



HHS Public Access

Author manuscript

Genet Med. Author manuscript; available in PMC 2016 May 18.

Published in final edited form as:

Genet Med. 2015 December ; 17(12): 1002–1006. doi:10.1038/gim.2015.9.

Determining the prevalence of McArdle disease from gene frequency by analysis of next generation sequencing data:

McArdle prevalence by NGS data

Mauricio De Castro, M.D¹, Jennifer Johnston, PhD¹, and Leslie Biesecker, M.D¹

¹National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA

Abstract

Purpose—McArdle disease is one of the most common glycogen storage disorders. Although the exact prevalence is not known, it has been estimated to be 1 in 100,000 patients in the United States. More than 100 mutations in *PYGM* have been associated with this disorder. McArdle disease has significant clinical variability with some patients presenting with severe muscle pain and weakness while others have only mild, exercise-related symptoms.

Methods—Next-Generation sequencing data allow estimation of disease prevalence with minimal ascertainment bias. We analyzed gene frequencies in two cohorts of patients from exome sequencing results. We categorized variants into three groups: a curated set of published mutations, variants of uncertain significance, and likely benign variants.

Results—An initial estimate based on the frequency of six common mutations predicts a disease prevalence of 1/7,650 (95% CI 1/5,362 to 1/11,108), which greatly deviates from published estimates. A second method using the two most common mutations predicts a prevalence of 1/42,355 (95% CI 1/24,536 - 1/76,310) in Caucasians.

Conclusions—These results suggest that the currently accepted prevalence of McArdle disease is an underestimate and that some of the currently considered pathogenic variants are likely benign.

Keywords

Prevalence; Next-generation sequencing; allele frequency; McArdle disease; *PYGM*

INTRODUCTION

McArdle disease (Glycogen storage disease type V) is an inherited disorder of glycogen metabolism that affects exclusively skeletal muscle. Initially described in 1951 by British physician Brian McArdle who described a patient with exercise intolerance that failed to

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Mauricio De Castro, M.D, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 35 Convent Drive, MSC 3717, Room 1B207, Bethesda, MD 20892-3717, (301) 451-7417 phone, (301) 496-7184 fax, ; Email: Mauricio.decastro-pretelt@nih.gov

produce lactate. Symptoms consist of rapid fatigue, myalgia, and cramping associated with exercise. There is clinical variability with some patients having mild symptoms (fatigue or poor stamina) related to exercise¹ while others have more pronounced proximal muscle weakness². A fatal, rapidly progressive neonatal form with widespread muscle weakness has also been reported³. A classic finding in patients with the disease is the rapid improvement of symptoms with rest (so called “second-wind phenomenon”). In mildly to moderately affected patients, the clinical diagnosis requires a high degree of suspicion, especially in older patients in which the only symptom can be exercise intolerance. The diagnosis is confirmed with identification of biallelic pathogenic variants in the *PYGM* gene that encodes for the muscle phosphorylase protein, the only gene known to be associated with McArdle disease⁴. If the results are unclear, muscle biopsy with measurement of phosphorylase enzyme activity can be helpful. A less invasive, recently described method includes the use of antibodies to determine the expression of *PYGM* in white blood cells⁵.

The prevalence of McArdle disease has been reported to be 1 in 100,000 in the US⁶, at least 1 in 170,000 in Spain⁷ and 1 in 350,000 in the Netherlands⁸. In Spain and the Netherlands, the calculations were based on the number of affected individuals from national McArdle disease registries. Because McArdle disease can cause mild symptoms, it is possible that an estimate of prevalence based on ascertainment by clinical presentation to a metabolic disease expert could severely underestimate the prevalence. Access to exome sequencing data allows us to estimate the prevalence of this disorder based on carrier frequency using the Hardy-Weinberg equilibrium, reducing the bias associated with clinical ascertainment.

MATERIALS AND METHODS

We evaluated variant call data from the ClinSeq[®] cohort (n=951) and the NHLBI GO Exome Sequencing Project (ESP) (n=4,297 EA and 2,201 AA). The ClinSeq[®] cohort is composed of 951 patients predominantly of Caucasian descent ascertained for their family history of cardiovascular disease, participants are otherwise healthy and were not selected for known muscular conditions or symptoms. The ESP cohort is composed of several groups of patients, most of the patients have a personal or family history of cardiovascular or pulmonary disease, some of them are healthy controls while others are affected with hyperlipidemia, cardiovascular disease, or other associated conditions. None of the cohorts were selected for primary muscle disease. We first analyzed variant calls for the *PYGM* gene in the ClinSeq[®] database, materials and methods for the ClinSeq[®] study are described elsewhere⁹; DNA isolation, library preparation, capture, sequencing and alignment and base calling were performed as described in previous reports¹⁰. *PYGM* variant analysis was performed in VarSifter v1.6¹¹. Variants were filtered for mutation type and population frequency.

Variants that met population frequency (MAF <0.5% in ClinSeq[®] and ESP) and quality filters were further classified by cross-referencing them with mutations in the Human Gene Mutation database (HGMD). The pathogenicity of these variants was evaluated by reviewing publications with clinical, functional, and/or genetic data. To be considered pathogenic, a variant had to be reported in the literature in a patient with classical manifestations of the disease with compatible ancillary testing (e.g., characteristic muscle biopsy, absent muscle

phosphorylase levels, or second-wind phenomenon on treadmill testing) and the identification of biallelic variants in *PYGM*. The phase of the variants had to be known and appropriate Mendelian segregation confirmed. For variants not described in the literature, further classification was limited to allele frequency in the general population and in-silico model predictions: PolyPhen-2, SIFT¹² and CADD (Combined annotation dependent depletion) score¹³. Variants that did not meet our criteria for classification as pathogenic, were predicted to be deleterious by all four models and had a MAF<0.5% were considered to be variants of uncertain significance (VOUS). Variants with a MAF>0.5% or unpublished variants predicted to be benign by one or more in silico models were considered to be likely benign.

Statistical analysis for the 95% confidence intervals was performed using the exact binomial method based on the beta distribution as described by Clopper and Pearson¹⁴. Variants p.Arg50* and p.Gly205Ser were Sanger verified for the ClinSeq® cohort, Sanger validation is not possible for variants in the ESP cohort.

RESULTS

The ClinSeq® data were evaluated first. Two variants were excluded (p.Thr395Met and p.Arg414Gly) because they were above the frequency limit. We were left with 59/951 ClinSeq® participants who had among them 27 *PYGM* variants (Table 1). No participant had two minor alleles. Fifteen participants were heterozygous for one of six published mutations. Thirteen participants were heterozygous for 12 VOUS and 31 participants were heterozygous for nine likely benign variants. We then evaluated the ESP dataset for European Americans for the mutations that we identified in ClinSeq®. In the ESP EA dataset, 105 participants were heterozygous for one of the six published mutations. Twenty-six participants were heterozygous for six of the 12 VOUS and 64 participants were heterozygous for one of the nine likely benign variants.

To increase power, we combined our results with data from the ESP project, which yielded 5,248 exomes. Although there were no homozygotes for any of these variants in the NHLBI ESP, we could not exclude compound heterozygosity because that database does not provide these data. A total of 27 variants were considered amongst 59 individuals from the 951 participants in ClinSeq®. Six of these 27 variants have been claimed to be pathogenic in prior publications. These six variants, which were present in a total of 15 participants for a MAF of 0.00789, predict a disease prevalence of 1/16,080 (95% CI 1/5,940-1/51,163). Because the confidence intervals of this estimate were so large, we expanded our dataset by analyzing the NHLBI ESP EA, for a total of 5,248 individuals. Between the two datasets, there were a total of 120 participants with one of the six pathogenic variants, for a MAF of 0.0114, which predicts a prevalence of 1/7,650 (95% CI 1/5,362 to 1/11,108).

Given the discrepancy with published estimates, we critically evaluated the evidence supporting the pathogenicity of the variants and rank ordered them from most evidence to least evidence. The p.Arg50* variant was the highest ranked since it is present in large numbers of affected individuals as compared to controls and has been shown to undergo non-sense mediated decay in muscle tissue from patients with McArdle disease¹⁵, we

calculated the predicted disease prevalence based on that variant alone. In the combined ClinSeq[®] and ESP EA data, the MAF for this variant was 0.00313, which predicts a disease prevalence of 1/101,166 (95% CI 1/51,349 – 1/213,345). We then took the variant with the next most strong evidence, p.Gly205Ser, and added the frequencies of that variant to p.Arg50* and estimated the frequency of the disease, this variant is located in a critical region for tetramerization of the PYGM enzyme and mutations in residue 205 have been shown to lead to misfolding of the protein in human cell lines¹⁶. The MAF of those two variants in the combined data set were 0.00352, which predicts a disease prevalence of 1/80,478 (95% CI 1/42,407 – 1/162,198). This series of calculations was continued for all six mutations, showing that the previous estimated prevalence of the disease is accounted for by only the p.Arg50* variant and that the upper 95% confidence interval of our calculations falls to about 1/100,000 when accounting for only three mutations (Figure 1). Indeed, by using all six of the published variants identified in ClinSeq[®], the predicted disease frequency is far more common than prior estimates. Although there are more than 100 reported *PYGM* mutations, we calculated a predicted disease prevalence of 1/7,650 (95% CI 1/5,362 to 1/11,108) using only six published mutations.

To provide yet another approach to these estimates, we calculated the prevalence by deriving the total fraction of all other pathogenic alleles using data from affected patients¹⁷. First, we tabulated the total mutation burden for the two most common mutations: p.Arg50* and p.Gly205Ser. The former is the most common mutation in McArdle disease, with the actual prevalence of the mutation varying among populations. The estimated prevalence in the US for p.Arg50* amongst patients with McArdle disease is 63%^{1,18}. p.Gly205Ser is the second most common mutation in Europe and the US, comprising about 9% of pathogenic alleles. The combination of these two alleles should account for 72% of alleles for McArdle disease in European Americans in the US. The prediction using both allele frequencies and assuming this accounted for 72% of causative alleles resulted in a prevalence of 1/42,355 (95% CI 1/24,536 - 1/76,310), which does not overlap with the currently estimated prevalence.

DISCUSSION

These data suggest that McArdle disease is significantly more common among European-derived Americans than the currently accepted 1/100,000 prevalence, and we conclude that the disorder is at least twice as common, in the range of 1/50,000. There are two potential explanations: 1) McArdle disease is under diagnosed and/or, 2) the penetrance of some of the variants in McArdle disease is overestimated. It is possible that some mutations in *PYGM* are not fully penetrant thus overestimating the prevalence when calculating from combined allele frequencies. We believe this is one of the strengths of the calculations that use only the two most common mutations (p.Arg50* and p.Gly205Ser), which all evidence to date suggests are fully penetrant. That both methods predict a higher frequency supports our thesis. Expressivity should also be considered – were there to be a wider range of expressivity than currently appreciated, there could be many patients who have a very mild form of this disease. This would be just as interesting and important – we suggest that a very mild form of McArdle disease could be present in a patient, not diagnosed as McArdle disease, but have significant implications for exercise tolerance. A separate issue to consider

is the possibility that many of the variants in McArdle disease are actually benign, which would erroneously increase the calculated prevalence (for instance the variant p.Ile513Val seems to be just as common as p.Arg50* in certain populations). We do not believe this to be valid, as our higher prevalence is supported by the method of extrapolating from only two variants that are essentially certain to be pathogenic, which makes the questions of individual pathogenicity assessment of other variants irrelevant. Nearly all variants other than p.Arg50* and p.Gly205Ser would have to be benign for the 95% CI of our estimates to overlap with the current prevalence estimate, which we think is an unreasonable hypothesis.

It is possible that some mutations in *PYGM* cause a very clinically mild phenotype of McArdle disease. This has been described for autosomal recessive metabolic disorders such as: biotinidase deficiency¹⁹, pyruvate kinase deficiency¹⁷ or Gaucher disease, but not for McArdle. Because McArdle disease is a condition with high clinical variability, symptoms can go unrecognized for many years before coming to diagnosis. It is possible that many affected patients develop an aversion to anaerobic exercise that does not limit their life enough to seek a diagnosis and as such, they are not included in current prevalence estimates.

There are some limitations to this approach. We assumed that McArdle is a monogenic condition and all variants can be accounted for by looking at *PYGM*. If locus heterogeneity were a possibility for McArdle disease then the prevalence of mutations would be higher than we are suggesting here. A second limitation is that for the NHLBI-ESP dataset, we are not able to ascertain the phase of the variants. Given that our estimates of prevalence are much higher than the inverse of the NHLBI-ESP dataset, we think this is unlikely to be an issue.

Finally, it is important to point out the technical limitations of identifying variants from next-generation sequencing data. Appropriate depth of coverage, deep intronic mutations, mutations in the promoter region and inability to detect large deletions or duplications would lead to under ascertainment of pathogenic variants. However, such an error would again make our estimate conservative, and the disease would be more common than we predict.

The estimation of disease frequency based on patients who present to specialty clinics is biased towards those with typical, recognizable, and more severe presentations. We predict that as sequencing is applied more widely in the clinic and in larger research cohorts that undiagnosed individuals with biallelic mutations in *PYGM* will be identified. This approach of genome-driven ascertainment (as opposed to phenotype-driven ascertainment) mitigates the inherent ascertainment bias towards more severe presentations. It will be important to identify patients by mutations and follow that with clinical research to elucidate the possible associated phenotype, which has been termed hypothesis-generating clinical research²⁰. Such identifications will allow a better appreciation of the true spectrum of clinical phenotypes associated with variation in this gene. We predict that a substantial number of such identified individuals will be found to have abnormal biochemistry and exercise tolerance, and that the full delineation of this phenotype will become a component of predictive medicine.

Acknowledgments

We would like to thank Mr. Neal Oden for his invaluable assistance with the statistical calculations.

References

1. el-Schahawi M, Tsujino S, Shanske S, DiMauro S. Diagnosis of McArdle's disease by molecular genetic analysis of blood. *Neurology*. 1996; 47:579–580. [PubMed: 8757044]
2. Wolfe GI, Baker NS, Haller RG, Burns DK, Barohn RJ. McArdle's disease presenting with asymmetric, late-onset arm weakness. *Muscle Nerve*. 2000; 23:641–64. [PubMed: 10716777]
3. Di Mauro S, Hartlage PL. Fatal infantile form of muscle phosphorylase deficiency. *Neurology*. 1978; 28:1124–1129. [PubMed: 101896]
4. Garcia-Consuegra I, Rubio JC, Nogales-Gadea G, et al. Novel mutations in patients with McArdle disease by analysis of skeletal muscle mRNA. *J Med Genet*. 2009; 46:198–202. [PubMed: 19251976]
5. de Luna N, Brull A, Lucia A, et al. PYGM expression analysis in white blood cells: A complementary tool for diagnosis McArdle disease? *Neuromuscul Disord*. 2014; 24:1079–86. [PubMed: 25240406]
6. Haller RG. Treatment of McArdle disease. *Arch Neurol*. 2000; 57:923–924. [PubMed: 10891971]
7. Lucia A, Ruiz JR, Santalla A, et al. Genotypic and phenotypic features of McArdle disease: insights from the Spanish national registry. *J Neurol Neurosurg Psychiatr*. 2012; 83:322–328. [PubMed: 22250184]
8. van Alfen N, de Bie HJ, Wevers RA, Arenas J, van Engelen BG. The prevalence and genetic characteristics of McArdle's disease in the Netherlands. *Neuromuscul Disord*. 2002; 12:718–783.
9. Biesecker LG, Mullikin JC, Facio FM, et al. The ClinSeq Project: Piloting large-scale genome sequencing for research in genomic medicine. *Genome Res*. 2009; 19:1665–1674. [PubMed: 19602640]
10. Johnston JJ, Rubinstein WS, Facio FM, et al. Secondary variants in individuals undergoing exome sequencing: Screening of 572 individuals identifies high-penetrance mutations in cancer susceptibility genes. *Am J Hum Genet*. 2012; 91:97–108. [PubMed: 22703879]
11. Teer JK, Green ED, Mullikin JC, Biesecker LG. Varsifter: visualizing and analyzing exome-scale sequence variation data on a desktop computer. *Bioinformatics*. 2012; 28:559–600.
12. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*. 2001; 11:863–874. [PubMed: 11337480]
13. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014; 46:310–315. [PubMed: 24487276]
14. Clopper C, Pearson E. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934; 26:404–413.
15. Nogales-Gadea G, Rubio J, Fernandez-Cadenas I, et al. Expression of the muscle glycogen phosphorylase gene in patients with McArdle disease: the role of nonsense-mediated mRNA decay. *Hum Mutat*. 2008; 29:277–83. [PubMed: 17994553]
16. Birch K, Quinlivan R, Morris G. Cell models for McArdle disease and aminoglycoside-induced read-through of a premature termination codon. *Neuromuscul Disord*. 2013; 23:43–51. [PubMed: 22818872]
17. Beutler E, Gelbart T. Estimating the prevalence of pyruvate kinase deficiency from the gene frequency in the general white population. *Blood*. 2000; 95:3585–3588. [PubMed: 10828047]
18. Tsujino S, Shanske S, DiMauro S. Molecular genetic heterogeneity of myophosphorylase deficiency (McArdle's disease). *N Engl J Med*. 1993; 329:241–245. [PubMed: 8316268]
19. Wolf B, Norrgard K, Pomponio RJ, et al. Profound biotinidase deficiency in two asymptomatic adults. *Am J Med Genet*. 1997; 73:5–9. [PubMed: 9375914]
20. Biesecker LG. Hypothesis-generating research and predictive medicine. *Genome Res*. 2013; 23:1051–1053. [PubMed: 23817045]

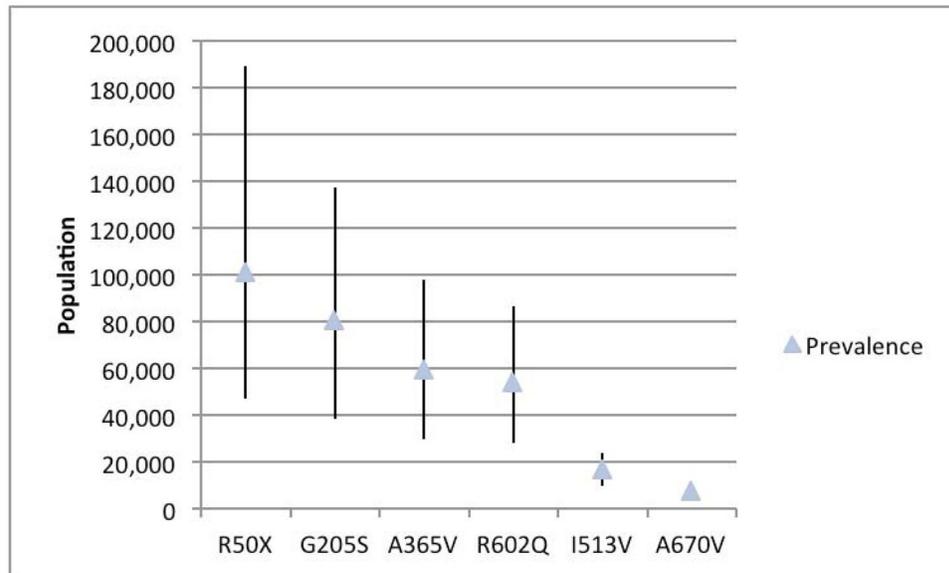


FIGURE 1. Ordinal mutation prevalence. Prevalence estimate with 95% CI starting with the mutation with the most evidence for pathogenicity (p.Arg50*) and subsequently adding published mutations in decreasing order of evidence for pathogenicity.

Table 1

Variants evaluated in this study

cDNA	AA change	Number of individuals with variant in ClinSeq®	Number of individuals with variant in E/A ESP	Total ClinSeq® + ESP EA	Number of individuals with variant in AA ESP	Variant Classification	Published in the literature	SIFT	Polyphen	CADD
c.148C>T	p.R50*	6	27	33	2	Pathogenic	Yes	LOF	LOF	40
c.613G>A	p.G205S	1	3	4	0	Pathogenic	Yes	DAMAGING	PROBABLY DAMAGING	36
c.1094C>T	p.A365V	2	4	6	2	Pathogenic	Yes	DAMAGING	PROBABLY DAMAGING	23.6
c.1537A>G	p.I513V	1	35	36	3	Pathogenic	Yes	TOLERATED	BENIGN	10.83
c.1805G>A	p.R602Q	1	0	1	0	Pathogenic	Yes	DAMAGING	PROBABLY DAMAGING	36
c.2009C>T	p.A670V	4	35	39	6	Pathogenic	Yes	DAMAGING	PROBABLY DAMAGING	35
c.100C>T	p.R34W	1	0	1	0	VOUS	No	DAMAGING	PROBABLY DAMAGING	35
c.209G>A	p.R70H	1	0	2	0	VOUS	No	DAMAGING	PROBABLY DAMAGING	35
c.482G>A	p.R161H	1	0	0	0	VOUS	No	DAMAGING	PROBABLY DAMAGING	34
c.832C>T	p.R278C	1	0	0	0	VOUS	No	DAMAGING	PROBABLY DAMAGING	22.1
c.848A>G	p.N283S	2	12	14	2	VOUS	No	DAMAGING	PROBABLY DAMAGING	24.2
c.1160G>A	p.R387H	1	0	1	0	VOUS	No	DAMAGING	PROBABLY DAMAGING	31
c.1558C>T	p.R520C	1	2	3	2	VOUS	No	DAMAGING	PROBABLY DAMAGING	22.8
c.1885G>T	p.D629Y	1	1	2	5	VOUS	No	DAMAGING	PROBABLY DAMAGING	24.89
c.2083G>A	p.G695R	1	0	1	0	VOUS	No	DAMAGING	PROBABLY DAMAGING	28.3
c.2446C>T	p.R816C	1	1	2	0	VOUS	No	DAMAGING	PROBABLY DAMAGING	20.9
c.2467C>T	p.R823W	1	1	2	0	VOUS	No	DAMAGING	PROBABLY DAMAGING	19.07
c.2500C>T	p.R834C	1	0	1	0	VOUS	No	DAMAGING	PROBABLY DAMAGING	20.4