Journal of Neuromuscular Diseases 1 (2014) 135–149
DOI 10.3233/JND-140032
IOS Press

Review

The Challenge of Next Generation Sequencing in the Context of Neuromuscular Diseases

Monkol Lek[a,b,c,d,e,*] and Daniel MacArthur[a,b,c]
[a]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
[b]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA
[c]Department of Medicine, Harvard Medical School, Boston, MA, USA
[d]Paediatrics and Child Health, University of Sydney, Sydney, NSW, Australia
[e]Institute for Neuroscience and Muscle Research, Sydney, NSW, Australia

Abstract. New genomic technologies, such as exome, whole-genome and transcriptome sequencing, are transforming the genetic diagnosis of neuromuscular diseases and dramatically accelerating the discovery of new disease-associated genes. The increasingly widespread availability of these technologies creates both opportunities and challenges for neuromuscular disease researchers. Here we survey the current literature on the application of new genomic technologies to the diagnosis of severe muscle diseases, with a focus on assessment of the approaches used for data processing, analysis and interpretation. We also highlight several key areas requiring improvement.

Keywords: exome sequencing, neurogenetics, muscle disease, next generation sequencing

INTRODUCTION

Next-generation sequencing (NGS) represents perhaps the most transformative technological advance in biomedical science since the development of the optical microscope. In less than a decade, the development and widespread adoption of short-read DNA sequencing has altered nearly every area of biomedical research, providing access to high-resolution biological data in areas ranging from DNA multiplex labeling for microscopy [1] to rapidly identifying pathogenic strains of *E. coli* during outbreaks [2].

One area that has been particularly impacted by NGS is the diagnosis of rare diseases. As in other disease areas, genomic approaches have dramatically accelerated gene discovery and diagnosis in muscle diseases (Fig. 1). These technologies have been used to identify the genetic cause of three subtypes of dominant limb girdle muscular dystrophy (LGMD1D [3–5], LGMD1F [6], LGMD1G [7]) that were localized by linkage analysis a decade ago. Similarly, it has identified the gene for Welander distal muscular dystrophy as a common cause of muscle disease in an isolated population [8], which was also localized a decade ago. In addition, NGS has enabled researchers to identify the disease gene modifier, *SMCHD1* in Facioscapulohumeral muscular dystrophy (FSHD) type 2 patients [9] that do not have the D4Z4 repeat contraction, which may shed light on non-manifesting carriers of this presumed pathogenic repeat contraction [10]. Also, two more genes from the Kelch-like (KLHL) family have been associated with nemaline myopathy [11, 12].

In the field of rare genetic disease research, there are three broad approaches enabled by NGS technology. Due to cost-effectiveness the most commonly employed is whole exome sequencing, which involves the targeted capture and sequencing of the coding regions of the genome (i.e. exome). Also, targeted capture may be limited to gene panels in which complete

*Correspondence to: Monkol Lek, Tel.: +1 617 726 5936; Fax: +1 617 643 3293; E-mail: lek@atgu.mgh.harvard.edu.

ISSN 2214-3599/14/$27.50

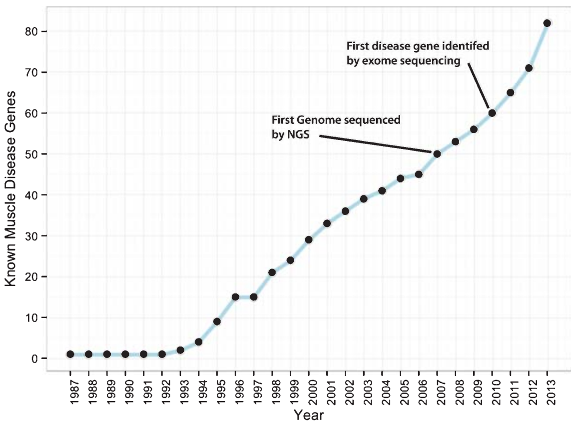

Fig. 1. The impact of Next Generation Sequencing in muscle disease gene discovery. The disease muscle disease genes were taken from the 2014 version of the gene table of monogenic neuromuscular disorders [75]. It was restricted to groups 1–5 and the first publication date the gene was identified.

and high coverage capture can be achieved and may be preferred in diagnostics to reduce false negatives. The next is whole genome sequencing, which provides a relatively unbiased survey of the genome including non-coding variants and structural variants. Lastly, RNA sequencing provides an unbiased survey of transcript isoform diversity provided by splicing. These technologies have been the subject of several recent reviews [13–16].

Genome and exome sequencing approaches provide several key advantages over traditional linkage/candidate gene approaches. Firstly, they can be used to discover genes in families with inheritance models inaccessible to linkage (such as *de novo* dominant diseases). Secondly, they overcome the "stopping problem" associated with serial testing of candidate genes in patients, in which a diagnosis is provided based on the first compelling candidate variant identified in a tested gene, rather than the most compelling variant across all genes. Finally, NGS has empowered researchers to comprehensively study very large genes – common among muscle disease genes – that are largely inaccessible with traditional approaches (Table 1).

Exome sequencing is a powerful approach that allows researchers to identify vast numbers of potential candidate mutations. However, this power also brings a major challenge: with hundreds of rare, potentially functional variants in every genome, it can be all too easy to build a compelling but false story about causal variants, a problem referred to as the "narrative potential" of human genomes [17]. To address this challenge a US National Human Genome Research Institute working group was formed to discuss this challenge and recently published guidelines for implicating sequence variants to disease [18].

In this review we survey the current literature on the application of genomic technology to gene discovery in neuromuscular diseases (Table 1). The manuscripts included in this survey were from a PubMed search of the keywords "muscle" and "exome sequencing". This list was then reduced by removing cardiomyopathies and Walker-Warburg syndromes. Thus the review does not aim to be fully comprehensive including all novel genes and mutations discovered but rather a survey to assess the degree in which selected studies empowered by exome sequencing conforms to best-practice standards. A secondary aim is to review overall informatics workflows with a focus on under-appreciated aspects of these workflows.

NGS INFORMATICS PIPELINE/ WORKFLOWS

The majority of publications reporting new neuromuscular disease genes document the method, tools and resources used for exome sequencing analysis (Table 1). In high impact publications, space limitations mean that such details are often consigned to the supplementary methods, and they can sometimes be entirely omitted from brief case reports. Moving from raw NGS data to interpretable variant calls requires a number of critical steps to be performed. Firstly, raw sequencing reads are aligned to the reference genome, and additional processing and calibration steps performed to improve these alignments. For the mapping of reads, the freely available tool, Burrows-Wheeler Aligner (bwa) is the most widely used tool in the study surveyed. In addition, missing from Table 1 are commonly used tools/options Picard, Genome Analysis Toolkit (GATK) and samtools in pre-processing prior to variant calling. Secondly, the aligned reads at each position along the reference are scanned for evidence of a variant – a location at which the individual differs from the reference sequence, either in the heterozygous or homozygous state. Variants are then scored with statistical confidence, and filtered to remove systematic artifacts. Amongst reviewed muscle research, GATK and samtools were the most popular variant callers for this step. Furthermore, GATK is more widely used due to its user friendly but powerful variant filtering. As each of these steps can be performed using a range and combination of different informatic tools, making

Table 1
Informatic worflows employed for the discovery of neuromuscular disease genes and variants

Author	Gene	Inheritance	Disease	Families	Alignment	Variant calling	Annotation	Filtering	Pathogenicity prediction
Novel Gene									
Belaya et al. [79]	DPAGT1	AR	Limb-girdle congenital myasthenic syndrome	4	Novoalign	Samtools, Platypus	ANNOVAR	dbSNP, EVS, 1KG	Conservation, Polyphen
Carss et al. [72]	GMPPB	AR	CMD and LMGD with Hypoglycosylation of a-Dystroglycan	8	BWA, NextGENe	NextGENEe, GATK	dbNSFP, VEP	EVS, in house exomes	Conservation, Polyphen, SIFT
Gupta et al. [12]	KLHL41	AR	Nemaline myopathy	5	BWA	NA	ANNOVAR	dbSNP, EVS, 1KG	Polyphen, SIFT, pMUT
Hicks et al. [80]	COL12A1	AD	Extracellular matrix-related myopathy	2	Mosaik	Samtools, dindel	NA	EVS, 1KG	Mutation taster, Polyphen, SIFT
Klar et al. [8]	TIA1	AD	Welander distal myopathy	35	Lifescope	Lifescope	ANNOVAR	in house exomes, population controls	Conservation
Majczenko et al. [81]	CCDC78	AD	Congenital Myopathy with Prominent Nuclei and atypical cores	1	NA	NA	NA	dbSNP, EVS, 1KG	NA
Muhammad et al. [82]	HACD1	AR	Congenital Myopathy	1	BWA	GATK	Custom	dbSNP, EVS, 1KG, population controls	NA
Ravenscroft et al. [70]	KLHL40	AR	Nemaline myopathy	28	BWA, Novoalign	Samtools, GATK	ANNOVAR	dbSNP, EVS, 1KG	NA
Stevens et al.[73]	B3GALNT2	AR	CMD with Hypoglycosylation of a-Dystroglycan	6	NA	NA	ANNOVAR, VEP	dbSNP, 1KG, in house exomes	NA
Torella et al. [6]	TNPO3	AD	Limb girdle muscular dystrophy (LGMD) 1F	2	BWA	GATK	ANNOVAR	dbSNP, EVS, 1KG, in house exomes	LRT, Mutation taster, Polyphen, SIFT
Weterman et al. [83]	MYL2	AR	Infantile type I muscle fibre disease and cardiomyopathy	9	Bioscope, Varscan	Bioscope, Varscan	Bioscope, Varscan	NA	NA

Table 1
(*Continued*)

Author	Gene	Inheritance	Disease	Families	Alignment	Variant calling	Annotation	Filtering	Pathogenicity prediction
Lemmers et al. [9]	SMCHD1	Digenic	Facioscapulohumeral muscular dystrophy (FSHD) 2	19	MAQ	MAQ	SeattleSeq	dbSNP, 1KG, in house exomes	GERP
Logan et al. [71]	MICU1	AR	Core myopathy	7	Novoalign	GATK	NA	dbSNP, EVS, 1KG, UK10K, ClinSeq	Condel, Polyphen, SIFT
Spiegel et al. [36]	FDX1L	AR	Mitochondrial muscle myopathy	1	NA	NA	NA	EVS, population control	Mutation taster, Polyphen
Vieira et al. [7]	HNRPDL	AD	Limb girdle muscular dystrophy (LGMD) 1G	2	BWA	GATK	NA	EVS, 1KG, in house exomes, population control	NA
Novel Variant									
Bohm et al. [84]	RYR1	AD	Samaritan congential myopathy	1	BWA	Samtools	NA	dbSNP, 1KG, SVA	NA
Ceyham-Bisroy et al. [85]	TTN	AR	Centronuclear myopathies	5	BWA	Samtools	NA	dbSNP, EVS, 1KG	NA
Chaouch et al. [86]	SLC25A1	AR	Congenital myasthenic syndrome	2	BWA	Samtools and dindel	ANNOVAR	dbSNP, EVS, 1KG, in house exomes	Conservation
Chauveau et al. [87]	TTN	AR	Core myopathy with heart disease	4	NextGENe	NextGENe	NA	dbSNP, 1KG, population controls	NA
Cortese et al. [88]	CLN3	AR	Autophagic vacuolar myopathy (AVM)	1	Novoalign	Samtools	NA	NA	Polyphen, SIFT
Couthouis et al. [4]	DNAJB6	AD	Limb girdle muscular dystrophy (LGMD) 1D	1	bowtie2	GATK	ANNOVAR	dbSNP, EVS, 1KG	NA
Davidson et al. [89]	TPM2	AD	Core-rod myopathy	1	BWA	NA	NA	dbSNP, EVS, 1KG, in house exomes	Polyphen, SIFT, SeattleSNP
Esposito et al. [90]	ITGA7	AR	Congenital fiber type disproportion	1	MAQ and NextGENe	MAQ and NextGENe	SeattleSeq	dbSNP, EVS, 1KG, population control	Polyphen, SIFT
Harms et al. [5]	DNAJB6	AD	Limb girdle muscular dystrophy (LGMD) 1D	2	Novoalign	Samtools	SeattleSeq	dbSNP, 1KG, in house exomes	SIFT
Hedberg et al. [65]	STIM1	AD	Tubular aggregate myopathy	3	BWA	GATK	ANNOVAR	dbSNP, 1KG, in house exomes	Mutation taster, Polyphen, SIFT
Izumi et al. [91]	TTN	AD	Hereditary myopathy with early respiratory failure (HMERF)	1	BWA	GATK	ANNOVAR	dbSNP, 1KG, population controls	Polyphen

Table 1
(*Continued*)

Author	Gene	Inheritance	Disease	Families	Alignment	Variant calling	Annotation	Filtering	Pathogenicity prediction
Jimenez-Escrig et al. [92]	LMNA	AR	Emery-Dreifuss muscular dystrophy	1	NA	NA	NA	200 control chromosomes	Conservation
Komlosi et al. [66]	MYH7	AD	Laing distal myopathy	1	NA	NA	NA	NA	NA
Leidenroth et al. [67]	CAPN3	AR	Limb girdle muscular dystrophy (LGMD) 2A	1	NA	GATK	SeattleSeq	dbSNP, 1KG	NA
Malfatti et al. [93]	TPM3	AD	Combined cap disease and Nemaline myopathy	1	NA	NA	NA	NA	Conservation
Maselli et al. [94]	GFPT1	AR	Limb-girdle myasthenia	1	BWA	GATK	SeattleSeq	EVS, 1KG	NA
Mitsuhashi et al. [95]	SMCHD1	Digenic	Facioscapulohumeral muscular dystrophy (FSHD) 2	1	BWA	GATK	ANNOVAR	dbSNP, EVS, 1KG	LFT, Polyphen, PhyloP, SIFT
Ohlsson et al. [96]	TTN	AD	Hereditary myopathy with early respiratory failure (HMERF)	1	BWA	GATK	NA	dbSNP, 1KG, in house	PhyloP, SIFT
Palmio et al. [68]	TTN	AD	Hereditary myopathy with early respiratory failure (HMERF)	12	BWA	GATK	NA	dbSNP, EVS, 1KG, population controls	Conservation
Park et al. [97]	MYH7	AD	Laing distal myopathy	1	BWA	Samtools	NA	dbSNP, 1KG, population controls	Mupro, Polyphen, SIFT
Pfeffer et al. [98]	TTN	AD	Hereditary myopathy with early respiratory failure (HMERF)	3	BWA	Varscan, dindel	NA	dbSNP, in house exomes	Mutation taster
Raphael et al. [69]	GMPPB	AR	Congenital muscular dystrophy	1	bowtie2	GATK	ANNOVAR	dbSNP, EVS, 1KG	Conservation
Romero et al. [99]	MYH7	AD	Central core disease	1	NA	NA	NA	dbSNP, EVS, 1KG	Polyphen, SIFT
Schessl et al. [100]	MYOT	AR	Myofibrillar myopathy	1	NA	NA	NA	Population control	Alamut, Mutation taster, Polyphen, SIFT
Renesse et al. [101]	POMK	AR	Congenital muscular dystrophy	1	BWA	GATK	NA	dbSNP, EVS, 1KG, in house exomes	Mutation Taster

it challenging to compare results between projects, or indeed often between batches in the same project. The majority of researchers have adopted open source software, while a few have chosen commercial software.

Careful reporting of methods is critical for NGS applications, since tools and resources for NGS are in constant development, including new software versions as well as updates to resources such as the human genome reference sequence (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human). In addition, choice of gene models for annotation [19] and subtle differences in workflows can result in large discrepancies in variant discovery [20]. There have been many comprehensive reviews on workflows and quality control in the context of rare diseases [14, 17, 21] and more specifically rare muscle diseases [22–28], therefore this section will focus on under appreciated but important themes within sample validation, quality and computational requirements.

Sample quality control

A critical step before proceeding to variant interpretation is quality control. The presence of family data in many rare disease studies provides a valuable resource for quality control, since it can be used to identify Mendelian violations such as *de novo* mutations and transmission errors, as well as discordance between inferred and expected pedigree structures (see below). Other valuable quality metrics can also be directly inferred from sequence data, such as transition to transversion ratio (TiTv), insertion to deletion ratio and total number of variants per sample, all of which can be compared to previous reference samples to identify gross problems with sequence quality. The data from next generation sequencing, similar to other experiments, is also at risk of sample contamination (the accidental mixture of samples from different sources) even when extreme care is taken in each step of the workflow. The tool verifyBAMID [29], which samples across many common SNPs, is one widely-used method for assessing such contamination directly from sequence data. The Genome Analysis Toolkit (GATK) assumes samples have 0–5% contamination and can reliably correct genotype calls for samples with up to 10% contamination.

In targeted gene sequencing there are very few common variants discovered in each sample; in contrast, exome sequencing allows the discovery of thousands of common variants across genes. These common variants or a subset then can be used as a means to identify samples and also infer relationship between samples [30]. First, using identity by descent metrics calculated by PLINK [31], we can confidently infer first and second degree relationships between samples. This is particularly useful in pedigree study designs in identifying duplicate samples, non-paternity and accidental sample mislabeling. Second, these variants can be used to estimate the deviation from expected heterozygosity represented as inbreeding coefficient (F). A significantly negative inbreeding coefficient (i.e. higher than expected heterozygosity) suggests a sample may be contaminated, while a positive F may suggests a sample is a child from consanguineous parents. Interestingly, high heterozygosity in contaminated samples will make them appear to be related to all samples [31, 32]. Thirdly, heterozygosity of common variants on chromosome X can be used to infer the gender of the sample. In addition, normalized coverage of chromosome X and Y can also be used to infer gender [32]. Lastly, common variants from samples can be used in a principle component analysis with samples of known ethnicity (eg. 1000 genomes samples [33]) to infer ethnicity of samples (Fig. 2). The latter two methods is particular useful for consistency checking in cases where only the proband was sequenced. The correct labeling of samples is absolutely crucial for the discovery of candidate mutations and the above techniques highlight how data from exome sequencing can be used to infer labeling inconsistencies. Also, early discovery of inconsistencies can save time and resources.

Sequencing coverage

Generating sufficient and uniform coverage across muscle disease genes is a challenge due to their overall size, number of exons, high similarity, low sequence complexity and high GC content [24–26]. Although within-sample coverage across genes is important, consistent coverage across samples is equally important in pedigree study designs. For example, in samples with highly variable coverage, there will be many cases where there is inadequate coverage for a candidate mutation across some members in the pedigree. Variable coverage becomes an even greater problem when samples from the same pedigree are sequenced at different sequencing centers. The coverage metric is a proxy to infer the ability to call variants at a site but does not take into account the quality of the sequencing reads and confidence of the mapping to a location in the genome, therefore a better metric would be the callability of a site, which takes base quality into account. There are regions in *TTN* and *NEB* where exons are

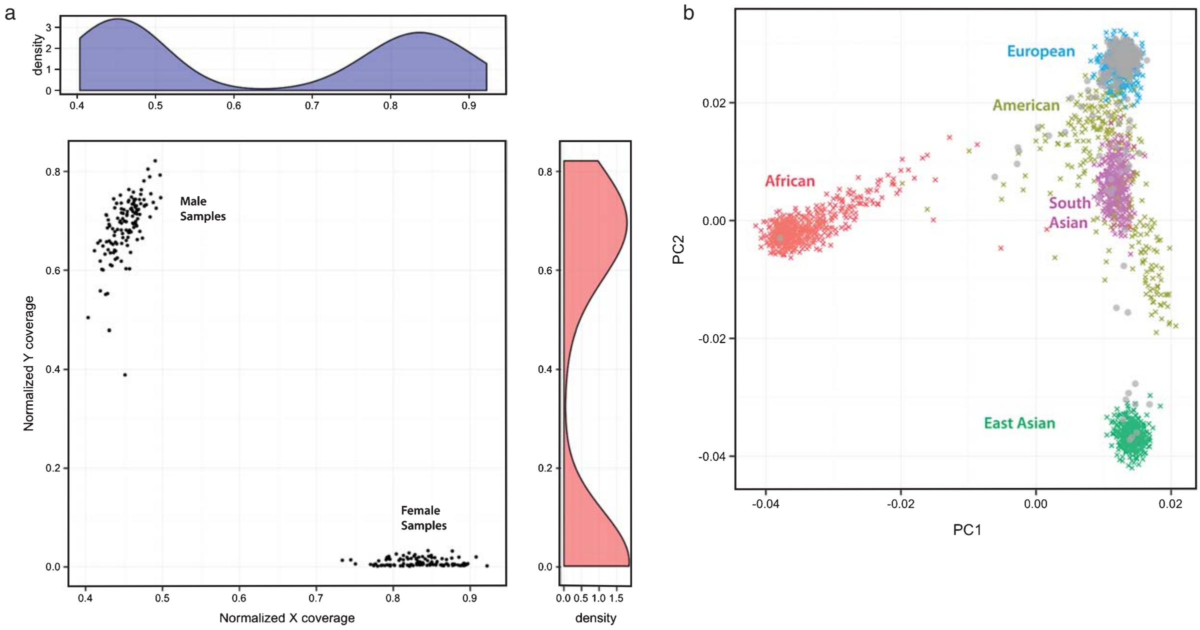

Fig. 2. Sample inference using common variants in exome sequencing data. a) The gender of Genetics of Inherited Muscle Disease samples (phs000655.v1.p1) can be determined from the normalized coverage over common variant in chromosome X and Y compared to chromosome 20. b) The ethicity of samples can be determined by performing principle component analysis with 1000 Genomes samples, where the ethicity of the samples are known. The Genetics of Inherited Muscle Disease samples are labeled in grey circles and the 1000 Genomes samples are labeled as colored crosses corresponding to the various population groups.

almost identical, making it difficult for the short reads from NGS to be mapped confidently. Another difficult region is the gene clusters containing *MYH2*, *MYH7* and including other myosin (*MYH*) genes. The genes in this cluster have been duplicated multiple times and some exons in each gene are still highly similar between genes. Lastly, genes such as *ACTA1*, *TPM2* and *FLNC* have pseudogenes scattered across the genome that can have off-target reads mapped to them. In the case of *FLNC* this has resulted in erroneous mutation reports [34]. Due to these challenges some researchers have preferred targeted gene capture panels or supplementation by traditional Sanger sequencing over poorly covered regions [26, 28]. However, despite these challenges the last two years have shown an increase in mutation reports in *TTN* and *MYH7* (Table 1), genes that have been challenging using traditional approaches.

Active storage costs

A typical high coverage exome fills approximately 8–20 Gb of hard disk space, which is a substantial increase from data usage required for Sanger sequencing of genes. Furthermore, 30x coverage whole genome sequenced samples can require up to 500 Gb of storage. The secure storage and transfer of this data is a huge challenge faced by researchers, for which most labs are inadequately prepared in terms of IT infrastructure. Many researchers also underestimate the substantial costs associated with the storage and processing of raw data from NGS, as well as intermittent re-processing as new algorithms become available. The development of compressed representation of NGS data as gVCFs and CRAM (compressed BAM) remains an active area of research with substantial potential to both reduce storage and processing costs. While whole genome sequencing now approaches the long-fabled $1000 target, lowering the costs of storage, processing and interpretation remains a challenge.

LIMITATIONS OF EXOME SEQUENCING

The overall success of exome sequencing in cohorts of undiagnosed muscle diseases is difficult to estimate due to widespread ascertainment and publication bias; in other rare diseases, success rates hover between 15 and 40% [35]. Discussed in this section are two broad areas of improvement: extracting full use of a patient's exome sequencing data and then possibly supplementing with other genomic technologies.

Study design

The study design for undiagnosed families is a compromise between cost and increasing genetic evidence. Starting with a proband-only approach will identify patients with known muscle disease mutations that were missed during screening and also identify strong gene candidates that are common amongst probands with a similar phenotype. If there are no strong candidate mutations, parents and additional family members can then be sequenced. There are notable limitations to this proband-first approach. Firstly, the timely results for some probands may be required for important medical or family planning decisions and therefore additional family members are usually sequenced. It is usually not possible in the case of recessive inheritance to determine if two mutations in a proband segregate with the disease (i.e. are compound heterozygous), or if candidate dominant mutations have arisen *de novo* in the proband, and there may be too many candidates to confirm using Sanger sequencing. Similarly, unless very fine mapping has been performed, probands with suspected dominant inheritance will often have large numbers of candidate mutations. Also, as discussed in the previous section, having no additional family members limits the ability to detect sample mislabeling and other inconsistencies. Lastly, sequencing additional family members in a different batch may result in batch artefacts and also reduces the sensitivity of tools such as copy number variant (CNV) detection.

Variant annotation

The choice of annotation tool and gene models can result in various discrepancies and has an unappreciated impact on our ability to interpret sequence variants [19]. The majority of mutations reported in recent neuromuscular disease gene discovery papers have been missense or truncating mutations caused by nonsense or frame shift mutations. Interestingly, in the case of *FDX1L* there was a start loss mutation rescued by a nearby methionine but resulting in lower protein levels [36]. Also, *TPNO3* has a stop loss mutation and depending on isoform involves an extension of 15 or 95 amino acids resulting in mislocalization [6]. The majority of publications do not explicitly mention the use of splice prediction for variants outside the essential donor and acceptor splice sites. Although, some capture kits target promoter and untranslated regions (UTRs) our ability to interpret is still limited. The recently characterized homozygous 9 bp deletion in the promoter of *POMGNT1* causing transcription repression is one such mutation that may be missed [37]. Lastly, another limitation is the correct annotation of multiple nucleotide polymorphisms (MNPs), which occurs when variants in the same codon are on the same haplotype and should not be treated as two independent annotations. This may result in missense variants annotated as loss of function [38].

Structural variant detection

In the studies reviewed, structural variants have not been implicated with muscle disease partly due to the challenges in detecting this class of variation in exome sequencing. Within the *DMD* gene, large deletions or duplications account for 74% and 87% of the total mutations in Duchenne muscular dystrophy and Becker's muscular dystrophy, respectively [39]. There currently is no comprehensive survey of structural variants in a large neuromuscular disease cohort. As a comparison, conclusive *de novo* CNVs account for 21% of cases in a severe intellectual disease cohort [40]. These copy number variations can be detected by analyzing significant read coverage deviation using tools such as exomeDepth [41] and XHMM [42] but there is still much room for improvement. The main challenge is distinguishing a read coverage deviation caused by true copy number and those caused by sample preparation, sequencing efficiency and read mapping. Using an appropriate reference panel that has been processed by the same workflow can help reduce false positives. Similarly, variant calling is difficult in short tandem repeat (STR) regions where there can be benefits in using specialized tools such as lobSTR [43]. The challenges with detecting variants in the STR region are that only a proportion of short reads will span the entire region and the PCR amplication used in sample preparation is prone to slippage in these regions. Lastly, retrotransposon insertion in *FKTN*, *POMT1* and *DMD* are known to cause muscle disease [44]. A systematic survey of mobile insertion in the 1000 genomes project has outlined how these events can be robustly detected [45], however in exome sequencing there may not be enough sequencing reads to convincingly support this event.

There are various structural variants that are difficult to detect using next generation sequencing but are known to occur in muscle diseases. Both myotonic dystrophies involve expansions of tandem repeats, in myotonic dystrophy (DM1) there is a expansion of CTG repeats in the 3′ UTR of *DPMK* [46], while in myotonic dystrophy (DM2) there is expansion of CCTG in intron 1 of *ZNF9* [47]. The contraction of D4Z4 repeat units in sub-telomeric region of chromosome 4 results in FSHD [48]. The insertion of SVA retrotransposon in the *FKTN* 3′ UTR causing Fukuyama-type congenital muscular dystrophy (FCMD) was the first human disease reported to be associated by this mechanism [49]. The SVA insertion results in a modified *FKTN* C-terminus affecting protein localization. Due to the common occurrence of DM1/2, FSHD and FCMD (amongst Japanese population), these mutations should be screened using more sensitive assays for patients with similar phenotypes.

Whole genome sequencing

The reducing costs of next generation sequencing will lead to wider use of RNA sequencing and whole genome sequencing. At the moment variants detected from whole genome sequencing are greatly limited by our ability to annotate and interpret non-coding sequence variants and there is also limited high coverage reference genomes to compare results. Apart from structural variants and known functional sites, ironically analysis is typically limited to genic regions of the genome. For instance, we note that an early publication demonstrating the use of genome sequencing for the diagnosis of Charcot-Marie-Tooth Neuropathy could have been successfully replicated using exome sequencing [50].

RNA sequencing

RNA sequencing is an orthogonal approach that addresses some key limitations in interpretation. It will allow us to detect the impact of expression and splicing in RNA isolated from disease muscle and in combination with DNA sequencing data will allow us to associate variants to these events. As a new and emerging technology there are several challenges associated with RNA sequencing. Firstly, informatic workflows are not as well-established as for DNA sequencing technologies, with little current consensus about best-practice workflows. Secondly, the process of obtaining RNA from primary muscle requires an invasive muscle biopsy, limiting its applicability. Fortunately, existing muscle biopies that have been stored properly are usually adequate for RNA sequencing. Thirdly, it is unclear what sequencing coverage is required to adequately detect pathogenic alterations. Finally, the large size of muscle genes combined with the bias of RNA sequencing data generated using standard poly(A) capture protocols towards the 3′ end of transcripts can substantially reduce coverage of the 5′ ends of many known muscle disease genes.

GENETIC EVIDENCE

In all reviewed publications, the minimum evidence presented was Sanger sequencing validation of the candidate mutation and segregation of this mutation amongst affected and unaffected available family members; linkage analysis was also typically performed for dominant diseases. In the case of novel disease gene association, mutations were discovered in multiple unrelated families except for the case of *CCDC78*, *HACD1* and *FDX1L*.

Large, publicly available data-sets of population control sequencing data have been rapidly adopted as a standard part of NGS analysis protocols by most muscle disease researchers, providing much more precise estimates of variant frequency than the traditional 100–200 control chromosomes. The majority of publications reviewed used variants from dbSNP, Exome Variant Server (http://evs.gs.washington.edu/EVS) and the 1000 Genomes Project to filter out variants in exomes. Interestingly, these data sets have empowered the development of frameworks to identify genes that are constrained [51] or intolerant [52] to mutation. These frameworks can be used to assess mutations in large genes such as *TTN* and *NEB* that are more likely to have rare variants by chance. Lastly, a variant that is absent or rare in population controls should not be used as the only evidence to conclude pathogenicity in a known disease gene. As more reliable and relevant high-throughput functional assays are developed, these assays can be employed with "saturation editing" to assess all possible variations of interest for a desired gene [53].

There are several limitations that muscle researchers should bear in mind when using these public resources. The variants within dbSNP are not limited to benign variants and also contain pathogenic mutations. The Exome Variant Server (EVS) hosts variants from the NHLBI GO ESP cohort, which contains individuals with various forms of cardiovascular disease and is limited to European Americans and African Americans. Although, the 1000 Genomes Project represents a more global cohort, the representation of various population groups is very limited in number [33]. Lastly, the public resources have not been processed and analyzed in the same way as the patient exomes, which may contribute to artefactual differences. To address these limitations, some of the studies have used appropriate population controls and also "in house" control exomes.

INFORMATIC SUPPORT

The use of pathogenicity prediction was reported in almost all missense variants in the publications reviewed. The most common prediction tools were SIFT [54], Polyphen2 [55] and Mutation taster [56]. These predictions should not be treated as independent as they are trained on similar data sets with overlapping training parameters. The recently available Combined Annotation-Dependent Depletion (CADD) method has attempted to address some of the limitations of previous tools including scores for non-coding variants and small insertion-deletions [57]. In contrast, splicing prediction was rarely employed; for proposed splice-disrupting mutations direct experimental evidence was presented instead. Lastly, molecular modeling using FoldX [58] and visualization using PyMOL (http://www.pymol.org) was presented in nemaline myopathy studies [12], where similar proteins structures are available for candidate genes.

The use of publicly available RNA expression and protein–protein interaction data was not widely reported by the studies reviewed. The pilot phase of the Genotype and Expression (GTEx) project has RNA sequencing from various human tissues [59], including skeletal and cardiac muscle. The expression data is publicly available through the GTEx portal (http://www.gtexportal.org) and can be used to address muscle tissue expression and whether candidate mutations are in muscle expressed transcripts. There are plans to expand the range of tissues, including multiple muscles with disease relevance. The differential expression patterns between muscle groups may help to explain the known difference in muscle wasting patterns amongst the muscle diseases [60]. There are numerous manually curated protein–protein interaction databases and also a recently published LGMD specific protein interaction network [61]. Using available tools such DAPPLE [62], new candidate genes can be tested for their association with other known disease proteins in the network. The use of these publicly available resources will reduce the need to present results from tissue expression experiments, however the under-representation of developmental isoforms is an obvious limitation of using these data.

EXPERIMENTAL EVIDENCE

All studies have comprehensively characterized a subset of affected patients and, in the case of novel genes, established the expression and protein localization in muscle and other relevant disease tissues. In implicating novel candidate genes, a subset of studies used morpholino knock down in zebra fish. Furthermore, in the case of *KLHL41* [12] an*d DNAJB6* [3] the knock down phenotype was rescued using mRNA from the human gene, ruling out the possibility that the phenotype was due to off target effects of the morpholino. Interestingly, the disease mechanism of FSHD2 is primate specific due to the interaction between *SMCHD1* and the D4Z4 repeat units [9]. Although there is a mouse model, it fails to recapitulate any of the muscle phenotype observed in human patients [63]. Besides zebra fish, no other animal models were used, thus showing the appeal of zebra fish models in obtaining rapid muscle disease phenotype results. Although zebra fish modeling is a path to more rapid publication, researchers must remain prudent that mouse models are more similar to human muscle function and also the complications introduced by duplicate functional gene copies in the zebra fish. In some studies, experiments on patient samples/mutations included a combination of disease specific functional assays (eg. enzymatic activity for mutations in genes that encode enzymes), splicing assays, protein mislocalization, truncation and reduced protein levels.

PUBLIC SHARING AND REPORTING

The correct reporting of disease variants and sharing of exome sequencing data would greatly empower muscle disease research and effectiveness of diagnostic labs around the world. There are two widely used free mutation databases for muscle researchers, the Leiden muscle dystrophy database (http://www.dmd.nl) and ClinVar [64]. In addition, exome and phenotype data can be stored on European Genome-phenome Archive (EGA) or the database of Genotypes and Phenotypes (dbGaP). Both dbGaP (http://www.ncbi.nlm.nih.gov/gap) and EGA (https://www.ebi.ac.uk/ega/home) are public archives for storing genotype and phenotype data in a structured format and also having a mechanism to approve other researchers to access the data. In the reviewed studies, a few mutations discovered have been previously reported [65–69]. Given, the existing reported mutations are known to include benign variants some scrutiny should be applied based on evidence demonstrated in the original report. Unfortunately only a few studies have explicitly stated the reporting of mutations in a public database [8, 70, 71] and only exomes sequenced from UK10K are available through EGA [71–73]. For our own research, muscle disease samples sequenced at the Broad Institute have been made available to the research community through dbGaP *as the Genetics of Inherited Muscle Disease* (Study Accession: phs000655.v1.p1). Two publications have already resulted from this cohort [72, 74]. Since its release in December 2013, it has been requested by 8 research groups and will be used for the development of better analytical techniques and tools in not only the domain of rare muscle diseases but also common diseases and cancer.

There are a few broad areas that researchers in the muscle disease community can improve upon on this front. First, locus-specific databases such as the Leiden muscular dystrophy database have assigned curators for each gene, but publication authors should also take responsibility for reporting and maintaining mutation reports; this effort will likely soon shift towards ClinVar, which we believe is emerging as the likely primary global repository of disease-causing variants. Furthermore, variants need to be carefully labeled with the evidence supporting pathogenicity, as outlined in several recent publications [18, 21].

Researchers must also face up to the perverse incentives against data sharing created by the current academic publication system. Many newly discovered genes languish for months or even years between confident identification and formal publication, while authors accumulate the additional families and experimental support required for a high-impact publications; this delay does a disservice to patients and their families. In addition, even after publication, many reports are locked behind journal subscription paywalls, hindering access by patients and their families. Given the value of this access to patients, and the generous donation of patient time and samples to the research process, we believe the research community has a moral imperative to improve this access, and urge researchers to favor open access publications, and to consider making their pre-formatted manuscripts available prior to publication through preprint servers such as bioRxiv (http://biorxiv.org).

Finally, we emphasize that sharing is greatly empowered by consistent nomenclature. The majority of publications reviewed did not use genomic coordinates and instead have used HGVS nomenclature, which is transcript-specific and frequently ambiguous. The rapid increase of subtle subcategories of phenotypes [75] have called for the use of more consistent description of phenotypes; formal structured vocabularies such as the Human Phenotype Ontology [76] and interface frameworks such as PhenoTips [77] will help to increase consistency. Lastly, there is a publication bias towards diagnosis in exome sequencing, with very few published unsolved cases. As a community we need to establish frameworks to make genetic data from unsolved cases available to collaborators to empower joint analysis with other cohorts. The RD-Connect platform used by the NeurOmics project is an example of such a framework [78].

CONCLUSIONS

New genomic technologies are unquestionably changing the landscape of rare disease genetics. While some challenges remain in terms of accuracy, especially for insertion and deletion variants, raw sequencing data is becoming increasingly accurate and complete; we expect that with increasing sequence accuracy and the advent of longer-read sequencing technologies most regions of the human genome will soon be fully accessible. However, major challenges remain to be resolved before NGS approaches can become a fully mature diagnostic toolkit. The most formidable of these challenges remain in the field of variant interpretation: taking the hundreds of rare, potentially functional variants in any individual's genome and confidently identifying the handful of variants that underlie a specific disease phenotype in that individual. Achieving perfect diagnostic rates for rare diseases will require consensus-building on standards for assessing variant pathogenicity; greatly improved statistical approaches for differentiating disease-causing variants from benign "noise"; and a strong commitment to data sharing, both before and after publication.

REFERENCES

[1] Jungmann, R., Avendaño, M. S., Woehrstein, J. B., Dai, M., Shih, W. M., and Yin, P. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. Nat Methods. 2014; 11:313-318.

[2] Zhao, M., Wang, P., Guan, Y., Cen, Z., Zhao, X., Christner, M., et al. Open-Source Genomic Analysis of Shiga-Toxin–Producing. N Engl J Med. 2011; 365:718-724.

[3] Sarparanta, J., Jonson, P. H., Golzio, C., Sandell, S., Luque, H., Screen, M., et al. Mutations affecting the cytoplasmic functions of the co-chaperone DNAJB6 cause limb-girdle muscular dystrophy. Nat Genet. 2012; 44:450-455, S1-S2.

[4] Couthouis, J., Raphael, A. R., Siskind, C., Findlay, A. R., Buenrostro, J. D., Greenleaf, W. J., et al. Exome sequencing identifies a DNAJB6 mutation in a family with dominantly-inherited limb-girdle muscular dystrophy. Neuromuscul Disord. 2014; 24:431-435.

[5] Harms, M. B., Sommerville, R. B., Allred, P., Bell, S., Ma, D., Cooper, P., et al. Exome sequencing reveals DNAJB6 mutations in dominantly-inherited myopathy. Ann Neurol. 2012; 71:407-416.

[6] Torella, A., Fanin, M., Mutarelli, M., Peterle, E., Del Vecchio Blanco, F., Rispoli, R., et al. Next-generation sequencing identifies transportin 3 as the causative gene for LGMD1F. PLoS One. 2013; 8: e63536.

[7] Vieira, N. M., Naslavsky, M. S., Licinio, L., Kok, F., Schlesinger, D., Vainzof, M., et al. A defect in the RNA-processing protein HNRPDL causes limb-girdle muscular dystrophy 1G (LGMD1G). Hum Mol Genet. 2014:1-8.

[8] Klar, J., Sobol, M., Melberg, A., Mäbert, K., Ameur, A., Johansson, A. C. V., et al. Welander distal myopathy caused by an ancient founder mutation in TIA1 associated with perturbed splicing. Hum Mutat. 2013; 34:572-577.

[9] Lemmers, R. J. L. F., Tawil, R., Petek, L. M., Balog, J., Block, G. J., Santen, G. W. E., et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet. 2012; 44:1370-1374.

[10] Scionti, I., Greco, F., Ricci, G., Govi, M., Arashiro, P., Vercelli, L., et al. Large-scale population analysis challenges the current criteria for the molecular diagnosis of fascioscapulohumeral muscular dystrophy. Am J Hum Genet. 2012; 90:628-635.

[11] Ravenscroft, G., Thompson, E. M., Todd, E. J., Yau, K. S., Kresoje, N., Sivadorai, P., et al. Whole exome sequencing in foetal akinesia expands the genotype-phenotype spectrum of GBE1 glycogen storage disease mutations. Neuromuscul Disord. 2013; 23:165-169.

[12] Gupta, V. A., Ravenscroft, G., Shaheen, R., Todd, E. J., Swanson, L. C., Shiina, M., et al. Identification of KLHL41 Mutations Implicates BTB-Kelch-Mediated Ubiquitination as an Alternate Pathway to Myofibrillar Disruption in Nemaline Myopathy. Am J Hum Genet. 2013; 93:1108-1117.

[13] Wang, Z., Gerstein, M., and Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10:57-63.

[14] Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011; 12:745-755.

[15] Hawkins, R. D., Hon, G. C., and Ren, B. Next-generation genomics: An integrative approach. Nat Rev Genet. 2010; 11:476-486.

[16] Metzker, M. L. Sequencing technologies – the next generation. Nat Rev Genet. 2010; 11:31-46.

[17] Goldstein, D. B., Allen, A., Keebler, J., Margulies, E. H., Petrou, S., Petrovski, S., et al. Sequencing studies in human genetics: Design and interpretation. Nat Rev Genet. 2013; 14:460-470.

[18] MacArthur, D. G., Manolio, T. A., Dimmock, D. P., Rehm, H. L., Shendure, J., Abecasis, G. R., et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014; 508:469-476.

[19] McCarthy, D. J., Humburg, P., Kanapin, A., Rivas, M. A., Gaulton, K., Cazier, J., et al. Choice of transcripts and software has a large effect on variant annotation. Genome Med. 2014; 6:26.

[20] O'Rawe, J., Jiang, T., Sun, G., Wu, Y., Wang, W., Hu, J., et al. Low concordance of multiple variant-calling pipelines: Practical implications for exome and genome sequencing. Genome Med. 2013; 5:28.

[21] Rehm, H. L., Bale, S. J., Bayrak-Toydemir, P., Berg, J. S., Brown, K. K., Deignan, J. L., et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013; 15:733-747.

[22] Vasli, N., and Laporte, J. Impacts of massively parallel sequencing for genetic diagnosis of neuromuscular disorders. Acta Neuropathol. 2013; 125:173-185.

[23] Matsuura, T., Kurosaki, T., Omote, Y., Minami, N., Hayashi, Y. K., Nishino, I., et al. Exome sequencing as a diagnostic tool to identify a causal mutation in genetically highly heterogeneous limb-girdle muscular dystrophy. J Hum Genet. 2013; 58:564-565.

[24] McDonald, K. K., Stajich, J., Blach, C., Ashley-Koch, A. E., and Hauser M. A. Exome analysis of two limb-girdle muscular dystrophy families: Mutations identified and challenges encountered. PLoS One. 2012; 7: e48864.

[25] Böhm, J., Vasli, N., Malfatti, E., Le Gras, S., Feger, C., Jost, B., et al. An integrated diagnosis strategy for congenital myopathies. PLoS One. 2013; 8: e67527.

[26] Dias, C., Sincan, M., Cherukuri, P. F., Rupps, R., Huang, Y., Briemberg, H., et al. An analysis of exome sequencing for diagnostic testing of the genes associated with muscle disease and spastic paraplegia. Hum Mutat. 2012; 33:614-626.

[27] Chen, Z., Wang, J.-L., Tang, B.-S., Sun, Z.-F., Shi, Y.-T., Shen, L., et al. Using next-generation sequencing as a genetic diagnostic tool in rare autosomal recessive neurologic Mendelian disorders. Neurobiol Aging. 2013; 34:2442.e11-7.

[28] Vasli, N., Böhm, J., Le Gras, S., Muller, J., Pizot, C., Jost, B., et al. Next generation sequencing for molecular diagnosis of neuromuscular diseases. Acta Neuropathol. 2012; 124:273-283.

[29] Jun, G., Flickinger, M., Hetrick, K. N., Romm, J. M., Doheny, K. F., Abecasis, G. R., et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet. 2012; 91:839-848.

[30] Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014; 506:185-190.

[31] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. a., R., Bender, D., et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559-575.

[32] Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., and Zondervan, K. T. Data quality control in genetic case-control association studies. Nat Protoc. 2010; 5:1564-1573.

[33] Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56-65.

[34] Odgerel, Z., van der Ven, P. F. M., Fürst, D. O., Goldfarb, L. G. DNA sequencing errors in molecular diagnostics of filamin myopathy. Clin Chem Lab Med. 2010; 48:1409-1414.

[35] Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013; 369:1502-1511.

[36] Spiegel, R., Saada, A., Halvardson, J., Soiferman, D., Shaag, A., Edvardson, S., et al. Deleterious mutation in FDX1L gene is associated with a novel mitochondrial muscle myopathy. Eur J Hum Genet. 2014; 22:902-906.

[37] Raducu, M., Baets, J., Fano, O., Van Coster, R., and Cruces J. Promoter alteration causes transcriptional repression of the POMGNT1 gene in limb-girdle muscular dystrophy type 2O. Eur J Hum Genet. 2012; 20:945-52.

[38] MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science 2012; 335:823-828.

[39] Tuffery-Giraud, S., Béroud, C., Leturcq, F., Yaou R Ben, Hamroun, D., Michel-Calemard, L., et al. Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: A model of nationwide knowledgebase. Hum Mutat. 2009; 30:934-945.

[40] Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W. M., Willemsen, M. H., et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014; 511:344-347.

[41] Plagnol, V., Curtis, J., Epstein, M., Mok, K. Y., Stebbings, E., Grigoriadou, S., et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics. 2012; 28:2747-2754.

[42] Fromer, M., Moran, J. L., Chambert, K., Banks, E., Bergen, S. E., Ruderfer, D. M., et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am J Hum Genet. 2012; 91: 597-607.

[43] Gymrek, M., Golan, D., Rosset, S., and Erlich Y. lobSTR: A short tandem repeat profiler for personal genomes. Genome Res. 2012; 22:1154-1162.

[44] Hancks, D. C., and Kazazian, H. H. Active human retrotransposons: Variation and disease. Curr Opin Genet Dev. 2012; 22:191-203.

[45] Stewart, C., Kural, D., Strömberg, M. P., Walker J. A., Konkel, M. K., Stütz, A. M., et al. A comprehensive map of mobile element insertion polymorphisms in humans. PLoS Genet. 2011; 7: e1002236.

[46] Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, a. J., Church, D., and Aburatani, H., et al. Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell. 1992; 69:385.

[47] Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science. 2001; 293:864-867.

[48] Van der Maarel, S. M., and Frants, R. R. The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2005; 76:375-386.

[49] Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Iida, E., Nomura, Y., et al. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. Nature. 1998; 394:388-392.

[50] Lupski, J. R., Reid, J. G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D. C.Y, Nazareth, L., et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med. 2010; 362:1181-1191.

[51] Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., et al. A framework for the interpretation of de novo mutation in human disease. Nat Genet. 2014; 46:944-950.

[52] Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S., and Goldstein, D. B. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013; 9: e1003709.

[53] Findlay, G. M., Boyle E. A., Hause, R. J., Klein, J. C., and Shendure J. Saturation editing of genomic regions by multiplex homology-directed repair. Nature. 2014; 513:120-123.

[54] Kumar, P., Henikoff, S., and Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009; 4: 1073-1081.

[55] Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248-249.

[56] Schwarz, J. M., Rödelsperger, C., Schuelke, M., and Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010; 7:575-576.

[57] Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014; 46:310-315.

[58] Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano L. The FoldX web server: An online force field. Nucleic Acids Res. 2005; 33: W382-W388.

[59] Consortium, T. G. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013; 45:580-585.

[60] Mercuri, E., and Muntoni F. Muscular dystrophies. Lancet. 2013; 381:845-860.

[61] Blandin, G., Marchand, S., Charton, K., Danièle, N., Gicquel, E., Boucheteil, J.-B., et al. A human skeletal muscle interactome centered on proteins involved in muscular dystrophies: LGMD interactome. Skelet Muscle. 2013; 3:3.

[62] Rossin, E. J., Lage, K., Raychaudhuri, S., Xavier, R. J., Tatar, D., Benita, Y., et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 2011; 7: e1001273.

[63] Krom, Y. D., Thijssen, P. E., Young, J. M., den Hamer, B., Balog, J., Yao, Z., et al. Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD. PLoS Genet. 2013; 9: e1003415.

[64] Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., et al. ClinVar: Public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014; 42: D980-D985.

[65] Hedberg, C., Niceta, M., Fattori, F., Lindvall, B., Ciolfi, A., D'Amico, A., et al. Childhood onset tubular aggregate myopathy associated with *de novo* STIM1 mutations. J Neurol. 2014; 261:870-876.

[66] Komlósi, K., Hadzsiev, K., Garbes, L., Martínez Carrera, L. A., Pál, E., Sigurethsson, J. H., et al. Exome sequencing identifies Laing distal myopathy MYH7 mutation in a Roma family previously diagnosed with distal neuronopathy. Neuromuscul Disord. 2014; 24:156-161.

[67] Leidenroth, A., Sorte, H. S., Gilfillan, G., Ehrlich, M., Lyle, R., and Hewitt, J. E. Diagnosis by sequencing: Correction of misdiagnosis from FSHD2 to LGMD2A by whole-exome analysis. Eur J Hum Genet. 2012; 20:999-1003.

[68] Palmio, J., Evilä A, Chapon, F., Tasca, G., Xiang, F., Br*rm*å*dvik, B., et al. Hereditary myopathy with early respiratory failure: Occurrence in various populations. J Neurol Neurosurg Psychiatry. 2014; 85:345-353.

[69] Raphael, A. R., Couthouis, J., Sakamuri, S., Siskind, C., Vogel, H., Day, J. W., et al. Congenital muscular dystrophy

and generalized epilepsy caused by GMPPB mutations. Brain Res. 2014:1-6.

[70] Ravenscroft, G., Miyatake, S., Lehtokari, V.-L., Todd, E. J., Vornanen, P., Yau, K. S., et al. Mutations in KLHL40 are a frequent cause of severe autosomal-recessive nemaline myopathy. Am J Hum Genet. 2013; 93:6-18.

[71] Logan, C. V., Szabadkai, G., Sharpe, J. A., Parry, D. A., Torelli, S., Childs, A.-M., et al. Loss-of-function mutations in MICU1 cause a brain and muscle disorder linked to primary alterations in mitochondrial calcium signaling. Nat Genet. 2014; 46:188-193.

[72] Carss, K. J., Stevens, E., Foley, A. R., Cirak, S., Riemersma, M., Torelli, S., et al. Mutations in GDP-mannose pyrophosphorylase B cause congenital and limb-girdle muscular dystrophies associated with hypoglycosylation of α-dystroglycan. Am J Hum Genet. 2013; 93:29-41.

[73] Stevens, E., Carss, K. J., Cirak, S., Foley a, R., Torelli, S., Willer, T., et al. Mutations in B3GALNT2 cause congenital muscular dystrophy and hypoglycosylation of α-dystroglycan. Am J Hum Genet. 2013; 92:354-365.

[74] Oates, E. C., Rossor, A. M., Hafezparast, M., Gonzalez, M., Speziani, F., MacArthur, D. G., et al. Mutations in BICD2 cause dominant congenital spinal muscular atrophy and hereditary spastic paraplegia. Am J Hum Genet. 2013; 92:965-973.

[75] Kaplan, J.-C., and Hamroun, D. The 2014 version of the gene table of monogenic neuromuscular disorders (nuclear genome). Neuromuscul Disord. 2013; 23:1081-1111.

[76] Köhler, S., Doelken, S. C., Mungall, C. J., Bauer, S., Firth, H. V., Bailleul-Forestier, I., et al. The Human Phenotype Ontology project: Linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014; 42:D966-D974.

[77] Girdea, M., Dumitriu, S., Fiume, M., Bowdin, S., Boycott, K. M., Chénier, S., et al. PhenoTips: Patient phenotyping software for clinical and research use. Hum Mutat. 2013; 34:1057-1065.

[78] Thompson, R., Johnston, L., Taruscio, D., Monaco, L., Béroud, C., Gut, I. G., et al. RD-Connect: An Integrated Platform Connecting Databases, Registries, Biobanks and Clinical Bioinformatics for Rare Disease Research. J Gen Intern Med. 2014; 29(Suppl 3):780-787.

[79] Belaya, K., Finlayson, S., Slater, C. R., Cossins, J., Liu, W. W., Maxwell, S., et al. Mutations in DPAGT1 cause a limb-girdle congenital myasthenic syndrome with tubular aggregates. Am J Hum Genet. 2012; 91:193-201.

[80] Hicks, D., Farsani, G. T., Laval, S., Collins, J., Sarkozy, A., Martoni, E., et al. Mutations in the collagen XII gene define a new form of extracellular matrix-related myopathy. Hum Mol Genet. 2014; 23:2353-2363.

[81] Majczenko, K., Davidson, A. E., Camelo-Piragua, S., Agrawal, P. B., Manfready R a, Li, X., et al. Dominant mutation of CCDC78 in a unique congenital myopathy with prominent internal nuclei and atypical cores. Am J Hum Genet. 2012; 91:365-371.

[82] Muhammad, E., Reish, O., Ohno, Y., Scheetz, T., Deluca, A., Searby, C., et al. Congenital myopathy is caused by mutation of HACD1. Hum Mol Genet. 2013; 22:5229-5236.

[83] Weterman, M. A. J., Barth, P. G., van Spaendonck-Zwarts, K. Y., Aronica, E., Poll-The, B.-T., Brouwer, O. F., et al. Recessive MYL2 mutations cause infantile type I muscle fibre disease and cardiomyopathy. Brain. 2013; 136:282-293.

[84] Böhm, J., Leshinsky-Silver, E., Vassilopoulos, S., Le Gras, S., Lerman-Sagie, T., Ginzberg, M., et al. Samaritan myopathy, an ultimately benign congenital myopathy is caused by a RYR1 mutation. Acta Neuropathol. 2012; 124: 575-581.

[85] Ceyhan-Birsoy, O., Agrawal, P. B., Hidalgo, C., Schmitz-Abe, K., DeChene, E. T., Swanson, L. C., et al. Recessive truncating titin gene, TTN, mutations presenting as centronuclear myopathy. Neurology. 2013; 81: 1205-1214.

[86] Chaouch, A., Porcelli, V., and Cox D. Mutations in the Mitochondrial Citrate Carrier SLC25A1 are Associated with Impaired Neuromuscular Transmission. J Neuromuscul Dis. 2014; 1:75-90.

[87] Chauveau, C., Bonnemann, C. G., Julien, C., Kho, A. L., Marks, H., Talim, B., et al. Recessive TTN truncating mutations define novel forms of core myopathy with heart disease. Hum Mol Genet. 2014; 23:980-991.

[88] Cortese, A., Tucci, A., Piccolo, G., Galimberti, C. a., Fratta, P., Marchioni, E., et al. Novel CLN3 mutation causing autophagic vacuolar myopathy. Neurology. 2014:2072-2076.

[89] Davidson, A. E., Siddiqui, F. M., Lopez, M. a., Lunt, P., Carlson, H. a., Moore, B. E., et al. Novel deletion of lysine 7 expands the clinical, histopathological and genetic spectrum of TPM2-related myopathies. Brain. 2013; 136: 508-521.

[90] Esposito, T., Sampaolo, S., Limongelli, G., Varone, A., Formicola, D., Diodato, D., et al. Digenic mutational inheritance of the integrin alpha 7 and the myosin heavy chain 7B genes causes congenital myopathy with left ventricular non-compact cardiomyopathy. Orphanet J Rare Dis. 2013; 8:91.

[91] Izumi, R., Niihori, T., Aoki, Y., Suzuki, N., Kato, M., Warita, H., et al. Exome sequencing identifies a novel TTN mutation in a family with hereditary myopathy with early respiratory failure. J Hum Genet. 2013; 58:259-266.

[92] Jimenez-Escrig, A., Gobernado, I., Garcia-Villanueva, M., and Sanchez-Herranz A. Autosomal recessive Emery-Dreifuss muscular dystrophy caused by a novel mutation (R225Q) in the lamin A/C gene identified by exome sequencing. Muscle Nerve. 2012; 45:605-610.

[93] Malfatti, E., Schaeffer, U., Chapon, F., Yang, Y., Eymard, B., Xu, R., et al. Combined cap disease and nemaline myopathy in the same patient caused by an autosomal dominant mutation in the TPM3 gene. Neuromuscul Disord. 2013; 23:992-997.

[94] Maselli, R., Arredondo, J., Nguyen, J., Lara, M., Ng, F., Ngo, M., et al. Exome sequencing detection of two untranslated G FPT1 mutations in a family with limb-girdle myasthenia. Clin Genet. 2013:166-171.

[95] Mitsuhashi, S., Boyden, S. E., Estrella, E. a., Jones, T. I., Rahimov, F., Yu, T. W., et al. Exome sequencing identifies a novel SMCHD1 mutation in facioscapulohumeral muscular dystrophy 2. Neuromuscul Disord. 2013; 23:975-980.

[96] Ohlsson, M., Hedberg, C., Brådvik, B., Lindberg, C., Tajsharghi, H., Danielsson, O., et al. Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin. Brain. 2012; 135:1682-1694.

[97] Park, J.-M., Kim, Y. J., Yoo, J. H., Hong, Y., Bin, Park, J. H., Koo, H., et al. A novel MYH7 mutation with prominent paraspinal and proximal muscle involvement. Neuromuscul Disord. 2013; 23:580-586.

[98] Pfeffer, G., Elliott, H. R., Griffin, H., Barresi, R., Miller, J., Marsh, J., et al. Titin mutation segregates with hereditary myopathy with early respiratory failure. Brain. 2012; 135:1695-1713.

[99] Romero, N. B., Xie, T., Malfatti, E., Schaeffer, U., Böhm, J., Wu, B., et al. Autosomal dominant eccentric core disease caused by a heterozygous mutation in the MYH7 gene. J Neurol Neurosurg Psychiatry. 2014. doi: 10.1136/jnnp-2013-306754

[100] Schessl, J., Bach, E., Rost, S., Feldkirchner, S., Kubny, C., Müller, S., et al. Novel recessive myotilin mutation causes severe myofibrillar myopathy. Neurogenetics. 2014; 15:151-156.

[101] Von Renesse, A., Petkova M, V., Lützkendorf, S., Heinemeyer, J., Gill, E., Hübner, C., et al. POMK mutation in a family with congenital muscular dystrophy with merosin deficiency, hypomyelination, mild hearing deficit and intellectual disability. J Med Genet. 2014; 51:275-282.