A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield

Arunkanth Ankala, PhD,[1] Cristina Da Silva, MS, [1] Francesca Gualandi, PhD,[2] Alessandra Ferlini, PhD,[2]

Lora J H Bean, PhD,[1] Christin Collins, PhD,[1] Alice K Tanner, PhD,[1] and Madhuri R Hegde, PhD,[1,*]

[1]Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, USA

[2]Department of Medical Sciences, University of Ferrara, Italy

[*]Corresponding author

Reprint Requests:
Madhuri R Hegde, PhD, FACMG
Department of Human Genetics
Emory University School of Medicine
615 Michael Street
Atlanta, GA 30322, USA
E-mail: mhegde@genetics.emory.edu
Fax: 404-727-3949
Phone: 404-727-3863

Running head: Comprehensive testing for NMDs

Number of characters in Title: 89

Number of characters in Running head: 27

Number of words in the abstract: 247

Number of words in the Text: 3,590

Number of figures: 2

Number of Color figures: 0

Number of tables: 5

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/ana.24303

Abstract

Objective

Neuromuscular diseases (NMDs) are a group of over 200 highly genetically as well as clinically heterogeneous inherited genetic disorders that affect the peripheral nervous and muscular systems, resulting in gross motor disability. The clinical and genetic heterogeneities of NMDs make disease diagnosis complicated and expensive, often involving multiple tests.

Methods

To expedite the molecular diagnosis of NMDs, we designed and validated several next-generation sequencing (NGS)-based comprehensive gene panel tests that include complementary deletion and duplication testing through comparative genomic hybridization arrays (aCGH). Our validation established the targeted gene panel test to have 100% sensitivity and specificity for single nucleotide variant detection. To compare the clinical diagnostic yields of single gene (NMD associated) tests with the various NMD NGS panel tests, we analyzed data from all clinical tests performed at the Emory Genetics Laboratory (EGL) from October 2009 through May 2014. We further compared the clinical utility of the targeted NGS panel test with that of exome sequencing (ES).

Results

We find that NMD comprehensive panel testing has a 3-fold greater diagnostic yield (46%) than single gene testing (15-19%). Sanger fill-in of low-coverage exons, copy number variation (CNV) analysis, and thorough in-house validation of the assay all complement panel testing and allow the detection of all types of causative pathogenic variants, some of which (about 18%) may be missed by ES.

Interpretation

Our results strongly indicate that for molecular diagnosis of heterogeneous disorders like NMD, targeted panel testing has the highest clinical yield and should therefore be the preferred first-tier approach.

Introduction

Neuromuscular diseases (NMDs) refer collectively to the many disorders that affect the peripheral nervous system, either by impairing the proper development or functioning of muscles or by damaging the associated nerves or neuromuscular junctions. Muscular dystrophies which form the majority of inherited NMDs share clinical, genetic, and pathological characteristics. Major clinical characteristics of the disease group include muscle degeneration and wasting, progressive muscle weakness, hypotonia, and although at very variable levels, elevated serum creatine kinase levels.[1] Very often cardiac involvement might also be present, accounting for higher morbidity and mortality. There are over 80 different genetically defined types of muscular dystrophies categorized into different subgroups based on the age of onset, the specific muscles involved, and common characteristic clinical features.[2,3] Congenital muscular dystrophies (CMDs) and limb-girdle muscular dystrophies (LGMDs) are the two major subgroups, the genetic heterogeneity of which has been expanding rapidly in recent years, with more and more genes being implicated.[4] Lack of pathognomonic signs or specific biochemical markers and the presence of high phenotypic overlap with other forms of NMDs often make clinical diagnosis difficult and molecular confirmation expensive. Molecular confirmation is critical not only to establish diagnosis, but also to allow participation in clinical trials of therapeutic treatments that are designed for a specific set of variants or variant types, like those for DMD.[5,6]

Oftentimes, to narrow down the genes to be tested for molecular confirmation, an extensive diagnostic work-up involving protein studies on muscle biopsy is necessary. Abnormal protein staining by immunohistochemical and immunoblotting may be highly specific, prompting single gene testing for certain subtypes, such as merosin-deficient CMD and dysferlin-deficient LGMD.[7,8] However, individuals with most other subtypes undergo a battery of invasive and expensive tests. Consequently, clinicians and patients often discontinue further testing without ever obtaining a diagnosis. Thanks to advances in technology, several new and comprehensive tests are available to expedite molecular diagnosis of the disease. These include NGS-based panel testing for sequence analysis of all disease-associated genes in a single test, comparative genomic hybridization arrays (aCGH) for analysis of deletions and duplications

within genes of interest in a single test, exome sequencing (ES), and genome sequencing (GS). However, it is critical that clinicians and geneticists understand the advantages and technical limitations of each of these tests, their specificities and sensitivities, and finally their diagnostic yields.

Emory Genetics Laboratory (EGL) began offering molecular testing for muscular dystrophies in 2009 and has since continuously developed and introduced several new tests, including single gene sequencing, comparative genomic hybridization arrays (aCGH) for deletion/duplication analysis, subtype-specific targeted gene panels, and a more comprehensive test to cover the entire coding region and all types of variants in the targeted gene panel for NMDs. Here we present our experience with molecular testing for NMD diagnosis and report the diagnostic yields of each testing methodology. Aspects of test design, validation, and performance of the targeted NGS panel compared with ES are also discussed.

Methods

Patients and diagnostic testing

All samples referred for either single gene or panel validation and clinical testing for CMDs, LGMDs, and NMDs, from the launch of the tests in 2009 through May 2014, are included in this study. Peripheral blood samples or isolated DNA were received for each patient referred for molecular testing. Two independent DNA isolations were performed on each blood sample using the commercially available Puregene Blood kit from Qiagen. All molecular diagnostic testing was performed at EGL, which is a Clinical Laboratory Improvement Amendments (CLIA)- and College of American Pathologists (CAP)-accredited laboratory.

Single gene testing

Single gene testing was available for all genes included in the various panels and was ordered by referring clinicians based on clinical diagnosis. All single gene testing was performed using conventional Sanger

sequencing. To avoid allelic dropout during PCR, multiple alternative primer sets were designed in regions with known single nucleotide polymorphisms (SNPs). An automated primer design script, developed and validated in house, was used to design primers.[9]

Targeted gene panel testing

We previously reported the complete details of panel design, target enrichment, and validation of the CMD targeted gene panel test.[10,11] The LGMD panel included 11 genes (Table 1), and the sequence analysis was performed by PCR and Sanger sequencing, rather than NGS. Panel design and clinical validation of the newly developed comprehensive NMD gene panel test targeting 41 genes (Table 1) are described in detail here. Two different target enrichment technologies namely micro-droplet based PCR method (RainDance Technologies, MA) and biotinylated-cRNA probe based hybridization method (Agilent Technologies, CA) were used for the NGS assay in this study. A comparison of the performances of the two enrichment methods has been detailed elsewhere.[11] NMD comprehensive gene panel included both sequence analysis by NGS and deletion/duplication analysis by array comparative genomic hybridization (aCGH) for all 41 genes. Additionally, all exons that lacked sufficient read and coverage depth (20X) were amplified using exon-specific primers by conventional PCR, and then Sanger-sequenced. An exon is referred to as being a "low-coverage exon" if at least a single nucleotide in the exon has a coverage below 20X, a cut-off independently determined at EGL prior to ACMG guidelines and during validation of the NGS platform.[12,13] For a list of low-coverage exons in the CMD panel, refer to Valencia et al,[10] and for the NMD panel, see supplementary data (Table S1).

Design and validation of comprehensive NMD Targeted gene panel

In view of the rapidly expanding genotypic and phenotypic heterogeneity of NMDs (Figure 1) we developed a comprehensive NMD panel to include all genes previously offered at EGL through individual CMD and LGMD panels and those associated with additional phenotypes and subtypes of NMDs (Table 1). A primer library was designed to amplify and enrich all targeted exons, as previously described.[10,11]

The analytical sensitivity and specificity of the NMD targeted gene panel assay were evaluated using known and blinded variant-positive samples. Known variant-positive samples (NMD-E01 through NMD-E05) included those that were previously sequenced by the Sanger method at EGL and found to have pathogenic variants in one of the targeted NMD genes. Alternatively, the blinded samples (NMD-A01 through NMD-A12), which potentially had pathogenic variants as well as SNPs in different genes, were provided by the Ferrara national reference diagnostic laboratory (Italy), without revealing the variants. Performance of the targeted gene panel assay was validated for all major steps, including target enrichment, sequencing, and bioinformatics analysis. After identifying the variants (benign and pathogenic – 53 total variants) by the developed assay and in-house bioinformatics pipeline, the results were compared to those previously Sanger sequenced at EGL or Ferrara national reference diagnostic laboratory (Table 2 and Table 3). While most of the detected variants were concordant between the two data sets, there was some discordance. However, as mentioned in the table against each discordant variant, explanations included either the difference in the reference sequence used between the two laboratories, the way the variant was called by the in-house pipeline (a single indel variant is referred to as two independent variants by bioinformatics pipeline), or insufficient (<20X) coverage for the specific nucleotide in targeted gene panel assay. Overall, only two variants stand out as false negatives (Table 2). However, as previously mentioned since the exons in which these two variants occur lacked sufficient coverage, these would be Sanger sequenced as part of the panel. Taken together, the targeted gene panel assay along with complementary Sanger fill-in of low-coverage exons detected all variants previously detected by Sanger sequencing in Ferrara's laboratory, thus establishing the analytical sensitivity and specificity of the assay for SNV detection to both be 100%.

NMD clinical testing

A total of 204 individual single gene tests were ordered for CMD genes altogether. These included 134 sequencing tests and 70 aCGH tests. A significant number of patients were referred for multiple gene testing, either in parallel or sequentially as a reflex test. For LGMDs, a total of 343 single gene tests were

ordered, which included 250 sequencing tests and 93 aCGH tests. The detection rate of aCGH tests for single gene testing was very low with 5 and 8 positive findings for CMDs and LGMDs respectively. During this period the overall panel tests ordered included 88 CMD, 96 LGMD, and 81 NMD tests.

Exome sequencing

ES was performed on genomic DNA using the NimbleGen (City. State) V3.0 targeted sequence capture method to enrich for the exome. These targeted regions are then sequenced using the Illumina (San Diego, CA) HiSeq 2000 sequencing system with 100 basepair (bp) paired-end reads at an average coverage of 100X in the target region. The targeted regions included the exon and 10 bp of flanking intronic sequence. In general, ES assays performed at EGL have an overall coverage of 92.9%, with as high as 94.8% coverage in the coding region. For comparative analysis of the performance and clinical utility of the targeted gene panel assay and ES assay, data from 20 routine clinical ES samples performed at EGL were selected randomly.

Bioinformatics analysis and variant classification

Bioinformatic analysis was performed using NextGENe™ software from SoftGenetics (State College, PA). The NextGENe output was customized to mine variants from EGL's internal variant database EmBase,[14] other public databases, and variant prediction tools, such as SIFT[15] and PolyPhen[16]. All variants detected were classified using population frequency data available from the Exome Variant Server (http://evs.gs.washington.edu/EVS/) and previous reports of disease association and pathogenicity available through the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/ac/index.php), NCBI PubMed (http://www.ncbi.nlm.nih.gov/pubmed/), and Google (https://www.google.com/). Based on available evidence, variants were classified as benign, pathogenic, or variants of unknown/uncertain significance (VOUS). A detailed overview of the bioinformatics pipeline and variant annotation protocol is described elsewhere.[12] All variants were curated and reviewed by board-certified laboratory directors

and maintained as an in-house variant database; they were made publicly accessible via EGL's online tool, EmVClass.[14]

Results

Versatility of variant detection: test validation and performance

The ability of the NMD panel assay to efficiently detect all kinds of pathogenic variants, including single nucleotide variants (SNVs), deletions, duplications, and small insertions/deletions, was evaluated using known samples that were previously Sanger sequenced and found to have variants (Table 3). These variant-positive samples represented all the different types of variants NGS based targeted gene panel assay is expected to detect. All confirmed variants had coverage of at least 20X and mutant allele percentages consistent with heterozygous and homozygous calls. Intragenic deletions presented in Table 3 were predicted based on lack of coverage for the specific exons and confirmed by complementary aCGH analysis. Duplications, though not readily detected by current bioinformatics algorithms, were detected by complementary aCGH.

Coverage

For single gene testing, Sanger-sequencing in both the forward and reverse directions is performed following amplification of the regions of interest, giving a perceived 2X coverage. In contrast, with NGS, each interrogated nucleotide is sequenced anywhere from several hundred to a thousand times (referred to as the read depth or coverage) and is prone to an extremely high rate of errors and false positives.[17-19] Based on our comparison of data from NMD targeted gene panel assays on 20 samples referred for clinical testing, the average coverage for the LGMD genes (Table 1) in the comprehensive NMD targeted gene panel assay is about 296X (range 125X to 658X), and the average coverage for the CMD genes is about 182X (range 77X to 286X). Such high coverage in the NMD panel assay gives us greater confidence that a detected variant is a true positive call and not a false call. The average coverage for the same set of genes in ES is much lower, about 63X (range 40X to 94X) for LGMD and about 54X (range

34X to 61X) for CMD genes. Regardless, several exons of multiple genes lack the minimal 20X recommended coverage in both the targeted NMD panel and ES. For NMD targeted gene panel the low-coverage exons totaled 68 which involve 22 genes (Supplementary table S1). However, with Sanger fill-in of low-coverage exons, targeted panels provide 100% coverage for all targeted exons to the nucleotide level. Low-coverage exons in ES accounted for about 5.4% (18/336) of CMD genes and 11% (16/140) of LGMD genes. Without Sanger fill-in, variants in low-coverage regions are missed in ES, contributing to a high false-negative rate.

Panel approach has higher diagnostic yield compared to single gene testing

Considering the higher clinical yields and diagnostic rates of the individual smaller panels, CMD (36%) and LGMD (26%), when compared to that of single gene testing at 15-19%, we predicted a comprehensive NMD panel including all interacting and co-existing NMD genes, will have even greater clinical yield and diagnostic utility. Data shows that NMD diagnosis by comprehensive NGS panel has a three-fold greater diagnostic yield than single gene testing (46% by NMD panel versus 15% by CMD and 19% by LGMD single gene testing) (Table 4). Multiple sequential negative single gene testing (requested as a result of confounding phenotypic overlap) was observed to contribute to the low diagnostic yields of single gene testing over panel testing.

Panel versus ES: comparison of diagnostic yields and false-negative rates

In analyzing the exomes for overall coverage across the entire targeted exome (all genes inclusive), we saw that on average 10% of an entire exome lacks 20X coverage. To further explore the effect of low-coverage regions on diagnostic yield in ES, we used our standard bioinformatics pipeline to determine the coverage (and therefore detectability) at the site of pathogenic variants (137 variants total) in multiple genes previously detected by Sanger sequencing in a set of 20 samples each of the NMD targeted panel and ES. This set of analyzed variants included all pathogenic variants ever detected and reported by EGL

by Sanger testing ordered for LGMD and CMD genes (23 genes listed in Table 1). Read depth and coverage were compared both at the exon and nucleotide levels. While 11 variants (7 in the CMD panel and 4 in the LGMD panel) had low coverage in panel testing, 24 variants (14 CMD-associated genes and 10 LGMD-associated genes) had low coverage among ES sample data (Table 5). This indicates that at least 11-18% of pathogenic variants in NMD-related genes included in this study would be missed by ES since Sanger fill-in is not performed. Of note, all undetectable variants within low-coverage regions are in exons listed in the NMD panel proactive list (*SEPN1* exon 1, *COL6A1* exons 9 and 10, *DYSF* exons 2, 17, and 36) and analyzed by Sanger sequencing irrespective of their coverage or detection by NGS. A comparison of complete gene coverage for the targeted panel and ES is shown in Figure 2.

Discussion

Here we report the design, validation, and clinical implementation of a comprehensive targeted NGS panel for NMD diagnosis. Using blinded and known samples, we established the versatility of our assay for variant detection and identified regions of low coverage that needed to be complemented with Sanger sequencing. We show that our targeted NMD targeted gene panel, which includes complementary Sanger sequencing of low-coverage exons, successfully detects all types of sequencing variants with sufficient coverage (≥20X), with no detected false negatives. Comparing data from targeted gene panel tests performed at EGL with data from single gene Sanger testing, we find that for the molecular diagnosis of NMDs, the targeted NGS panel has higher clinical yield, about three times (46%) the 15-19% yield of single gene testing. Moreover, the performance and coverage of NMD genes is significantly better with the targeted panel approach compared to ES. With complementary and confirmatory Sanger and aCGH assays, EGL's NMD targeted gen panel test is therefore a comprehensive singleton clinical test with versatility in variant detection, high analytical sensitivity and specificity (100%) and maximum clinical yield. Taken together, NGS based targeted gene panel testing is the preferred first-tier approach for NMD diagnosis.

Though targeted panel testing may be recommended for most other diseases that are equally or similarly heterogeneous, we recently reported that for congenital disorders of glycosylation (CDGs), single gene testing had higher clinical yield than targeted panel testing (21.2% versus 5.8%).[13] The higher diagnostic rate of single gene testing for CDGs is attributable to the patients referred for single gene testing having either a phenotype consistent with a specific gene or biochemical testing that implicated a specific gene or part of the pathway as being causative. While biochemical testing and clinical phenotype can be very specific for certain CDGs, this is often not so for NMDs. For a detailed overview of the different clinical indications of when and for what conditions a single gene testing may be preferred over a targeted panel testing or exome sequencing, refer to the recent review article by Xue et al.[20]

Further challenges in NMD disease diagnosis include the large number of causative genes (genetic heterogeneity), the association of multiple genes with similar phenotype or a single gene with multiple phenotypes (phenotypic heterogeneity), occurrence of various number and types of pathogenic variants along the length of each gene (allelic heterogeneity), and finally the significantly larger sizes of most NMD genes: *DMD, DYSF, TTN, LMNA, RYR1*, and many more (reviewed by Laing[21]). In a recent report comparing the various sequencing approaches for analyzing *TTN*, the gene with the most coding exons (363), the authors elaborate on the challenges of molecular diagnosis for *TTN* associated cardiomyopathies.[22] The authors also allude to the fact that the cost and challenges involved in analyzing this titanic gene by Sanger sequencing refrains most laboratories from offering single gene testing. Consequently, the mutation spectrum and disease prevalence of *TTN*-associated LGMD and cardiomyopathies remain undetermined. With the inclusion of this and other similar less-characterized genes in the cost-effective targeted NGS gene panel assay, these disease subtypes and associated genes can hence be better defined, at the same time providing molecular diagnoses to those with defects in these genes. On the other hand is *DMD*, where intragenic deletions and duplications (CNVs), not comprehensible by Sanger sequencing, are more frequent (65-70%). As current clinical NGS assays are not yet reliable for CNV detection, targeted panel tests are complemented with aCGH analysis for comprehensive pathogenic-variant detection. A variety of bioinformatics tools, each based on one or more

of several sequence specifics, such as relative read depth, split-read analysis, and paired-end mapping, are being validated for CNV detection from NGS data. [23,24] Going forward, complementary aCGH analysis for CNV detection may no longer be needed for comprehensive NGS targeted panel assays, adding to the current advantages of panel testing over single gene testing and further reducing the test price.

In validating targeted gene panel assays, we determined their limitations in terms of coverage for each single exon of genes included. The one major limitation is the lack of coverage for a significant number of exons involving multiple genes, regardless of the target capture or enrichment technology used.[11] Especially for the comprehensive NMD panel, there was insufficient coverage for, on average, 60-70 exons involving multiple genes (Supplementary table 1). This is true for most of the clinically available NGS based targeted panel tests and is in fact a limitation of the available sequencing chemistries.[13,22,25] These low-coverage exons were more or less the same exons for most samples and included GC-rich first exons of various genes, exons with AT-rich or other sequence-specific complexities, and exons or genes that have highly homologous pseudoexons or pseudogenes. On the other hand, at least 180 exons of these NMD genes had low coverage by ES. Therefore the validation process involved in a targeted panel design is crucial in identifying its limitations and complementing with ancillary assays to establish molecular diagnosis or to give a true negative result and help rule out genes.

As Rehm[26] has discussed, the size of a gene or its relative contribution to a disease is no longer a factor limiting its inclusion in a clinical test, rather the discovery of its association with the disease in question is. Therefore, as genes are identified and established with disease specificity, these can be added to existing targeted panels, thereby expanding the particular panel and increasing its clinical diagnostic yield, as well. However, as presented here and elsewhere,[25] every single targeted panel should be thoroughly validated before its implementation for clinical diagnostics and should comply with the detailed guidelines and recommendations outlined by the various organizations and committees.[27,28] This entire approach has a broader outcome: gearing the shift in paradigm from conventional multi-tier test, involving sequential analysis of one gene after another, or biochemical testing before molecular confirmation, or protein studies before genetic testing, to a simpler two-tier testing procedure for

genetically heterogeneous diseases. This would include disease-related targeted panel testing when clinically available, and if this comes back negative, it is most likely that the gene is either a gene not known to be associated with the disease or known but not included in the targeted panel. The next immediate option would then be to opt for ES, which might identify the causative gene and variant, and in some cases even implicate a new gene in the disease in question. Rightly so, ACMG recommends opting for ES and WGS approaches only when disease-targeted testing is negative, not clinically available, or is unlikely to return a positive result in a timely and cost-effective manner.[29]

In conclusion, for genetically and phenotypically heterogeneous diseases with minimal clinical specificity, targeted gene panel testing should be the preferred approach to rule in or rule out the causative gene and disease subtype. With optimal coverage depth and thorough validation of assays, the expectation is that targeted gene panels can eventually be offered as stand-alone tests, without complementary Sanger confirmation. Though this is currently not the practice in clinical labs, studies such as this one and others[25,30,31] are sure to make it a reality.

Conflict of interest

The authors have nothing to disclose.

References

1. Flanigan KM. The muscular dystrophies. Semin Neurol 2012;32:255-63.

2. Mercuri E, Muntoni F. Muscular dystrophies. Lancet 2013;381:845-60.

3. Rocha CT, Hoffman EP. Limb-girdle and congenital muscular dystrophies: current diagnostics, management, and emerging technologies. Curr Neurol Neurosci Rep 2010;10:267-76.

4. Mercuri E, Muntoni F. The ever-expanding spectrum of congenital muscular dystrophies. Ann Neurol 2012;72:9-17.

5. Cossu G, Sampaolesi M. New therapies for Duchenne muscular dystrophy: challenges, prospects and clinical trials. Trends Mol Med 2007;13:520-6.

6. Mercuri E, Muntoni F. Muscular dystrophy: new challenges and review of the current clinical trials. Curr Opin Pediatr 2013;25:701-7.

7. Cacciottolo M, Numitone G, Aurino S, et al. Muscular dystrophy with marked Dysferlin deficiency is consistently caused by primary dysferlin gene mutations. Eur J Hum Genet 2011;19:974-80.

8. Wewer UM, Engvall E. Merosin/laminin-2 and muscular dystrophy. Neuromuscul Disord 1996;6:409-18.

9. Jones MA, Bhide S, Chin E, et al. Targeted polymerase chain reaction-based enrichment and next generation sequencing for diagnostic testing of congenital disorders of glycosylation. Genet Med 2011;13:921-32.

10. Valencia CA, Ankala A, Rhodenizer D, et al. Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. PLoS One 2013;8:e53083.

11. Valencia CA, Rhodenizer D, Bhide S, et al. Assessment of target enrichment platforms using massively parallel sequencing for the mutation detection for congenital muscular dystrophy. J Mol Diagn 2012;14:233-46.

12. Chin EL, da Silva C, Hegde M. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. BMC Genet 2013;14:6.

13. Jones MA, Rhodenizer D, da Silva C, et al. Molecular diagnostic testing for congenital disorders of glycosylation (CDG): detection rate for single gene testing and next generation sequencing panel testing. Mol Genet Metab 2013;110:78-85.

14. Bean LJ, Tinker SW, da Silva C, Hegde MR. Free the data: one laboratory's approach to knowledge-based genomic variant classification and preparation for EMR integration of genomic data. Hum Mutat 2013;34:1183-8.

15. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic acids research 2003;31:3812-4.

16. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nature methods 2010;7:248-9.

17. Costello M, Pugh TJ, Fennell TJ, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res 2013;41:e67.

18. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 2010;20:1297-303.

19. Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. Genome Biol 2011;12:R112.

20. Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genet Med 2014.

21. Laing NG. Genetics of neuromuscular disorders. Crit Rev Clin Lab Sci 2012;49:33-48.

22. Punetha J, Hoffman EP. Short read (next-generation) sequencing: a tutorial with cardiomyopathy diagnostics as an exemplar. Circ Cardiovasc Genet 2013;6:427-34.

23. Hayes J, Tzika A, Thygesen H, et al. Diagnosis of Copy Number Variation by Illumina Next Generation Sequencing is comparable in performance to Oligonucleotide Array Comparative Genomic Hybridisation. Genomics 2013.

24. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 2013;14:S1.

25. Cottrell CE, Al-Kateb H, Bredemeyer AJ, et al. Validation of a Next-Generation Sequencing Assay for Clinical Molecular Oncology. J Mol Diagn 2013.

26. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. Nat Rev Genet 2013;14:295-300.

27. Gargis AS, Kalman L, Berry MW, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol 2012;30:1033-6.

28. Aziz N, Zhao Q, Bry L, et al. College of American Pathologists' Laboratory Standards for Next-Generation Sequencing Clinical Tests. Arch Pathol Lab Med 2014.

29. ACMG Board of Directors. Points to consider in the clinical application of genomic sequencing. Genet Med 2012;14:759-61.

30. Ankala A, Hegde M. Genomic Technologies and the New Era of Genomic Medicine. J Mol Diagn 2013.

31. Sikkema-Raddatz B, Johansson LF, de Boer EN, et al. Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. Hum Mutat 2013;34:1035-42.

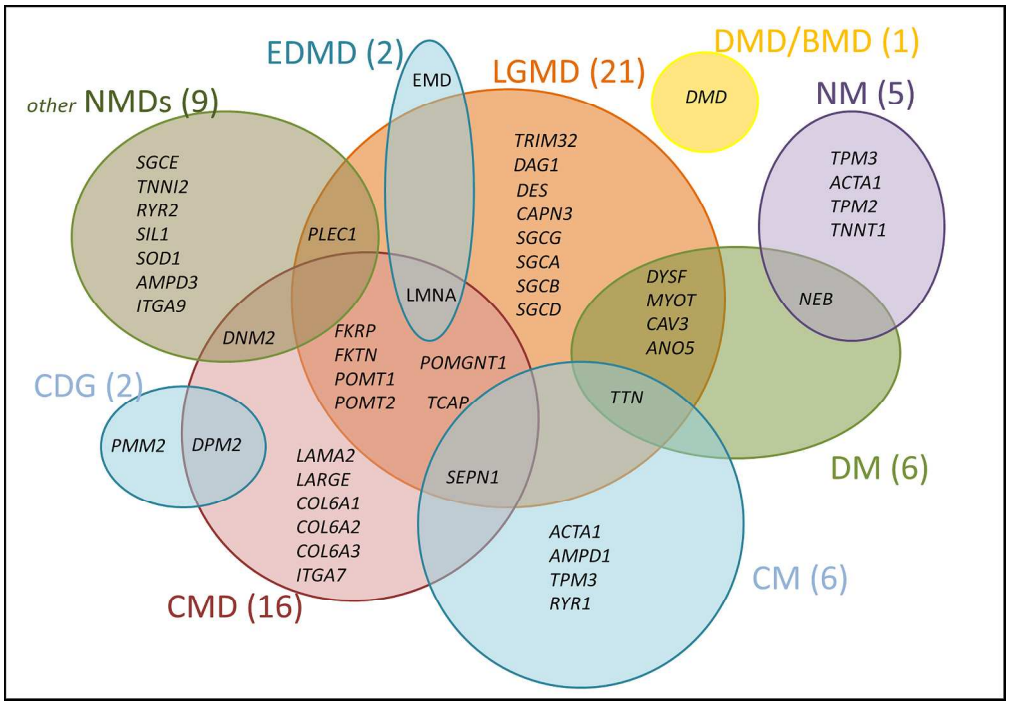


Figure 1. DMD – Duchenne muscular dystrophy, BMD – Becker muscular dystrophy, CDG – congenital disorders of glycosylation, LGMD – limb girdle muscular dystrophy, EDMD – Emery-Dreifuss muscular dystrophy, CM – congenital myopathy, DM – distal myopathy, NM – Nemaline myopathy, Other NMDs – include dystonias, metabolic myopathies, myasthenias.
318x220mm (300 x 300 DPI)

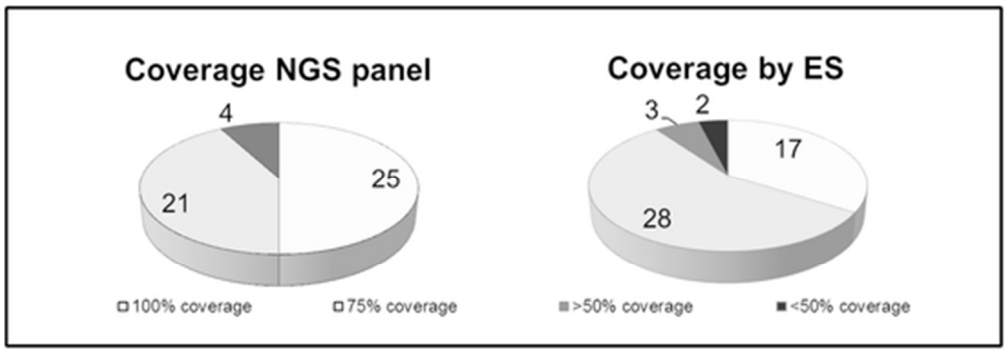


Figure 2: Comparison of coverage for NMD exons by NMD NGS panel and ES. For panels, an exon is considered to have low coverage if even a single nucleotide within the region of interest has coverage of <20X (for exons ± 2 bp consensus splice sites) or of <10X (for ± 3 bp through ± 10 bp flanking the exons). These low-coverage exons are eventually covered by Sanger sequencing to provide complete coverage for all exons of the genes included in a panel. While mutations in low-coverage exons would be missed by ES, they will be picked up by the Sanger fill-ins in panels where all low- and no-coverage exons are analyzed by complementary Sanger sequencing
46x16mm (300 x 300 DPI)

Table 1. Genes included in the different NGS panels offered

DMD	LGMD	CMD	NMD*
DMD	*CAPN3*	*COL6A1*	*ACTA1*
	CAV3	*COL6A2*	*AMPD1*
	DYSF	*COL6A3*	*AMPD3*
	LMNA	*FKRP*	*ANO5*
	MYOT	*FKTN*	*DES*
	SGCA	*ITGA7*	*EMD*
	SGCB	*LAMA2*	*NEB*
	SGCD	*LARGE*	*PLEC*
	TCAP	*POMGNT1*	*PMM2*
	SGCG	*POMT1*	*RYR1*
	TRIM32	*POMT2*	*RYR2*
		SEPN1	*SIL1*
			TTN
			TNNI2
			TNNT1
			TPM2
			TPM3

*NMD panel includes the genes listed under it, as well as all the other genes in the DMD, LGMD, and CMD lists

Table 2. Sanger confirmed variants interrogated by NGS panel testing for assay validation

Sample ID	Gene	Variant detection by NGS	Sample ID	Gene	Variant detection by NGS
Concordant calls					
NMD-A03	*DMD*	Het c.7223_7224delCT	NMD-A09	*COL6A3*	Het c.6855G>C
NMD-A03	*DMD*	Het c.7728T>C	NMD-A09	*COL6A3*	Het c.7929G>A
NMD-A04	*DMD*	Hemi c.5574_5575delTA	NMD-A09	*COL6A3*	Het c.8780T>C
NMD-A04	*DMD*	Hemi c.1635A>G	NMD-A09	*COL6A3*	Het c.9034G>C
NMD-A06	*DYSF*	Het c.3516_3517delTT	NMD-A09	*COL6A3*	Het c.9206C>T
NMD-A06	*DYSF*	Het c.1920T>C	NMD-A09	*COL6A3*	Het c.*7G>C
NMD-A06	*DYSF*	Het c.2676A>T	NMD-A10	*COL6A2*	Hom c.2572C>T
NMD-A06	*DYSF*	Het c.4101C>A	NMD-A11	*COL6A2*	Het c.1769C>T
NMD-A07	*DYSF*	Het c.265C>T	NMD-A11	*COL6A2*	Het c.663C>T
NMD-A07	*DYSF*	Hom c.1827T>C	NMD-A11	*COL6A2*	Het c.1333-8T>C
NMD-A07	*DYSF*	Hom c.2583A>T	NMD-A11	*COL6A2*	Het c.1609-10C>T
NMD-A07	*DYSF*	Hom c.4008C>A	NMD-A11	*COL6A2*	Het c.2803G>A
NMD-A09	*COL6A3*	Het c.6156+2T>G	NMD-A11	*COL6A2*	Het c.2979C>T
NMD-A09	*COL6A3*	Het c.8009C>T	NMD-A08	*EMD*	Hemi c.346delG
NMD-A09	*COL6A3*	Het c.6653C>T	NMD-A09	*COL6A3*	Het c.4436A>T
NMD-A09	*COL6A3*	Het c.6753+52T>G	NMD-A09	*COL6A3*	Het c.768C>T
NMD-A09	*COL6A3*	Het c.6753+53T>C			

Sample ID	Gene	Variant detection by Sanger	Variant detection by NGS	Explanation
Discordant or undetected calls				
NMD-A05	*DMD*	Hemi c.2169-1_2169delinsAC	c.2169-1G>A, c.2169G>C	Indel is represented as two independent SNVs
NMD-A04	*DMD*	Hemi c.2645G>A	No call	Reference has A*
NMD-A09	*COL6A3*	Hom c.6369G>A	No call	Reference has A*
NMD-A09	*DYSF*	Het c.1766C>T	No call	Reference has T*
NMD-A10	*COL6A2*	Hom c.2094G>A	No call	Zero coverage
NMD-A10	*COL6A2*	Hom c.2097C>T	No call	Zero coverage

SNV- Single nucleotide variants. As mentioned in the text, regions or exons with low coverage (<20X) are analyzed by Sanger sequencing as part of the test and so any such variants missed by NGS part of the panel test (like the *COL6A2* variants above) will still be analyzed by targeted panel assay. *Different reference sequences seem to have been used by the two testing laboratories. The refore, while the detected allele is the same between the two laboratories, there is discordance in the variant call made.

Table 3. List of pathogenic variants assessed for validating the NMD NGS panel test for its versatility of variant detection.

Sample ID	Gene	Nucleotide change	AA change	Detected	Coverage	Mutation type
Detected variants						
NMD-A01	*DMD*	c.5697_5698insA	FS	Yes	34	1-bp insertion
NMD-A02	*DMD*	c.5697_5698insA	FS	Yes	26	1-bp insertion
NMD-A06	*DYSF*	c.2202_2206delCCACC	FS	Yes	80	5-bp deletion
NMD-A06	*DYSF*	c.2200A>T	p.T734S	Yes	80	base substitution
NMD-A06	*DYSF*	c.3516_3517delTT	FS	Yes	419	2-bp deletion
NMD-A09	*COL6A3*	c.8009C>T	p.A2670V	Yes	188	base substitution
NMD-A10	*COL6A2*	c.2572C>T	p.Q858*	Yes	37	base substitution
NMD-A11	*COL6A2*	c.1769C>T	p.T590M	Yes	22	base substitution
NMD-A12	*DES*	c.1398_1399delinsTT	FS	Yes	263	indel
NMD-E01	*DMD*	c.9854_9863delTGAGACTGGA	FS	Yes	37	10-bp deletion
NMD-E05	*DMD*	c.4412_4413insGTCT	FS	Yes	72	4-bp insertion
Undetected Variants*						
NMD-E02	*DMD*	deletion of exon 43	-	No	0	single exon deletion
NMD-E04	*DMD*	deletion of exon 56	-	No	0	single exon deletion
NMD-E03	*DMD*	duplication of exon 49-51	-	No	-	intragenic duplication

*Current NGS testing and analysis algorithms are not designed to confidently detect large deletions and duplications. However, complementary comparative genomic hybridization arrays (aCGH), included with targeted gene panel testing, are highly efficient at detecting heterozygous as well as homozygous deletions and duplications. This is an added advantage of panel testing over single gene testing or clinical exome sequencing. In the table, lack of coverage (zero coverage) for deletions in *DMD* in males may indirectly indicate that these exons were deleted. However, this may not always be true and confirmatory results are therefore obtained by aCGH analysis. FS, frameshift.

Table 4. Diagnostic yields of the different NMD clinical tests

Clinical test ordered	No. of tests with diagnosis / No. of tests performed	Diagnostic yield (%)
CMD single gene testing	31/204	15
CMD *COL6A* subpanel	15/71	21

CMD comprehensive panel	32/88	36
LGMD single gene testing	67/343	19
LGMD comprehensive panel	25/96	26
NMD comprehensive panel	37/81	46

Table 5. Pathogenic variants of disease subtypes potentially missed by targeted NGS or ES method

	# Variants	# Exons	# Genes	Genes
CMD by panel	5	3	2	*SEPN1, COL6A1*
CMD by ES	11	10	6	*LAMA2, SEPN1, POMT1, POMT2, COL6A2, FKRP*
LGMD by panel	4	4	1	*DYSF*
LGMD by ES	10	8	5	*DYSF, SGCG, LMNA, SGCA, SGCB*

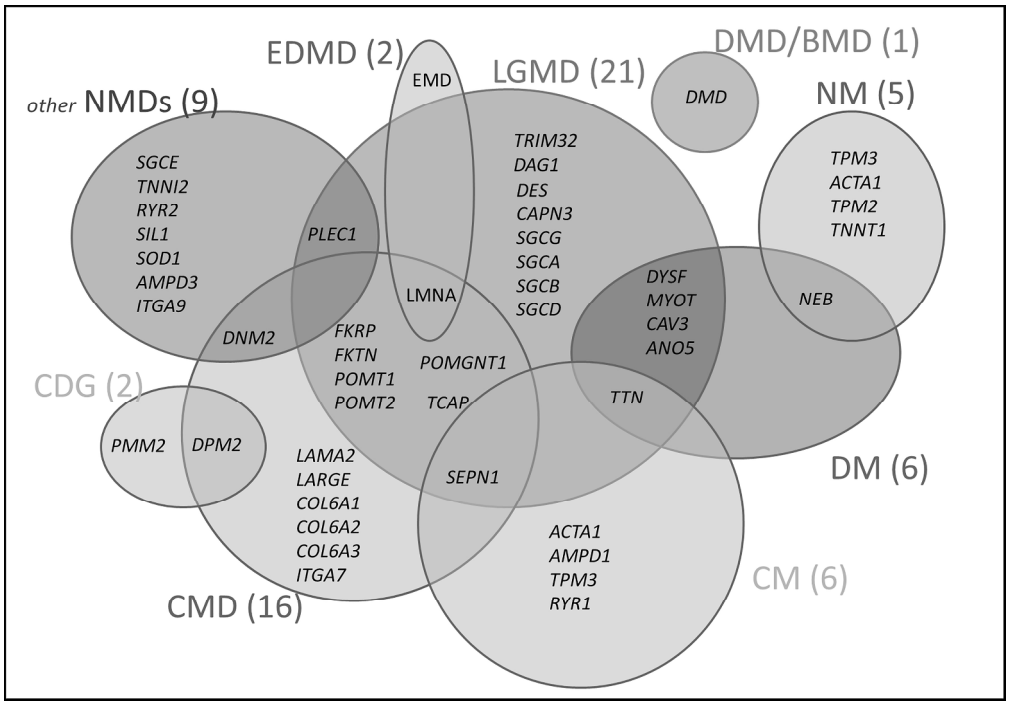
Figure legends:

Figure 1. DMD – Duchenne muscular dystrophy, BMD – Becker muscular dystrophy, CDG – congenital disorders of glycosylation, LGMD – limb girdle muscular dystrophy, EDMD – Emery-Dreifuss muscular dystrophy, CM – congenital myopathy, DM – distal myopathy, NM – Nemaline myopathy, Other NMDs – include dystonias, metabolic myopathies, myasthenias.

Figure 2. Comparison of coverage for NMD exons by NMD NGS panel and ES. For panels, an exon is considered to have low coverage if even a single nucleotide within the region of interest has coverage of <20X (for exons ± 2 bp consensus splice sites) or of <10X (for ± 3 bp through ± 10 bp flanking the exons). These low-coverage exons are eventually covered by Sanger sequencing to provide complete coverage for all exons of the genes included in a panel. While mutations in low-coverage exons would be missed by ES, they will be picked up by the Sanger fill-ins in panels where all low- and no-coverage exons are analyzed by complementary Sanger sequencing.

Supplementary Table S1. Low coverage (<20X) genes and exons of Targeted NMD panel

Gene	Exon	Gene	Exon
CAPN3	12	*PLEC*	18
CAPN3	14	*PMM2*	7
COL6A1	7	*POMGNT1*	19
COL6A1	8	*POMT2*	11
COL6A1	9	*RYR1*	61
COL6A1	10	*RYR1*	70
COL6A1	11	*RYR1*	91
COL6A1	18	*RYR1*	99
COL6A1	24	*RYR2*	4
COL6A1	25	*RYR2*	43
COL6A1	27	*EMD*	5
COL6A1	28	*FKRP*	4
COL6A1	29	*LMNA*	10
COL6A1	34	*LMNA*	11
COL6A2	8	*LMNA*	12
COL6A2	9	*NEB*	4
COL6A2	10	*SEPN1*	1
COL6A2	16	*SGCA*	8
COL6A2	17	*SGCD*	2
COL6A2	18	*TPM3*	7
COL6A2	20	*TTN*	107
COL6A2	23	*TTN*	153
COL6A3	16	*TTN*	154
COL6A3	18	*TTN*	155
COL6A3	19	*TTN*	156
COL6A3	24	*TTN*	157
DES	2	*TTN*	158
DES	7	*TTN*	166
DMD	62	*TTN*	234
DMD	63	*TTN*	244
DMD	71	*TTN*	259
DYSF	2	*TTN*	274
DYSF	9	*DYSF*	35
DYSF	17	*DYSF*	36



318x220mm (300 x 300 DPI)