD through targeted LGMD gene-panel testing in a very large strictly clinically LGMD-characterized or -suspected patients across US with multiple ethnicities indicate that LOPD should not be overlooked when considering potential diagnoses for LGMD patients even if clinical subtyping can be performed, that in turn will assist in ERT or recruitment for other clinical trials.

This large-scale study confirmed that LGMD2A (calpainopathy), characterized by severe-to-mild, symmetric, progressive weakness of proximal limb-girdle muscles with 2–40 years onset-age, as the most common LGMD subtype with the higher prevalence of dominant LGMD forms. LGMD2A was considered a strictly autosomal-recessive-LGMD subtype for many years, but recently patients carrying a single pathogenic variant in the CAPN3 have been reported. This variant, a 21 bp in-frame deletion c.643_663del21, cosegregates in 37 individuals from 10 families with muscle disease and dominant-transmission through several generations. Moreover, it is associated with severe loss of calpain3 expression shown by immunoblotting, suggesting a dominant-negative effect on calpain3 homodimer.29 In our large-scale study, the identification of the same 21 bp in-frame deletion in 17 individuals (Table 2) without any reportable second variant further confirms the c.643_663del21 as a common, dominant mutation, but further deletion-duplication analysis did not occur. Even CAPN3 inheritance is considered only autosomal recessive in OMIM (https://www.omim.org/entry/114240). Hence it is important to emphasize that through our very large sequencing program, we now know that dominant inheritance of CAPN3 is far more prevalent than previously thought, and should be widely considered in molecular diagnostic pipeline and NGS data analysis and interpretation in the clinics. Our finding of a similar 15 bp in-frame deletion c.598_612del15 in 16 patients with ages ranging from 48 to 76 years (Table 2) without a second pathogenic variant indicate a possible autosomal-dominant form of calpainopathy. The same 15 bp in-frame deletion Table 2. CAPN3 patients with in-frame deletions associated with autosomal-dominant subtype.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Gender</th>
<th>Age (years)</th>
<th>CAPN3 variant 1</th>
<th>CAPN3 variant 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C100</td>
<td>Male</td>
<td>24</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C200</td>
<td>Male</td>
<td>72</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C300</td>
<td>Male</td>
<td>44</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C400</td>
<td>Female</td>
<td>70</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C500</td>
<td>Female</td>
<td>58</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C600</td>
<td>Female</td>
<td>73</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C700</td>
<td>Male</td>
<td>57</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C800</td>
<td>Male</td>
<td>60</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C900</td>
<td>Female</td>
<td>61</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C111</td>
<td>Male</td>
<td>69</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C222</td>
<td>Female</td>
<td>17</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C333</td>
<td>Male</td>
<td>57</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C444</td>
<td>Female</td>
<td>34</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C555</td>
<td>Male</td>
<td>45</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C666</td>
<td>Male</td>
<td>13</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C777</td>
<td>Female</td>
<td>78</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>C888</td>
<td>Female</td>
<td>73</td>
<td>c.643_663del21</td>
<td>—</td>
</tr>
<tr>
<td>D001</td>
<td>Male</td>
<td>59</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D002</td>
<td>Female</td>
<td>57</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D003</td>
<td>Female</td>
<td>48</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D004</td>
<td>Male</td>
<td>69</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D005</td>
<td>Female</td>
<td>59</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D006</td>
<td>Male</td>
<td>57</td>
<td>c.598_612del15</td>
<td>c.794C&gt;T (p.S265F), VUS</td>
</tr>
<tr>
<td>D007</td>
<td>Female</td>
<td>58</td>
<td>c.598_612del15</td>
<td>c.1477C&gt;T (p.R493W), VUS</td>
</tr>
<tr>
<td>D008</td>
<td>Female</td>
<td>54</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D009</td>
<td>Female</td>
<td>57</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D010</td>
<td>Male</td>
<td>76</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D011</td>
<td>Female</td>
<td>53</td>
<td>c.598_612del15</td>
<td>c.1505T&gt;C (p.I502T), VUS</td>
</tr>
<tr>
<td>D012</td>
<td>Female</td>
<td>54</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D013</td>
<td>Male</td>
<td>39</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D014</td>
<td>Female</td>
<td>59</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D015</td>
<td>Female</td>
<td>69</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
<tr>
<td>D016</td>
<td>Male</td>
<td>66</td>
<td>c.598_612del15</td>
<td>—</td>
</tr>
</tbody>
</table>

VUS, Variant of uncertain significance.

splice variant (Table S6) with most of them located in the mutational-hotspot region of exon-5 in the G/F-rich linker domain. Only one reported variant c.525C>G(p.F175L) is located in exon-7 and no nonsense/frameshift variants have been reported. We report 29 patients with pathogenic missense variants in the exon-5 hotspot; but also identified 13 patients with novel missense variants outside of exon-5 (exons 4, 6, 7, 8, 9, 10) of DNAJB6 (Table S4) with no other reportable variants in an LGMD gene. These data indicate a broad mutation spectrum for DNAJB6 and a need for natural-history studies to understand the penetrance and nature of disease association. These 13 missense variants are further classified as VUS due to lack of published studies on LGMD1E and signifies the need for functional validation for reclassification of the VUSs and natural history studies on these patients to determine nature of variant contribution to disease.

Our targeted NGS panel data identified at least 31 individuals with pathogenic variants in more than one LGMD gene suggesting potential digenic/multiple gene contributions to LGMD disease presentation/progression (Fig. 4; Table S1). The explicit possible association of multigenic pathogenicity to unusual disease presentation with rapid/very slow progression seen in all 31 cases (Table S1) in our cohort including examples given in above (Results), suggests a potential high prevalence of gene-gene epistatic interaction and synergistic heterozygosity in LGMD. Albeit, further functional and segregation studies are needed to validate the complex inheritance patterns of LGMD subtypes. Moreover, we expect whole genome sequencing is needed on these patients to confirm if there are any other reportable pathogenic variant that could definitively diagnose these patients, before performing functional and segregation studies on them to understand multigenic contribution.

The findings that 72% identified reportable variants are classified as VUSs, and 23% of the variants are classified as pathogenic variants indicate the importance of further functional studies including RNA sequencing, enzyme assays or mass spectrometry to prove pathogenicity of identified VUSs. This will further aid in resolving the diagnosis of patients with at least one VUS either in DYSF or CAPN3 or any of the three collagen genes (Table S7).

Overall, the current large-scale LGMD sequencing project has improved our knowledge of the gene-variant spectrum, mutational hotspots, genetic etiologies, and relative prevalence of different LGMD subtypes across the US using a large clinically characterized patient cohort, allowing timely management, participation of definitively diagnosed individuals in ongoing clinical trials through disease-specific registries. This large US-cohort not previously described for LGMD has also showed evidence for a new mechanism which has been previously described in handful of cases in the literature. We identified the higher prevalence of dominant forms of LGMD, CAPN3, and DNAJB6 in particular and novel digenic/multigenic (more than 2) possible contribution to unusual LGMD clinical presentations/progressions. Adult-onset Pompe patients, also identified in our LGMD-suspected cohort, now have the opportunity of seeking enzyme replacement therapy. The slightly higher diagnostic yield of this study compared to our previous exome studies on NMD patients and exomes of neurological disorders in general as mentioned before, as well as the identification of 24% negative cases suggests yet-to-be discovered genes or genetic events associated with LGMD-like clinical symptoms. To resolve these cases, we predict the need for clinical genome sequencing (GS) to capture the full genetic variation spectrum as seen in recent trends in large clinical diagnostics studies as a first-tier approach. We expect a shift from panel sequencing to clinical-GS with better, uniform coverage therefore leading fewer exon drop-out, to detect deep-intronic and copy-number variants, and ability to interrogate periodically without resequencing.

Acknowledgments
We are greatly indebted to all the study participants and their families without them this research would not have been possible. This work was jointly funded by the Muscular Dystrophy Association, and the Jain Foundation.

Author Contributions
MH had full access to all of the data in the study and had responsibility for the integrity and accuracy of the data and its analysis. MH, SC, BRRN, LR, PM, MW, and MH designed the study concept. All authors contributed to the acquisition, analysis, or interpretation of data. BRRN, SC, and MH drafted the manuscript. All authors contributed to the critical revision of the manuscript for important scientific or intellectual content. BRRN and SC performed statistical analysis. MH obtained funding for this study. MH provided administrative support and supervised the overall study. The corresponding author MH had final responsibility for the decision to submit for publication.

Conflict of Interest
Nothing to report.

References


**Supporting Information**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Methods.** Questionnaires for patient recruitment.

**Figure S1.** Pictorial representation of different muscle-related proteins associated with various subtypes of LGMD is shown here pictorially in an interlinked chemomechanical network of sarcomere, sarcoplasm, sarcolemma, and extracellular matrix.
Figure S2. Contribution of different LGMD genes. Molecular diagnosis has been established in 27% of the patients with the majority having pathogenic variants identified in one of the following genes CAPN3 (17%), DYSF (16%), FKRP (9%), and ANO5 (7%) indicating that these genes are likely the major contributors to LGMD like phenotype.

Table S1. Summarized list of some of the affected cases with pathogenic variants in multiple genes and unusual clinical presentations.

Table S2. List of genes in the targeted LGMD panel testing with their respective LGMD subtype or other associated neuromuscular disease with/without overlapping LGMD subtype.

Table S3. Summary of late onset patients with one pathogenic variant and one variant of uncertain significance in the GAA gene.

Table S4. Summary of patients with DNAJB6 variants.

Table S5. List of patients with pathogenic variants in COL6A1, COL6A2, and COL6A3 genes.

Table S6. Mutation spectrum of DNAJB6 gene. List of variants reported in the Human Gene Mutation Database (HGMD).

Table S7. List of variants of uncertain significance (VUS) identified in major LGMD genes.