Review

Long-read sequencing for diagnosis of genetic myopathies

Dennis Yeow [1,2,3,4] Laura Ivete Rudaks [1,2,3,5] Ryan Davis,[2,3,6] Karl Ng,[2,7] Roula Ghaoui,[8,9,10] Pak Leng Cheong,[2,11,12,13] Gianina Ravenscroft,[14,15] Marina Kennerson,[2,11,16] Ira Deveson,[17,18] Kishore Raj Kumar[1,2,3,18]

ABSTRACT

Genetic myopathies are caused by pathogenic variants in >300 genes across the nuclear and mitochondrial genomes. Although short-read next-generation sequencing (NGS) has revolutionised the diagnosis of genetic disorders, large and/or complex genetic variants, which are over-represented in the genetic myopathies, are not well characterised using this approach. Long-read sequencing (LRS) is a newer genetic testing technology that overcomes many of the limitations of NGS. In particular, LRS provides improved detection of challenging variant types, including short tandem repeat (STR) expansions, copy number variants and structural variants, as well as improved variant phasing and concurrent assessment of epigenetic changes, including DNA methylation. The ability to concurrently detect multiple STR expansions is particularly relevant given the growing number of recently described genetic myopathies associated with STR expansions. LRS will also aid in the identification of new myopathy genes and molecular mechanisms. However, use of LRS technology is currently limited by high cost, low accessibility, the need for specialised DNA extraction procedures, limited availability of LRS bioinformatic tools and pipelines, and the relative lack of healthy control LRS variant databases. Once these barriers are addressed, the implementation of LRS into clinical diagnostic pipelines will undoubtedly streamline the diagnostic algorithm and increase the diagnostic rate for genetic myopathies. In this review, we discuss the utility and critical impact of LRS in this field.

INTRODUCTION

The genetic myopathies, including the muscular dystrophies, are a large group of clinically and genetically heterogeneous muscle disorders. Collectively, they have a prevalence of up to 1 in 4500 individuals in the general population.[1] Genetic diagnosis of this group of disorders can be complex due to a number of factors: the large number of potentially causative genes (>300), genotypic and phenotypic heterogeneity (ie, variants in different genes presenting with similar clinical features and variants within a single gene presenting with different clinical features in individual patients) and the diverse range of genetic variant types eg, small variants, such as single nucleotide variants (SNVs) and small insertions and deletions (indels), as well as larger structural variants (SVs), such as copy number variants (CNVs), insertions, inversions, translocations and short tandem repeat (STR) expansions) and epigenetic alterations (eg, altered DNA methylation) that cause disease.[2]

Short-read next-generation sequencing (NGS) has revolutionised the diagnosis of genetic myopathies by providing cost-efficient, high-throughput, massively parallel genetic sequencing.[3] However, while NGS accurately identifies SNVs and indels, it is limited in its ability to resolve larger or more complex genetic variants, many of which are relevant to genetic myopathies and muscular dystrophies. In current clinical practice, a number of alternative genetic testing technologies are used to identify these other genetic variant classes, including multiplex ligation-dependent probe amplification (MLPA), microarray and/or Bionano optical genome mapping (OGM) for SVs, repeat-primed PCR (RP-PCR) and/or Southern blot (SB) for STR expansions or macrosatellite contractions, and bisulphite sequencing to assess altered methylation, among many others (figure 1A).[2 4] The diversity of methods required to identify these various causative genetic variants, combined with the genotypic and phenotypic heterogeneity of this group of disorders, means that it is not uncommon for patients with suspected genetic myopathies to undergo multiple rounds of genetic testing. This complex, multistep, iterative process of diagnostic testing may contribute to the significant diagnostic delay that many patients with genetic myopathy experience.[2] Furthermore, there remains a substantial diagnostic gap

For numbered affiliations see end of article.

Correspondence to
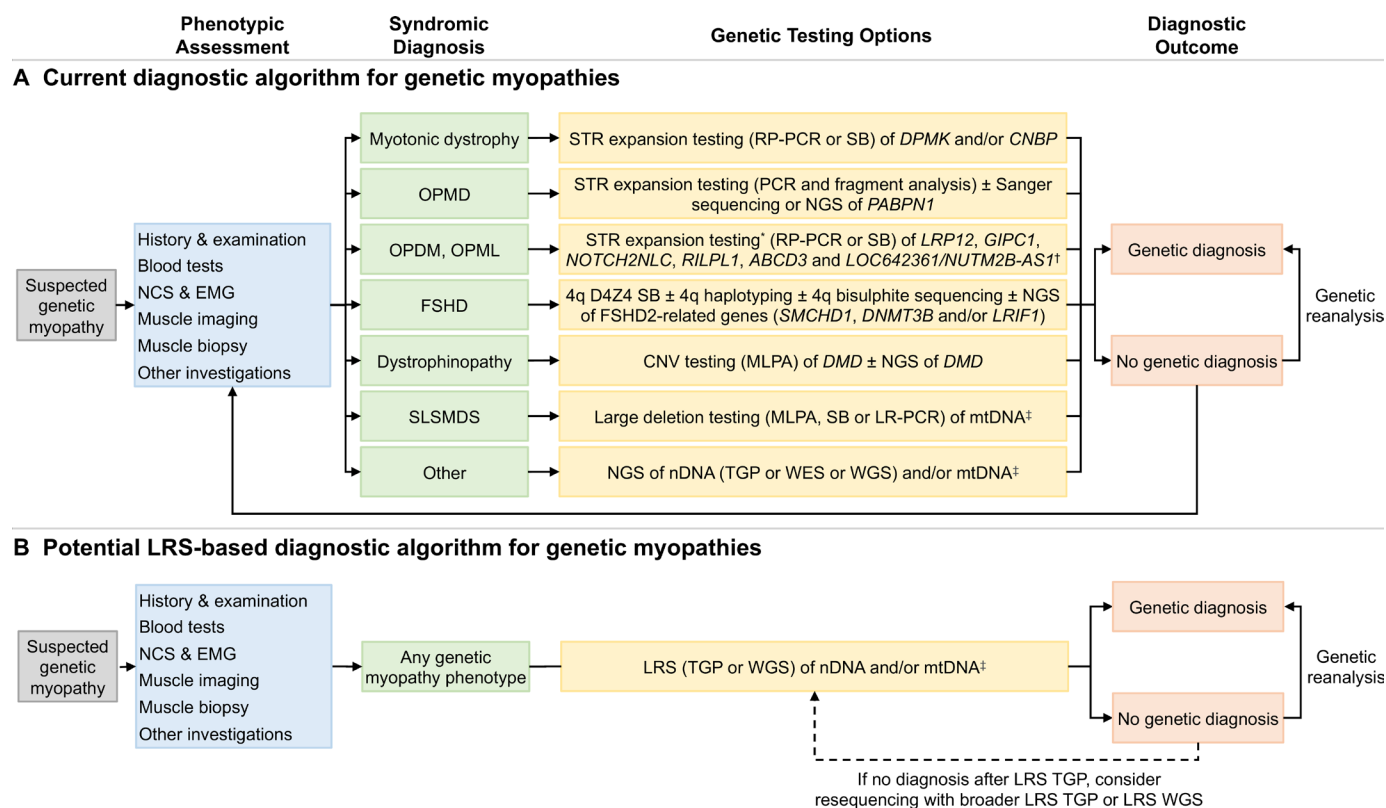Dr Dennis Yeow;
d.yeow@neura.edu.au

Figure 1 Genetic testing algorithms for genetic myopathies. (A) The current genetic testing algorithm typically involves thorough phenotypic evaluation with the goals of identifying a syndromic diagnosis and guiding selection of the appropriate genetic testing technology. A negative genetic test result may be followed by re-evaluation of the phenotype and, potentially, selection of alternative genetic testing technologies in an iterative manner, contributing to diagnostic delay. Alternatively, reanalysis of previously acquired genomic data can also result in genetic diagnosis. (B) Potential streamlined genetic testing algorithm takes advantage of the ability of long-read sequencing to detect multiple genetic variant types in a single assay. CNV, copy number variant; EMG, electromyography; FSHD, facioscapulohumeral muscular dystrophy; MLPA, multiplex ligation-dependent probe amplification; mtDNA, mitochondrial DNA; NCS, nerve conduction studies; nDNA, nuclear DNA; NGS, next-generation sequencing; OPDM, oculopharyngodistal myopathy; OPMD, oculopharyngeal muscular dystrophy; OPML, oculopharyngeal myopathy with leukoencephalopathy; PCR, polymerase-chain reaction; RP-PCR, repeat-primed PCR; SB, Southern blot; SLSMDS, single large-scale mtDNA deletion syndrome; TGP, targeted gene panel; WES, whole exome sequencing; WGS, whole-genome sequencing. *STR expansion testing for OPDM/OPML loci may not be widely available on a clinical diagnostic basis. †The FAM193B locus is another putative OPDM-causing gene that could be assessed if/when it is validated as disease-causing. ‡Sequencing of mtDNA from muscle may be more sensitive than sequencing of mtDNA from blood for detection of pathogenic mtDNA variants.

with only 30%–60% of patients with suspected genetic myopathies receiving a molecular diagnosis using standard genetic sequencing technologies, although diagnostic rates may be augmented by the addition of RNA sequencing and/or proteomics.[5 6]

Long-read DNA sequencing (LRS) is a newer sequencing modality that addresses many of the limitations of NGS, potentially providing a single, unified assay to replace the complex combination of genetic tests that are currently used (figure 1B). In particular, in addition to massively parallel sequencing and identification of SNVs and indels, LRS is capable of concurrently characterising SVs and STR expansions,[7 8] phasing distant variants without the need for parental sequencing data[9] and assessing methylation status,[10] all within a single PCR-free assay. In this review, we present a summary of the advantages of LRS over NGS and how this improves the diagnosis of

genetic myopathies. No patient or public involvement was sought for this review.

OVERVIEW OF LONG-READ SEQUENCING

Sequencing by NGS typically involves fragmentation of native DNA into short fragments followed by either PCR amplification-dependent or PCR-free library preparation.[3] This library of DNA fragments is then sequenced, typically producing short reads (~150 bps) that are then aligned to a reference genome (figure 2A). Limitations of NGS include difficulty in aligning short reads to repetitive or structurally complex regions of DNA, difficulty phasing distant variants and molecular biases (eg, GC bias) introduced during PCR amplification or library preparation.[3] In contrast, LRS sequences long strands of native DNA, thus, obviating amplification biases and facilitating concurrent assessment of epigenetic changes
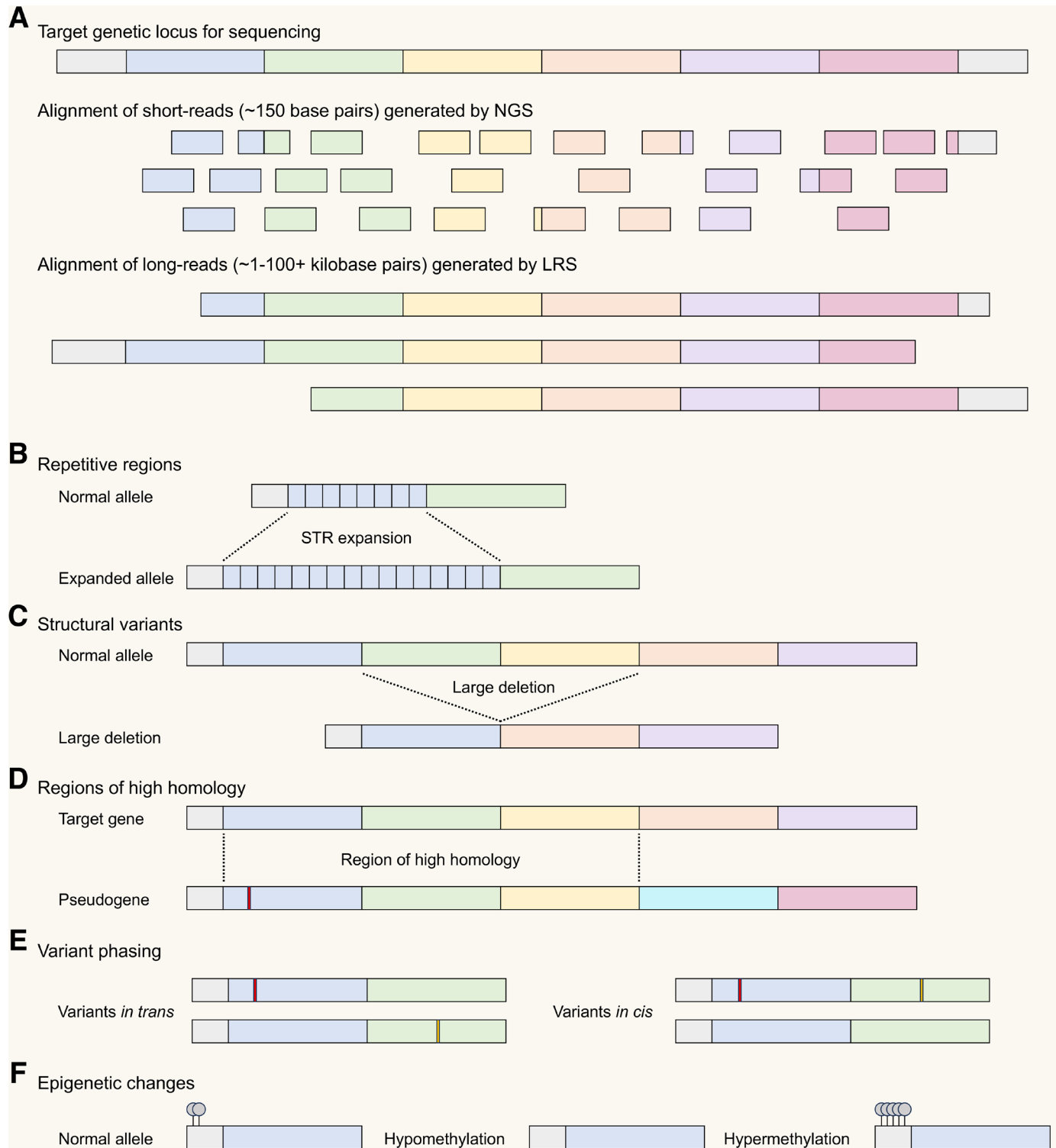
Figure 2 Comparison of short-read vs long-read sequencing for detecting challenging genetic variants. (A) Schematic showing alignment of short reads generated by next-generation sequencing (NGS) compared with long reads generated by long-read sequencing (LRS). The long read length produced by LRS facilitates accurate sequencing and alignment of genetic variants that are often difficult to characterise by NGS including changes in repetitive regions of the genome such as short tandem repeat (STR) expansions (B) and large structural variants (C). LRS also facilitates distinction between genomic regions that share regions of high homology, for example, allowing determination of whether an apparently pathogenic variant is present in a target gene or in a related pseudogene (D). The long-read length also facilitates phasing of relatively distant genetic variants, for example, in the context of an autosomal recessive disorder, helping to determine if two genetic variants are in trans (potentially disease-causing) or in cis (not expected to be disease-causing) (E). Separate to advantages related to increased read length, LRS also facilitates concurrent assessment of epigenetic changes including potentially disease-related changes in DNA methylation (F).

(eg, DNA methylation). LRS produces reads up to many thousands of base pairs in length, allowing resolution of complex genetic variants that NGS is unable to characterise (figure 2B–F).

The two current major LRS platforms are Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) HiFi sequencing. Detailed comparisons of the two technologies are documented elsewhere.[11] Briefly, PacBio HiFi LRS provides higher per base accuracy compared with ONT LRS while ONT LRS has the potential to generate longer read lengths than PacBio HiFi (up to several Mbp with ONT 'ultra-long' read sequencing). While throughput and sequencing accuracy were previous major limitations for PacBio HiFi and ONT LRS, respectively, newer iterations of their respective sequencing hardware, flow cell technology and sequencing chemistries combined with improvements in LRS bioinformatic pipelines have resulted in rapid improvements in these parameters.[11] Both LRS platforms support projects of various sizes, with mid-throughput bench-top instruments (PacBio Vega; ONT GridION or P2) and high-throughput instruments (PacBio Revio; ONT PromethION P24 or P48) now available.

Like NGS, LRS can be applied to the whole genome or used for targeted sequencing of specific genes or loci. As an alternative to PCR amplification for target enrichment, amplification-free targeted LRS can be achieved using the clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 system although this requires careful design and optimisation of guide RNA molecules for the loci of interest.[12] An alternative approach is ONT's 'adaptive sampling' functionality, wherein the nanopores dynamically 'accept' or 'reject' individual DNA molecules for sequencing in real-time based on a set of programmed genomic coordinates for the genetic loci of interest.[8] This enables targeted sequencing without additional molecular processes for target enrichment. While ONT adaptive sampling facilitates targeted sequencing with minimal additional sample preparation, it is more computationally demanding and the on-target coverage obtained is generally lower than CRISPR/Cas9-based targeting methodologies.[13] Therefore, use of ONT adaptive sampling may be preferable for large, targeted gene panels encompassing up to hundreds of loci; whereas, CRISPR/Cas9-based targeted LRS may be more suitable when higher coverage of a smaller number of genetic loci is desired.

In addition to characterisation of complex genomic variants in individual patients with genetic disorders, LRS has also contributed to an improved understanding of the normal human genome. Use of ONT and PacBio LRS was crucial to generating the most complete human genome references to date, known as the Telomere to Telomere-CHM13 reference and the human pangenome.[14 15] These more accurate human reference genomes facilitate improved variant calling and detection of pathogenic variants in individual patients.[15]

LRS FOR GENETIC MYOPATHY DIAGNOSIS
Repeat expansion myopathies

Large STR expansions in the *DMPK* and *CNBP* genes are the cause of myotonic dystrophy type 1 (DM1) and type 2 (DM2).[16] Currently, these disorders are typically diagnosed using RP-PCR and/or SB assays specifically designed for each of these genetic loci, due to the difficulty of accurately characterising large STR expansions using NGS. In contrast, the $(GCN)_n$ STR expansions in the *PABPN1* gene that cause oculopharyngeal muscular dystrophy (OPMD) are relatively short (11–18 repeats) and are often diagnosed using PCR and fragment analysis.[17] Sanger sequencing and NGS can also detect these short STR expansions in patients with OPMD[18] and have the added benefit of also being able to detect rare cases of OPMD caused by SNVs in *PABPN1*.[19] In just the last 5 years, six additional large STR expansions have been described as the cause for a number of genetic myopathies: $(CGG/CCG)_n$ STR expansions in *LRP12*, *GIPC1*, *NOTCH2NLC*, *RILPL1*, *ABCD3* and *LOC642361/NUTM2B-AS1* cause oculopharyngodistal myopathy (OPDM) types 1–4, *ABCD3*-related OPDM and oculopharyngeal myopathy with leukoencephalopathy (OPML), respectively.[20–23] Even more recently, a similar $(CGG/CCG)_n$ STR expansion in the *FAM193B* gene has been described as the putative cause of OPDM in a pair of siblings.[24] RP-PCR assays for these recently described STR expansion loci are not widely available on a clinical diagnostic basis. Although OPDM types 1–4 and OPML were initially only described in East Asian and South-East Asian populations, this group of disorders is increasingly being identified in other ethnic groups. For example, *NOTCH2NLC* STR expansions, which cause OPDM type 3 as well as neuronal intranuclear inclusion disease (NIID), have recently been found in New Zealand and Cook Island Māori populations[25] and also in occasional cases of NIID in European patients,[26 27] while pathogenic $(CGG)_n$ STR expansions in *ABCD3* have been shown to cause OPDM in individuals with European ancestry.[23] The rapidly increasing number of myopathy-causing STR expansion loci and the phenotypic similarity between many of these disorders (ie, OPMD, OPDM and OPML) necessitates the development of an assay capable of sequencing all these STR loci in parallel, as opposed to the traditional, low-throughput approach of sequential, single locus RP-PCR. Targeted LRS is a simple solution to this problem and is capable of concurrently sequencing all disease-causing STR loci across the human genome in a single assay.[8]

In addition to high-throughput identification of STR expansions, LRS, unlike NGS, is able to characterise the full sequence of expanded STR alleles including STR motif variations and the presence of interruptions within STR expansions.[12] Detection of these subtle changes is clinically relevant as there is increasing recognition that such STR interruptions and motif variations can modulate the clinical phenotype across many neurologic STR expansion disorders.[12 28] For example, in patients with DM1, LRS can identify the presence of CGG interruptions

within the $(CTG)_n$ STR expansion in *DMPK*.[29] These interruptions are associated with a milder phenotype than otherwise expected for a given $(CTG)_n$ STR length. Similarly, in patients with DM2, LRS can identify the presence of interruptions, motif variations and the degree of somatic mosaicism of the $(CCTG)_n$ STR expansion in *CNBP*.[30] Although previous studies have found that $(CCTG)_n$ STR expansion length in *CNBP*, as assessed by SB, does not correlate with age at onset of DM2,[31] it is possible that the additional information provided by LRS regarding STR expansion content may improve genotype-phenotype correlations.[30] Therefore, by more comprehensively characterising the size and content of STR expansions, LRS may be able to provide important prognostic information to patients with STR expansion myopathies that would not be available from traditional RP-PCR or SB assays.

Beyond STR expansions, LRS has also aided in the discovery of pathogenic expansions of longer tandem repeat structures. In particular, ONT LRS helped to identify the expansion of a 99 base pair repeat in the *PLIN4* gene as the cause of a newly described rimmed vacuolar distal myopathy.[32] Patients with this myopathy have ≥39 repeats compared with ~30 repeats in healthy controls.[32] The large size and highly repetitive nature of this genetic locus makes it relatively refractory to NGS, again highlighting the advantages of LRS in the diagnosis of genetic myopathies.

PacBio HiFi and ONT LRS are both capable of sequencing expanded STRs. However, PacBio HiFi appears to achieve higher base calling accuracy within the STRs, likely due to the lower per base accuracy of ONT.[33] On the other hand, the greater read length achievable with ONT, including 'ultra-long' (>100 kbps) reads, is preferable for sequencing very large STR expansions (eg, in DM2, where the STR expansion may extend up to 50 kb) which may be difficult for PacBio HiFi LRS (~15–20 kbp reads) to span entirely.[30]

Facioscapulohumeral muscular dystrophy

Facioscapulohumeral muscular dystrophy (FSHD) is a genetically complex disorder driven by aberrant expression of the *DUX4* gene, which is buried within a D4Z4 macrosatellite repeat array located in the subtelomeric region of the long arm of chromosome 4 (4q).[34] *DUX4* gene expression is normally repressed in adult muscle tissue by hypermethylation of the D4Z4 array but is facilitated in FSHD by relative hypomethylation at this locus. In the most common form, FSHD type 1 (FSHD1), D4Z4 array hypomethylation is caused by a contraction in the number of D4Z4 repeats and expression of the *DUX4* gene requires this contraction to occur on a permissive 4q haplotype, called 4qA.[34] In FSHD type 2 (FSHD2), D4Z4 array hypomethylation results from pathogenic sequence variant(s) in one of several distantly located genes involved in methylation (*SMCHD1*, *DNMT3B* or *LRIF1*), but disease is, again, only manifest in the presence of at least one permissive 4qA haplotype, such that FSHD2 is digenically inherited.[34] Current diagnostic approaches to FSHD typically involve SB (to assess for D4Z4 repeat array contraction), and when necessary, 4q haplotyping via specialised SB (to determine the presence of a permissive 4qA haplotype), NGS (to sequence *SMCHD1*, *DNMT3B* and/or *LRIF1*) and/or bisulphite sequencing to profile D4Z4 methylation.[34] More complex forms of FSHD have been described resulting from large proximally extended D4Z4 deletions and translocations between the D4Z4 array on 4q and a homologous D4Z4 repeat array on chromosome 10q—these cases can elude standard SB testing.[35][36] While alternative genomic mapping technologies, including Bionano OGM and molecular combing, can identify these more complex D4Z4 region SVs,[4][37] these methods cannot assess D4Z4 methylation or identify sequence variants in the FSHD2 genes. In contrast, LRS can concurrently determine the size of the D4Z4 repeat array and the presence of SVs within the array, analyse and phase the 4q haplotype, profile allele-specific D4Z4 methylation and identify standard sequence variants in FSHD2-related genes, making it well suited to

Table 1 Genetic and epigenetic changes identified by different genetic testing technologies in the evaluation of facioscapulohumeral muscular dystrophy (FSHD)

Genetic testing technology	4q D4Z4 repeat array size	Phased 4q haplotype	D4Z4 methylation	Sequence variants in FSHD2-related genes
Southern blot	✓	✓		
Molecular combing	✓	✓		
Optical genome mapping	✓	✓		
Bisulphite sequencing			✓	(✓)*
Next-generation sequencing				✓
Long-read sequencing	✓	✓	✓	✓

*In the context of FSHD, targeted bisulphite sequencing is most commonly performed to assess D4Z4 array methylation, although could also theoretically be used to identify sequence variants in FSHD2-related genes.

comprehensively assessing FSHD genetics (table 1).[38] A good example of this was the recent use of LRS to identify and describe a novel genetic form of FSHD caused by in cis D4Z4 repeat array duplications, which were not detected by SB assays or standard methylation assays.[39]

Genetic myopathies due to structural variants

Pathogenic SVs may be responsible for up to 5% of all genetic myopathies[40] and this proportion is even higher for specific myopathies, including those related to *DMD* and *LAMA2*, where CNVs account for up to 75% and 18% of pathogenic variants, respectively.[41][42] Traditionally, NGS has had difficulty resolving large and/or complex SVs due to the inability of short reads to span distantly located genomic breakpoints. However, improvements in bioinformatic analysis have increased the detection of SVs from NGS data.[43] Despite this, for genetic myopathies where CNVs are a common cause of disease, such as DMD, MLPA is often considered a first-line diagnostic genetic test.[2] However, MLPA cannot detect SNVs and indels which cause up to 25% of DMD cases, necessitating reflex NGS of *DMD* in cases where MLPA is negative. Additionally, MLPA may miss small CNVs and copy number neutral SVs (eg, inversions, translocations). As an example of the utility of LRS for detection of SVs, in a cohort of nine families with muscular dystrophy that had undergone standard clinical genetic testing, ONT whole-genome LRS identified four previously undetected pathogenic variants, including SVs in *DMD* and *LAMA2* as well as deep intronic SNVs in *DMD*.[44] Another example involves the group of mitochondrial myopathies caused by single large-scale mitochondrial DNA (mtDNA) deletions, including Kearns-Sayre syndrome.[45] Clinical testing for these large mtDNA deletions, which are often present at low levels of heteroplasmy, typically relies on low-throughput assays like SB or long-range PCR, although they can also be detected by NGS with appropriate bioinformatic analysis.[46] Similar to other SVs, LRS provides a fast and accurate method of identifying these large-scale mtDNA deletions.[47]

Another type of complex SV that can cause genetic myopathy is a mobile element insertion (MEI), wherein a transposable genomic element integrates itself into a distantly located genetic locus and disrupts a myopathy-related gene. The prototypical MEI-related genetic myopathy is Fukuyama congenital muscular dystrophy, which is caused by a SINE-VNTR-Alu retrotransposon insertion into the 3′ untranslated region of the *FKTN* gene.[48] This particular MEI event in the *FKTN* gene is suspected to have originated ~3000 years ago in a Japanese founder.[49] However, transposable elements remain active within the human genome and, when integrated into a myopathy-related gene, can be the cause of sporadic cases of other types of genetic myopathy.[50][51] Modern bioinformatic analysis of NGS data can detect a subset of MEI events, but LRS approaches are more accurate and comprehensive, detecting and localising MEI events that are missed by NGS platforms.[52] For example, ONT LRS was able to identify the presence of a pathogenic LINE-1 retrotransposon MEI into intron 1 of *DMD* in a case of Becker muscular dystrophy, where MLPA and NGS failed to identify a causative pathogenic variant.[53]

While both PacBio HiFi and ONT LRS perform well in detecting SVs, ONT generates more artefactual SVs than PacBio HiFi, likely as a result of base-calling errors.[7] However, ONT appears to be superior for the detection of particularly large SVs, especially large insertions, likely related to its longer read length, enabling improved read alignment and assembly.[11]

Highly homologous genomic regions

Genes with high levels of internal homology due to the presence of repetitive DNA sequences (eg, *TTN*, *NEB*) may be difficult to accurately sequence using NGS. In contrast, LRS of *TTN* provides accurate localisation of genetic variants, particularly within the repetitive segments of the gene, leading to improved diagnosis of *TTN*-related myopathies.[54] For similar reasons, genes that have highly homologous pseudogenes can be challenging to accurately sequence using NGS. For example, exons 46–48 of the *FLNC* gene share high sequence homology with the pseudogene *LOC392787*. NGS of *FLNC* without thorough consideration of the *LOC392787* pseudogene can result in the misdiagnosis of filamin C-related myopathy.[55] The issue of sequencing genes with highly homologous pseudogenes is easily overcome by the use of LRS due to greater alignment specificity of longer read lengths.[56]

Variant phasing

Genetic sequencing technologies are generally unable to reliably phase variants that are separated by regions of homozygosity longer than the read length of the technology. This means NGS is largely unsuitable for phasing distant variants due to its short-read length, except when familial sequencing data is available. Phasing is important in autosomal recessive disorders to prove that two variants are in trans (ie, compound heterozygosity), where they would be potentially disease-causing, and not in cis, where they would not be expected to cause an autosomal recessive disease. To date, phasing of distantly located gene variants by NGS has, in large part, relied on segregation analysis involving testing of parents and/or siblings to infer phasing of the variants in the proband. However, familial DNA is not always readily available, particularly for adult myopathy patients. In contrast, by virtue of the long-read lengths, LRS can determine the phase of relatively distant variants, leading to improved genetic diagnosis without the need for parental/sibling sample testing for phasing.[9]

Methylation

NGS-based techniques for assessment of DNA methylation, such as bisulphite sequencing, can characterise the presence of 5-methylated cytosine (5mC) but this typically requires additional preparation of the DNA substrate

(eg, bisulphite treatment).[57] In contrast, LRS can identify 5mC as well as a host of other epigenetic changes (eg, 6-methylated adenine) without the need for additional sample preparation.[10] Furthermore, due to the long-read lengths, LRS can also characterise allele-specific methylation changes even in highly repetitive or complex genomic regions. In the context of myopathies, this is of clear relevance to FSHD, which is pathogenetically driven by hypomethylation of the D4Z4 repeat array (see above). Assessment of methylation status of highly repetitive DNA is also relevant for determining pathogenicity and penetrance of certain OPDM-related STR expansions. For example, individuals with extremely long STR expansions in certain OPDM genes are paradoxically asymptomatic, likely due to hypermethylation of these extremely long expansions.[58] Interestingly, due to the meiotic instability of these STR expansions, these asymptomatic individuals may have children that are affected by NIID or OPDM due to contraction of the extremely long, hypermethylated STR expansion in the parent to slightly shorter but less methylated STR expansion in the offspring.[58] As such, the ability of LRS to concurrently characterise STR expansion size and allele-specific methylation status may provide important information to OPDM patients and their families regarding disease penetrance and inheritance.

Concurrent sequencing and assessment of DNA methylation may also aid in functional validation of variants of uncertain significance. For example, pathogenic variants in *DNMT3A*, which encodes a DNA methyltransferase, cause a spectrum of neurodevelopmental disorders including Tatton-Brown-Rahman syndrome (TBRS)[59] and are associated with a specific, recognisable epigenetic methylation profile.[60] Recently, a patient with a congenital myopathy was reported with a novel de novo missense variant in *DNMT3A*.[61] Use of a separate EpiSign methylation assay identified a methylation profile in this patient that matched the known profile of TBRS, supporting pathogenicity of the novel *DNMT3A* variant and expanding the phenotypic spectrum of *DNMT3A*-related disorders to include prominent congenital myopathy. Although LRS was not used here, theoretically, LRS would be well suited to diagnose this and similar cases through concurrent identification of both the *DNMT3A* variant itself and the associated abnormal genome-wide methylation profile in a single assay.

OTHER APPLICATIONS OF LRS

Although this review has primarily focused on LRS of DNA, ONT LRS can also be applied to RNA and protein. ONT RNA sequencing can generate amplification-free, long-read transcriptomic and epitranscriptomic (ie, post-transcriptional RNA modification) data, which has advantages over current short-read, amplification-dependent RNA sequencing methodologies, analogous to the advantages of LRS compared with NGS of DNA. As an example, ONT RNA sequencing in *DUX4*-expressing muscle cells has revealed novel downstream transcriptomic alterations

induced by *DUX4* expression, contributing to a better understanding of FSHD pathogenesis.[62] Direct protein sequencing using ONT is also emerging as a novel method of generating proteomic and epiproteomic (ie, post-translational protein modification) data with single molecule resolution.[63] Further exploration and integration of these lines of long-read multiomic data is likely to aid functional validation of novel genetic variants (eg, functional assessment of variants of uncertain significance), improve understanding of disease pathogenesis and contribute to development of novel disease biomarkers.

Limitations of LRS

Although LRS has many advantages over current clinical sequencing technologies, it is also associated with significant limitations that currently prevent its ready utilisation and adoption into clinical diagnostic pipelines. LRS typically requires specialised DNA extraction methods to generate high-quality, high molecular weight DNA for sequencing, meaning that historical/stored DNA samples obtained using extraction methods optimised for NGS are not optimal for LRS. LRS data also require specialised bioinformatic pipelines to identify, filter and interpret variants. While LRS bioinformatic tools do exist, they are relatively underdeveloped compared with NGS bioinformatic pipelines and need to be streamlined and scaled for large volume clinical work. The currently limited size of LRS variant databases compared with NGS datasets (eg, gnomAD) also renders interpretation and curation of variants identified by LRS much more challenging. Lastly, the current high cost of LRS (~2–5 times NGS in terms of cost per gigabyte of sequencing data generated)[64] also limits accessibility. However, costs are continually falling and the increased diagnostic yield of LRS compared with NGS will soon reach a threshold where comparable or even superior cost-effectiveness is achieved in certain contexts, particularly for targeted LRS.[65] As a result of these limitations, LRS is currently largely limited to research contexts. As was needed for adoption of NGS into clinical settings, significant foundational work is required to address these barriers before LRS is ready for widespread, routine, standalone use in a clinical diagnostic setting. However, use of targeted LRS as a complement to NGS-based assays has already begun in some clinical diagnostic laboratories.[66]

The relative limitations (and strengths) of each LRS technology (PacBio HiFi vs ONT) and sequencing approach (targeted LRS vs long-read whole-genome sequencing) as well as the properties of the sequencing target (eg, the number, size and structure of the loci of interest) will inform the optimal choice of LRS methodology for a given diagnostic situation. For example, PacBio HiFi LRS may be preferable when high base call accuracy is required (eg, detecting single-base interruptions within repetitive genomic regions); whereas, ONT LRS may be preferable when needing to generate particularly large read lengths (eg, to determine STR length in DM2).

CONCLUSIONS

Although LRS technologies have been under development for more than a decade, reductions in cost, improvements in sequencing accuracy and ongoing maturation of long-read bioinformatic tools have led to increasing utilisation of LRS in both research and clinical settings. LRS offers the ability to accurately sequence long stretches of native DNA molecules with concurrent assessment of variant phasing and epigenetic modifications and is therefore well suited to the challenge of sequencing the diverse range of causative variants seen in the genetic myopathies. The advantages of LRS over NGS technologies have already resulted in discovery and characterisation of novel myopathy genes, as well as previously undetected complex genetic variants, and will facilitate continued advances in this field. As a result, the adoption of LRS into clinical practice has the potential to vastly simplify the diagnostic algorithm for genetic myopathies, reduce diagnostic delay, increase the overall diagnostic rate and improve understanding of disease pathobiology. Just as NGS technologies have revolutionised the field of genetic myopathies compared with the previous Sanger sequencing methods, so too will the increasing use of LRS technologies. However, ongoing development of LRS bioinformatic pipelines, resources and workflows, as well as optimisation of LRS accuracy, throughput and cost-effectiveness, is required for large-scale clinical implementation.

Author affiliations

[1]Neurology Department and Molecular Medicine Laboratory, Concord Repatriation General Hospital, Sydney, New South Wales, Australia
[2]The University of Sydney Faculty of Medicine and Health, Sydney, New South Wales, Australia
[3]Translational Neurogenomics Group, Garvan Institute of Medical Research, Sydney, New South Wales, Australia
[4]Neurodegenerative Service, Prince of Wales Hospital and Neuroscience Research Australia, Sydney, New South Wales, Australia
[5]Clinical Genetics Unit, Royal North Shore Hospital, Sydney, New South Wales, Australia
[6]Neurogenetics Research Group, Kolling Institute, School of Medical Sciences, Faculty of Medicine and Health, University of Sydney and Northern Sydney Local Health District, Sydney, New South Wales, Australia
[7]Department of Neurology, Royal North Shore Hospital, Sydney, New South Wales, Australia
[8]Department of Neurology, Royal Adelaide Hospital, Adelaide, South Australia, Australia
[9]The University of Adelaide School of Medicine, Adelaide, South Australia, Australia
[10]Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia
[11]Molecular Medicine Laboratory, Concord Repatriation General Hospital, Sydney, New South Wales, Australia
[12]Royal Prince Alfred Hospital, Sydney, New South Wales, Australia
[13]Institute of Precision Medicine and Bioinformatics, Sydney Local Health District, Sydney, New South Wales, Australia
[14]Rare Disease Genetics and Functional Genomics, Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia
[15]Centre for Medical Research, The University of Western Australia Faculty of Health and Medical Sciences, Perth, Western Australia, Australia
[16]Northcott Neuroscience Laboratory, ANZAC Research Institute, Sydney Local Health District, Sydney, New South Wales, Australia
[17]Genomic Technologies Lab, Garvan Institute of Medical Research, Sydney, New South Wales, Australia
[18]Faculty of Medicine and Health, University of New South Wales, Sydney, New South Wales, Australia

ORCID iDs
Dennis Yeow http://orcid.org/0000-0001-7534-7007
Laura Ivete Rudaks http://orcid.org/0000-0002-1100-319X

REFERENCES

1. Theadom A, Rodrigues M, Poke G, *et al*. A Nationwide, Population-Based Prevalence Study of Genetic Muscle Disorders. *Neuroepidemiology* 2019;52:128–35.
2. Nicolau S, Milone M, Liewluck T. Guidelines for genetic testing of muscle and neuromuscular junction disorders. *Muscle Nerve* 2021;64:255–69.
3. Kumar KR, Cowley MJ, Davis RL. Next-Generation Sequencing and Emerging Technologies. *Semin Thromb Hemost* 2019;45:661–73.
4. Stence AA, Thomason JG, Pruessner JA, *et al*. Validation of Optical Genome Mapping for the Molecular Diagnosis of Facioscapulohumeral Muscular Dystrophy. *J Mol Diagn* 2021;23:1506–14.
5. Ghaoui R, Cooper ST, Lek M, *et al*. Use of Whole-Exome Sequencing for Diagnosis of Limb-Girdle Muscular Dystrophy: Outcomes and Lessons Learned. *JAMA Neurol* 2015;72:1424–32.
6. Beecroft SJ, Yau KS, Allcock RJN, *et al*. Targeted gene panel use in 2249 neuromuscular patients: the Australasian referral center experience. *Ann Clin Transl Neurol* 2020;7:353–62.
7. Sedlazeck FJ, Rescheneder P, Smolka M, *et al*. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 2018;15:461–8.
8. Stevanovski I, Chintalaphani SR, Gamaarachchi H, *et al*. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci Adv* 2022;8:eabm5386.
9. Maestri S, Maturo MG, Cosentino E, *et al*. A Long-Read Sequencing Approach for Direct Haplotype Phasing in Clinical Settings. *Int J Mol Sci* 2020;21:9177.
10. Lv H, Dao FY, Zhang D, *et al*. Advances in mapping the epigenetic modifications of 5-methylcytosine (5mC). *Biotechnol Bioeng* 2021;118:4204–16.
11. Harvey WT, Ebert P, Ebler J, *et al*. Whole-genome long-read sequencing downsampling and its effect on variant-calling precision and recall. *Genome Res* 2023;33:2029–40.
12. Chintalaphani SR, Pineda SS, Deveson IW, *et al*. An update on the neurological short tandem repeat expansion disorders and the

emergence of long-read sequencing diagnostics. *Acta Neuropathol Commun* 2021;9:98.

13 Iyer SV, Goodwin S, McCombie WR. Leveraging the power of long reads for targeted sequencing. *Genome Res* 2024;34:1701–18.

14 Nurk S, Koren S, Rhie A, *et al*. The complete sequence of a human genome. *Science* 2022;376:44–53.

15 Liao W-W, Asri M, Ebler J, *et al*. A draft human pangenome reference. *Nature New Biol* 2023;617:312–24.

16 Kamsteeg E-J, Kress W, Catalli C, *et al*. Best practice guidelines and recommendations on the molecular diagnosis of myotonic dystrophy types 1 and 2. *Eur J Hum Genet* 2012;20:1203–8.

17 Brais B, Bouchard JP, Xie YG, *et al*. Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat Genet* 1998;18:164–7.

18 Nallamilli B, McCarty A, Kesari A, *et al*. Genetic basis of oculopharyngeal muscular dystrophy: detection of alanine repeats in PABPN1 gene by next generation sequencing. *Mol Genet Metab* 2021;132:S277.

19 Robinson DO, Wills AJ, Hammans SR, *et al*. Oculopharyngeal muscular dystrophy: a point mutation which mimics the effect of the PABPN1 gene triplet repeat expansion mutation. *J Med Genet* 2006;43:e23.

20 Ishiura H, Shibata S, Yoshimura J, *et al*. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat Genet* 2019;51:1222–32.

21 Deng J, Yu J, Li P, *et al*. Expansion of GGC Repeat in GIPC1 Is Associated with Oculopharyngodistal Myopathy. *Am J Hum Genet* 2020;106:793–804.

22 Yu J, Shan J, Yu M, *et al*. The CGG repeat expansion in RILPL1 is associated with oculopharyngodistal myopathy type 4. *Am J Hum Genet* 2022;109:533–41.

23 Cortese A, Beecroft SJ, Facchini S, *et al*. A CCG expansion in ABCD3 causes oculopharyngodistal myopathy in individuals of European ancestry. *Nat Commun* 2024;15:6327.

24 Fazal S, Danzi MC, Xu I, *et al*. RExPRT: a machine learning tool to predict pathogenicity of tandem repeat loci. *Genome Biol* 2024;25:39.

25 Zhang T, Chancellor A, Liem B, *et al*. Neuronal intranuclear inclusion disease in New Zealand: A novel discovery. *J Neurol Sci* 2024;460:122987.

26 Chen Z, Yan Yau W, Jaunmuktane Z, *et al*. Neuronal intranuclear inclusion disease is genetically heterogeneous. *Ann Clin Transl Neurol* 2020;7:1716–25.

27 Podar IV, Gutmann DAP, Harmuth F, *et al*. First case of adult onset neuronal intranuclear inclusion disease with both typical radiological signs and *NOTCH2NLC* repeat expansions in a Caucasian individual. *Euro J of Neurology* 2023;30:2854–8.

28 Rajan-Babu I-S, Dolzhenko E, Eberle MA, *et al*. Sequence composition changes in short tandem repeats: heterogeneity, detection, mechanisms and clinical implications. *Nat Rev Genet* 2024;25:476–99.

29 Cumming SA, Hamilton MJ, Robb Y, *et al*. De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur J Hum Genet* 2018;26:1635–47.

30 Alfano M, De Antoni L, Centofanti F, *et al*. Characterization of full-length *CNBP* expanded alleles in myotonic dystrophy type 2 patients by Cas9-mediated enrichment and nanopore sequencing. *Elife* 2022;11:e80229.

31 Day JW, Ricker K, Jacobsen JF, *et al*. Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum. *Neurology (ECronicon)* 2003;60:657–64.

32 Ruggieri A, Naumenko S, Smith MA, *et al*. Multiomic elucidation of a coding 99-mer repeat-expansion skeletal muscle disease. *Acta Neuropathol* 2020;140:231–5.

33 Ebbert MTW, Farrugia SL, Sens JP, *et al*. Long-read sequencing across the C9orf72 "GGGGCC" repeat expansion: implications for clinical use and genetic discovery efforts in human disease. *Mol Neurodegener* 2018;13:46.

34 Giardina E, Camaño P, Burton-Jones S, *et al*. Best practice guidelines on genetic diagnostics of facioscapulohumeral muscular dystrophy: Update of the 2012 guidelines. *Clin Genet* 2024;106:13–26.

35 Lemmers R, Osborn M, Haaf T, *et al*. D4F104S1 deletion in facioscapulohumeral muscular dystrophy: phenotype, size, and detection. *Neurology (ECronicon)* 2003;61:178–83.

36 Lemmers RJLF, van der Vliet PJ, Blatnik A, *et al*. Chromosome 10q-linked FSHD identifies *DUX4* as principal disease gene. *J Med Genet* 2022;59:180–8.

37 Vasale J, Boyar F, Jocson M, *et al*. Molecular combing compared to Southern blot for measuring D4Z4 contractions in FSHD. *Neuromuscul Disord* 2015;25:945–51.

38 Huang M, Zhang Q, Jiao J, *et al*. Comprehensive genetic analysis of facioscapulohumeral muscular dystrophy by Nanopore long-read whole-genome sequencing. *J Transl Med* 2024;22:451.

39 Lemmers RJLF, Butterfield R, van der Vliet PJ, *et al*. Autosomal dominant *in cis* D4Z4 repeat array duplication alleles in facioscapulohumeral dystrophy . *Brain (Bacau)* 2024;147:414–26.

40 Giugliano T, Savarese M, Garofalo A, *et al*. Copy Number Variants Account for a Tiny Fraction of Undiagnosed Myopathic Patients. *Genes (Basel)* 2018;9:524.

41 Tuffery-Giraud S, Béroud C, Leturcq F, *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;30:934–45.

42 Oliveira J, Gonçalves A, Oliveira ME, *et al*. Reviewing Large LAMA2 Deletions and Duplications in Congenital Muscular Dystrophy Patients. *J Neuromuscul Dis* 2014;1:169–79.

43 Minoche AE, Lundie B, Peters GB, *et al*. ClinSV: clinical grade structural and copy number variant detection from whole genome sequencing data. *Genome Med* 2021;13:32.

44 Bruels CC, Littel HR, Daugherty AL, *et al*. Diagnostic capabilities of nanopore long-read sequencing in muscular dystrophy. *Ann Clin Transl Neurol* 2022;9:1302–9.

45 Zeviani M, Moraes CT, DiMauro S, *et al*. Deletions of mitochondrial DNA in Kearns-Sayre syndrome. *Neurology (ECronicon)* 1988;38:1339–46.

46 Davis RL, Kumar KR, Puttick C, *et al*. Use of Whole-Genome Sequencing for Mitochondrial Disease Diagnosis. *Neurology (ECronicon)* 2022;99:e730–42.

47 Frascarelli C, Zanetti N, Nasca A, *et al*. Nanopore long-read next-generation sequencing for detection of mitochondrial DNA large-scale deletions. *Front Genet* 2023;14:1089956.

48 Kobayashi K, Nakahori Y, Miyake M, *et al*. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature New Biol* 1998;394:388–92.

49 Colombo R, Bignamini AA, Carobene A, *et al*. Age and origin of the FCMD 3'-untranslated-region retrotransposal insertion mutation causing Fukuyama-type congenital muscular dystrophy in the Japanese population. *Hum Genet* 2000;107:559–67.

50 Bychkov I, Baydakova G, Filatova A, *et al*. Complex Transposon Insertion as a Novel Cause of Pompe Disease. *Int J Mol Sci* 2021;22:10887.

51 Akman HO, Davidzon G, Tanji K, *et al*. Neutral lipid storage disease with subclinical myopathy due to a retrotransposal insertion in the PNPLA2 gene. *Neuromuscul Disord* 2010;20:397–402.

52 Zhou W, Emery SB, Flasch DA, *et al*. Identification and characterization of occult human-specific LINE-1 insertions using long-read sequencing technology. *Nucleic Acids Res* 2020;48:1146–63.

53 Xie Z, Liu C, Lu Y, *et al*. Exonization of a deep intronic long interspersed nuclear element in Becker muscular dystrophy. *Front Genet* 2022;13:979732.

54 Perrin A, Van Goethem C, Thèze C, *et al*. Long-Reads Sequencing Strategy to Localize Variants in TTN Repeated Domains. *J Mol Diagn* 2022;24:719–26.

55 van der Ven PFM, Odgerel Z, Fürst DO, *et al*. Dominant-negative effects of a novel mutation in the filamin myopathy. *Neurology (ECronicon)* 2010;75:2137–8.

56 Grosz BR, Stevanovski I, Negri S, *et al*. Long read sequencing overcomes challenges in the diagnosis of SORD neuropathy. *J Peripher Nerv Syst* 2022;27:120–4.

57 Gouil Q, Keniry A. Latest techniques to study DNA methylation. *Essays Biochem* 2019;63:639–48.

58 Yu J, Deng J, Guo X, *et al*. The GGC repeat expansion in *NOTCH2NLC* is associated with oculopharyngodistal myopathy type 3 . *Brain (Bacau)* 2021;144:1819–32.

59 Tatton-Brown K, Seal S, Ruark E, *et al*. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat Genet* 2014;46:385–8.

60 Aref-Eshghi E, Kerkhof J, Pedro VP, *et al*. Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *Am J Hum Genet* 2020;106:356–70.

61 Ghaoui R, Ha TT, Kerkhof J, *et al*. Expanding the phenotype of DNMT3A as a cause a congenital myopathy with rhabdomyolysis. *Neuromuscul Disord* 2023;33:484–9.

62 Mitsuhashi S, Nakagawa S, Sasaki-Honda M, *et al*. Nanopore direct RNA sequencing detects DUX4-activated repeats

and isoforms in human muscle cells. *Hum Mol Genet* 2021;30:552–63.

63 Motone K, Kontogiorgos-Heintz D, Wee J, *et al*. Multi-pass, single-molecule nanopore reading of long protein strands. *Nature New Biol* 2024;633:662–9.

64 Espinosa E, Bautista R, Larrosa R, *et al*. Advancements in long-read genome sequencing technologies and algorithms. *Genomics* 2024;116.

65 Wang Y, Zhu G, Li D, *et al*. High clinical utility of long-read sequencing for precise diagnosis of congenital adrenal hyperplasia in 322 probands. *Hum Genomics* 2025;19:3.

66 Kaplun L, Krautz-Peterson G, Neerman N, *et al*. ONT long-read WGS for variant discovery and orthogonal confirmation of short read WGS derived genetic variants in clinical genetic testing. *Front Genet* 2023;14:1145285.