Applying genomic and transcriptomic advances to mitochondrial medicine

William L. Macken[1], Jana Vadrovcova[1], Michael G. Hanna[1], Robert D.S. Pitceathly[1,†]

[1]Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology and The National Hospital for Neurology and Neurosurgery, London, United Kingdom

[†]email: r.pitceathly@ucl.ac.uk

Abstract

Next generation sequencing (NGS) has increased our understanding of the molecular basis of many primary mitochondrial diseases (PMDs). Despite this progress, many patients with suspected PMD remain without a genetic diagnosis, which limits their access to in-depth genetic counselling, reproductive options and clinical trials, in addition to hampering our efforts to understand the underlying disease mechanisms. Although a considerable improvement over their predecessors, current methods for sequencing the mitochondrial and nuclear genomes have important limitations, and molecular diagnostic techniques are often manual and time consuming. However, recent advances offer realistic solutions to these challenges. In this Review, we discuss the current genetic testing approach for PMDs and the opportunities that exist for increased use of whole-genome NGS of nuclear and mitochondrial DNA (mtDNA) in the clinical environment. We consider the possible role for long-read approaches in sequencing of mtDNA and in the identification of novel nuclear genomic causes of PMDs. We examine the expanding applications of RNA sequencing, including the detection of cryptic variants that affect splicing and gene expression, as well as the interpretation of rare and novel mitochondrial transfer RNA variants.

[H1] Introduction

Primary mitochondrial diseases (PMD) comprise a group of rare genetic conditions characterised by impaired mitochondrial oxidative phosphorylation (OXPHOS), leading to energy deficiency and organ dysfunction. These diseases are defined by the identification of DNA variants that are known to cause dysfunction of OXPHOS or to disturb mitochondrial structure and function in another way[1]. Over 300 genes across the mitochondrial and nuclear genomes have been associated with PMD[2]. Many of these disorders are ultra-rare at the molecular level[3], but collectively the prevalence of PMDs in adults is approximately 1 in 4,300 (1 in 11,500 in children <16 years), establishing them as one of the most common groups of inherited neurological diseases[3,4].

Mitochondria are a central hub for metabolism in almost all cell types. Consequently, the clinical manifestations of impaired mitochondrial function have the potential to be extremely heterogenous, posing substantial diagnostic challenges. Symptoms typically associated with PMD include psychomotor regression, dystonia, failure-to-thrive, bone marrow dysfunction, muscle fatigue, exercise intolerance, myopathy, ophthalmoplegia, migraine, encephalopathy and stroke-like episodes, seizures, optic neuropathy, cardiomyopathy and cardiac conduction defects, deafness, and endocrinopathies (in particular diabetes)(Fig. 1)[5] The clinical manifestations of PMDs are often multisystemic, particularly in in children. However, neurological complications, including seizures, encephalopthy, stroke-like episodes, peripheral neuraopthy, and myopathy predominate in adults. Consequently, adult neurologists are usually the primary care providers for this group of patients. Historically, PMDs have been considered to have syndromic presentations[6]. However, single system disease (for example, skeletal myopathy or peripheral neuropathy) and overlapping symptoms that extend beyond the 'classical' recognised syndromes are also common[7,8]. Although the vast majority of PMDs currently lack disease-modifying interventions, an expanding portfolio of pharmacological and genetic treatments are at the preclinical and early clinical stages of development[9].

[H1] Diagnostic challenges

The investigation of PMD presents unique challenges when compared with other neurogenetic disorders (Table 1). These challenges arise principally because mitochondria contain their own genome, which consists of mitochondrial DNA (mtDNA) and is distinct from nuclear DNA (nDNA). mtDNA is a short double-stranded circular molecule, approximately 16.6kb in length[10]. Multiple copies of mtDNA are present in each mitochondrion and the number of mitochondria per cell varies widely among cell types[11,12] Of the mtDNA genes, 13 encode mitochondrial proteins and 24 encode non-coding RNA[10]. mtDNA contains no large introns or intergenic sequences, instead most protein-coding genes are separated by genes that encode transfer RNA (tRNA)[10]. The entire mitochondrial genome is transcribed as a polycistronic transcript **[G]**, which is then divided into its constituent parts before translation into a protein or modification into functioning ribosomal RNA (rRNA) or tRNA[13].

mtDNA is exclusively transmitted via the ovum, thus mtDNA-related PMDs are maternally inherited[14]; however, >1,000 nDNA (nDNA) genes (inherited from both parents) also encode proteins required for mitochondrial function[15]. Consequently, PMD can result from pathogenic genetic variants within either mtDNA or nDNA. Indeed, a large and growing number of nDNA genes have been implicated in PMDs[16]. The situation is further complicated by the presence of a substantial subgroup of nuclear genes that control mtDNA maintenance**[G]**. Mutations in these maintenance genes can trigger downstream replication errors across the mitochondrial genome. These errors include point mutations and/or polyclonal deletions in the mtDNA and, in some instances, depletion of mtDNA copy number[17].

A further consideration in the molecular diagnosis of PMD is the concept of heteroplasmy, that is, the presence of mixed populations of mitochondria — some carrying mutated mtDNA and others carrying non-mutated mtDNA — within a single cell, tissue or organism. In tissues with a high cell turnover (for example, blood) the selective pressure to remove faulty mitochondria is especially high, which can

result in lower levels, or complete absence, of mutated mtDNA present in these tissue types [18]. Similarly, mtDNA rearrangements are less reliably detected in the blood of adults than in the blood of children[19]. These characteristics often mandate sampling of a post-mitotic tissue **[G]** (for example, skeletal muscle) to fully exclude the presence of a pathogenic mtDNA variant. One particularly important implication of mtDNA heteroplasmy is the necessity for deep sequencing **[G]** of mtDNA to ensure that even low levels of heteroplasmy are detected[20]. Typically, a threshold effect exists, whereby a heteroplasmic variant needs to be present in a certain proportion of mtDNA copies in a tissue or cell before a phenotypic or biochemical effect manifests[21]

Although the threshold for disease manifestations is relatively high for most mtDNA variants (>60% of mtDNA copies in a cell), lower levels can result in clinical phenotypes. For instance, pancreatic m.3243A>G levels of well below 60% have been reported in people with mitochondrial diabetes.[23,24] Furthermore, a surprisingly high proportion of the general population (1 in 200) harbour potentially pathogenic mtDNA variants, albeit a very low levels[21,22].

Tissue heteroplasmy warrants careful consideration during PMD diagnosis. For example, an mtDNA variant that is present at low levels in blood, fibroblast or muscle might not meet the threshold to cause a biochemical or histological abnormality in those tissues, giving the false impression that the variant is not disease-causing. However, the same variant might be present at high levels in a difficult-to-access tissue, such as brain or cardiac muscle, where most of the disease burden might lie. In such cases, expert consideration of the variant by a specialist clinical scientist, consideration of the phenotype by a specialist clinician and, where possible, inclusion of research-based functional studies

(such as cell hybrids) is necessary (Supplementary Table 1). A further curiosity of the mitochondrial genome is the high proportion of tRNA genes it contains. Of the genes contained within mtDNA, 60% encode tRNA molecules[10] and confirming the pathogenic effects of rare and novel variants in these genes presents unique challenges.

[H1] Approaches to molecular diagnosis

Next-generation sequencing **[G]** (NGS) of nuclear genes, mainly with targeted gene panels and whole exome sequencing (WES), has become a mainstay of PMD research and has led to a substantial increase in the number of molecular diagnoses achieved[25–30]. This technology offers accurate sequencing at very competitive costs and is now widely used in clinical laboratories. NGS is also now routinely applied to sequencing of the entire mitochondrial genome[31–33]. However, short-reads **[G]** have inherent limitations for the identification and confirmation of disease-causing variants in PMD. These limitations include difficulty detecting structural variants (SVs) **[G]** and short tandem-repeat variants, the inability to phase **[G]** variants and recognise epigenetic **[G]** changes, and the absence of transcript data, all of which could hamper variant discovery and interpretation.

Exome, genome and transcriptome studies in PMD have reported diagnostic rates of 10–67%[27–30,34,35], but these numbers belie a complex picture. PMDs are a heterogenous group, encompassing well-researched conditions with established genotype–phenotype correlations, childhood-onset diseases amenable to whole exome trio **[G]** resolution, and adult-onset disorders that overlap with other neurogenetic, metabolic and acquired neurological disorders. In our experience, the rate of genetic diagnosis in adults with suspected PMD following routine molecular testing — that is, mtDNA sequencing, large-scale rearrangement analysis, and mitochondrial nuclear gene panels — remains low (~20%), even after extended candidate gene-based whole genome sequencing (WGS) (personal communication). This observation emphasises the importance of combining emerging genomic and

transcriptomic strategies to identify cryptic molecular causes and PMD mimics, thus facilitating the next diagnostic uplift for this group of disorders.

In this article, we review the current stepwise approach to investigating mtDNA-related PMD (Fig. 2) and discuss the use of WGS in the clinical setting, which is on the cusp of widespread introduction. We examine the utility of long-read sequencing when studying mtDNA and consider its potential application in deciphering unresolved nDNA causes of PMD, including SVs, short tandem repeat variants, and epigenetic changes, none of which have yet been established as causes of PMD but are increasingly recognized as causative in other neurogenetic conditions[36]. Finally, we discuss the use of NGS-based transcriptomics (RNA-seq) in identifying pathogenic variants and helping confirm causality, including for mitochondrial tRNA (mt-tRNA)-related disorders. Importantly, unlike NGS gene panels and WES, in depth interpretation of WGS (including long-reads) and RNAseq data is currently beyond the resources of most clinical service laboratories and is only feasible through partnership with and support from research groups.

[H1] Current approach to genetic testing

Clinical laboratories currently employ multiple techniques to test mtDNA from patients with suspected PMD and often also conduct parallel nDNA sequencing studies (Fig. 2)**.** Some PMD phenotypes (for example, Leber hereditary optic neuropathy) are well-defined and have targeted testing approaches[37]. However, the majority of PMDs cause substantially overlapping phenotypes and a broad, agnostic approach is necessary for their diagnosis.

In the first instance, our specialist mitochondrial diagnostic centre excludes the most common mtDNA pathogenic point mutations (for example, m.3243A>G, m.8344A>G, and m.8993T>G/C; table 2) using PCR and restriction fragment length polymorphism **[G]** testing. Although genotype–phenotype correlations exist for these mutations, they are often non-specific; therefore, simultaneous analysis of

multiple loci is performed in patients with a 'mitochondrial' presentation. If none of the most common mtDNA mutations are present, or if the phenotype is suggestive of a single mtDNA deletion, we move on to mtDNA large-scale rearrangement **[G]** analysis (long-range PCR **[G]** and Southern blotting, or a quantitative PCR technique), sequencing of the entire mitochondrial genome and, if appropriate, mtDNA copy number analysis (real-time quantitative PCR or array CGH)[38–41]. These investigations are often undertaken sequentially, although some centres now proceed directly to NGS of the entire mitochondrial genome for single nucleotide variants (SNVs) and single mtDNA deletions, without excluding common mutations first. However, irrespective of the step-wise approach applied to mtDNA analysis, the diagnostic odyssey of PMD is often protracted[42].

Genetic testing for PMD is initially undertaken in blood and uroepithelial cells with the caveat that, if an mtDNA mutation is not detected, a muscle biopsy might be required for further genetic analysis. Muscle tissue is examined for histological and histochemical changes suggestive of mitochondrial dysfunction, such as the presence ragged-red and cytochrome *c* oxidase (COX) -negative fibres[43]. The activity of mitochondrial respiratory chain enzymes is also measured, in addition to the re-analysis of mtDNA for sequence changes and large-scale rearrangements. In presentations such as chronic progressive external ophthalmoplegia, Kearns–Sayre syndrome, and primary mitochondrial myopathy, the genetic basis of disease is frequently only detectable in muscle mtDNA[19]. If appropriate, interrogation of nDNA for variants in genes involved with mtDNA maintenance, or broader testing including nuclear genes encoding proteins required for OXPHOS, can be performed in parallel to these mtDNA studies via gene panels. The Genomics England PanelApp contains a useful virtual nuclear gene panel for possible mitochondrial disorders.

[H2] Mitochondrial genome sequencing

Deep NGS of the mitochondrial genome is now in wide clinical use and is considered the gold standard approach for sequencing mtDNA[20,31–33]. This method enables very deep sequencing and thus has major

advantages over its predecessors in accurate identification and quantification of low-level heteroplasmy[32]. This accurate analysis of heteroplasmy is important for variant classification[44], for example, the presence of a mutation at very low levels in an unaffected mother and at high levels in an affected child supports a conclusion of pathogenicity. mtDNA can also be analysed with WES, either by use of off-target WES data, or by specifically capturing mtDNA[45–47]. However, in clinical practice, diagnostic laboratories usually sequence the mitochondrial genome and then use separate gene panels to probe for variants in nuclear DNA. The utility of mtDNA sequencing is generally limited to SNVs and large-scale single mtDNA deletions. Although, in our experience, other rearrangements, such as multiple mtDNA deletions and low-level heteroplasmic rearrangements, are detectable in NGS data, parsing these molecular changes from coverage **[G]** issues with sufficient certainty to issue a diagnostic, clinically actionable report, remains difficult. Consequently, rearrangement analysis continues to rely upon other techniques, such as Southern blotting or digital droplet PCR. This situation is suboptimal given that, these techniques do not provide high-resolution assessment of deletion breakpoints.

[H3] Mitochondrial DNA haplogroups

By identifying patterns in mitochondrial polymorphisms, individuals can be categorised into 'mitochondrial haplogroups' that are characterised by shared variants from a common ancestor[48]. Mitochondrial haplogroups have clinical relevance, for example, if a rare variant is identified, but is known to contribute to that individual's haplogroup signature, it is unlikely to be the underlying cause of the patient's disease and can be disregarded from further analysis. The detection of divergent haplotypes in different samples from the same individual can also be used to identify sample contamination. Finally, mitochondrial haplogroups can potentially influence expression of specific mtDNA variants, for example, Haplogroup J (Western Eurasian) is associated with an increased penetrance of some pathogenic variants that cause Leber Hereditary Optic Neuropathy[49,50].

[H3] Enrichment of mitochondrial DNA for targeted sequencing

For targeted mitochondrial sequencing, the mtDNA present in a sample is enriched over nDNA before sequencing[51]. The most common enrichement approach involves the use of PCR to selectively amplify mtDNA, but not nDNA. However, PCR can cause artefactual variants especially when small templates are used[52,53]. Among PCR techniques, long-range PCR performs best, ensuring uniform coverage and facilitating identification of heteroplasmic variants[32,54]. nDNA is known to incorporate nuclear copies of mtDNA as it evolves over time[55,56] and these near-identical nuclear sequences can be co-amplified with actual mtDNA leading to spurious identification of heteroplasmic variants during analysis[57]. The presence of nuclear copies of mtDNA underlines the importance of effective enrichment of mtDNA over nDNA. Non-PCR based amplification techniques, such as rolling circle amplification, organelle selection and enzymatic degradation of nDNA are available[52,58–64] However, further discussion of these techniques is beyond the scope of this article.

[H3] Obstacles to streamlined genetic testing for PMD

The use of multiple techniques for what would ideally be a single investigation is time consuming and expensive. Many of these processes are technically challenging, require large quantities of DNA, and are subject to failure for technical reasons. Also, current strategies are reliant on PCR, which has several important biases, including preferential amplification of short fragments (which might over-represent heteroplasmic molecules with large deletions) and poor performance at loci with extreme GC content **[G]**[65,66]. Furthermore, because traditional PCR involves imperfect exponential amplification, the PCR products are not directly proportional to the input[67]. This disparity affects the accuracy of mtDNA quantification, rendering the method unsuitable for the measurement of mtDNA copy number, which is an important indicator of mtDNA depletion syndromes[68].

The optimal method for genetic diagnosis of PMD would combine the identification of pathogenic mtDNA point mutations and/or large-scale rearrangements with an accurate quantification of mutant

heteroplasmy. More broadly, for patients with a high probability of an nDNA or blood-identifiable mtDNA variant being the cause for their disease, concurrent analysis of nDNA and mtDNA through WGS is potentially the most effective approach for identifying the underlying causative mutation.

[H2] WGS

Given the improved quality and falling cost of WGS, some experts have suggested the use of a genome-first method to improve diagnostic efficiency[69]. The high coverage of mtDNA by WGS facilitates the assessment of heteroplasmic variants while also enabling the identification of molecular changes in nuclear mitochondrial genes. However, not all WGS approaches achieve sufficient read depth for diagnostic purposes; our laboratory recommends a minimum depth of 500X in order to exclude the presence of low-level heteroplasmy in clinical samples. WGS also has the advantage of revealing non-mitochondrial phenocopies **[G]** of PMD.

WGS for PMD can be performed without a PCR because mtDNA naturally has a higher copy number than nDNA, which enables identification of heteroplasmy, for example, leukocytes contain only two copies of nDNA (that is, maternal and paternal homologous chromosomes) but they possess 100s of copies of mtDNA. Ideally, this process would also avoid the need for invasive tests such as muscle biopsy, although it would only be suitable when a blood-identifiable mtDNA variant is suspected. Careful patient selection would maximise the rate of detection, for example, paediatric patients are generally more likely to harbour nDNA mutations and retain higher level of heteroplasmic mtDNA variants in blood than adult patients[70]. In a study published in 2020, investigators performed WGS on blood samples from well-phenotyped parent–child trios who had clinical presentations consistent with PMD and detected pathogenic variants in 68% of probands[34]. These variants included mutations (nDNA and mtDNA) in known PMD-associated genes as well in previously non-morbid genes. Some of the identified variants in known disease genes had not previously been linked to PMD presentations. Thus, the findings from WGS might challenge current dogmas of what defines a PMD .

[H3] Phenocopies and 'double trouble'

PMD phenotypes are often heterogenous and multi-systemic. However, in some instances the presence of multiple pathological processes, unrelated to PMD, can result in a complex phenotype that mimics a PMD. For example, a patient can have muscle disease resulting from a pathogenic variant in a myopathy gene and hearing impairment owing to an unrelated mutation in a deafness gene. In one study, this co-occurrence, sometimes informally referred to as 'double trouble', was identified in 4.9% of WES-solved rare disease cases[71]. The genes involved in these PMD mimics can be unrelated to mitochondrial function, so two non-mitochondrial gene mutations can result in a "compound phenocopy" that resembles PMD. In our experience, WGS using multiple virtual gene panels helps identify these PMD mimics, thereby underlining the benefit of adopting a broad approach to genetic testing. Such patients might be labelled as having "possible mitochondrial disease" before a molecular diagnosis is identified. However, this terminology could contribute towards inaccurate genetic counselling, distress and misdirection, or cessation of the diagnostic journey[72]. The term "diagnosis uncertain" qualified with the established abnormalities (for example, "complex I deficiency", or "variant of uncertain significance in *POLG*") has been recommended[72], although some clinicians continue to favour terms such as "suspected PMD" or "clinically-defined PMD" to indicate strong clinical and/or biochemical evidence of PMD, at least in scientific communications.

[H3] Digenic-digenomic inheritance

The influence the nDNA background can exert on the penetrance of mtDNA variants adds further weight to the argument for use of WGS in the clinical setting [73] For example, the mtDNA variant m.1630A>G in *MT-TV*, which encodes a mitochondrial tRNA molecule responsible for carrying the amino acid valine (mt-tRNA^{Val}), was reported to be present at high levels in the fibroblasts of an unaffected mother and an affected child.[73] The child presented at 15 years with a MELAS phenotype (stroke-like episode, seizures and a raised plasma lactate). Exome sequencing revealed a second nuclear variant in the *VARS2* gene present in the child, but absent in the mother. This truncating *VARS2*

variant resulted in loss of a protein domain that charges valine to the mt-tRNA^{Val} molecule, suggesting that the interplay between both variants resulted in pathogenicity. Alhough the study did not use WGS, this digenic-digenomic phenomenon underlines the potential utility of a WGS-first strategy.

Another perplexing phenomenon is the variable tissue expression and penetrance of homoplasmic (present at 100% mutant load) mtDNA pathogenic variants. One example of this is the m.14674T>C mutation in *MT-TE,* which encodes the mitochondrial tRNA molecule responsible for carrying the amino acid glutamine (mt-tRNA^{Glu}). m.14674T>C is homoplasmic, but only causes reversible infantile respiratory chain deficiency in a fraction of carriers[74,75]. WES in healthy and affected m.14674T>C carriers found that the majority of affected individuals also harboured heterozygous variants in nuclear genes known to interact with mt-tRNA^{Glu}. For example, some individuals carried deleterious mutations in *EARS2*, which encodes a protein that aminoacylates the 3' end of the tRNA molecule[76].

Identification of these nuclear modifier variants is extremely challenging and requires detailed analysis of a clearly defined cohort or family to identify recurrent variants that co-occur with the penetrant phenotype. At present, such an undertaking would be impractical within the diagnostic setting and given the huge numbers of candidate variants in WGS data, it would be unfeasible to undertake at scale, even in research environments.

Although ancillary mitochondrial investigations are useful in ruling in PMD, they have a limited capacity to rule out non-PMD, and so cannot always distinguish between true-PMD and phenocopies of the disorders. For example, a patient with myalgia and weakness and a muscle biopsy that shows COX-negative fibres, borderline low complex IV activity, and mtDNA deletions might have PMD, or might instead have secondary mitochondrial dysfunction such as that resulting from advanced age, inactivity or an overlapping disease (for example, such as inclusion body myositis)[77].

[H3] Interpreting variants in WGS

As with exome sequencing, the prioritisation and interpretation of the large number of variants identified during WGS is a substantial challenge to its widescale adoption in the clinical sector. Most centres follow the American College of Medical Genetics and Genomics (ACMG) guidance for variant interpretation, which involves application of a set of criteria to assess the pathogenicity of a candidate variant[78]. Detailed discussion of these criteria is outside the scope of this Review; however, a number of peculiarities specific to PMD warrant consideration. For example, the ACMG criterion "Patient's phenotype or family history is highly specific for a disease with a single genetic etiology" (PP4)[78] might not be applicable in disorders where variants in a number of genes cause a similar phenotype, as is typical for PMD. The matrilineal pattern of inheritance and the influence of mtDNA heteroplasmy makes it difficult to apply *de novo* and segregation criteria when interpreting variants associated with PMD. Furthermore, mtDNA genes are enriched for non-protein-coding (tRNA and rRNA) genes that are challenging to interpret using ACMG criteria. In particular, the ACMG computational predictive data criteria, which mainly relate to loss-of-function and missense predictions in proteins, do not readily pertain to tRNA and rRNA genes, although guidelines for the assessment of tRNA variants within the mitochondrial genome have been developed[44]. Given the substantial complexities and subtle nuances that arise from molecular changes in the mitochondrial genome, the interpretation and reporting of mtDNA variants should be undertaken in specialist centres. Supplementary table 1 provides a summary of important resources and considerations for interpreting mtDNA and nDNA variants in PMD.

[H2] Large-scale introduction of WGS

In several parts of the world, WGS is on the cusp of widespread introduction as a diagnostic tool in mainstream medicine and the technique has an expanding role in infectious disease medicine, public health, and pharmacogenomics.[79–81] In the UK, the '100,000 Genomes Project' has bridged the gap between the clinical, research and even commercial sectors. It has demonstrated that the large-scale

generation and interpretation of WGS data is feasible in the public sector, and a national genomic medicine service is being created on the basis of the framework and infrastructure established during the project[82]. Clinicians, not least those within the neurology community, must be at the forefront of this process to ensure the accurate interpretation of genetic findings in a manner that is nuanced, clinically meaningful and beneficial to patients.

[H3] Barriers to blood-based WGS in mitochondrial medicine

Although the increased use of blood-based WGS as a 'catch-all' investigation might improve diagnostics in PMD, several challenges remain. First, a subgroup of PMDs lack blood-identifiable variants. On the basis of data from two large studies[2,3] Lucy Raymond and colleagues estimated the rate of solely muscle-identifiable mtDNA mutations to be 11.5% across paediatric and adult patients[60]. However, in adult neuromuscular clinics, where late-onset myopathy presentations are common, muscle-identifiable mutations are likely to be substantially over-represented. Second, whether identified in blood or muscle DNA, novel variants must have sufficient functional evidence to validate their pathogenicity. Unfortunately, as OXPHOS defects are not usually detectable in blood, this validation frequently requires the biopsy of an affected post-mitotic tissue, such as skeletal muscle or liver, to analyse mitochondrial respiratory chain enzyme activity through histochemical staining or spectrophotometric analysis. Thus, invasive tests may not be avoidable in all cases of suspected PMD. Finally, the use of WGS data to study rearrangements and depletion of mtDNA is difficult. Advances have been made in bioinformatic methods to identify SVs and copy number variants in WGS data[83,84]; however, further progress is required before these techniques can be adopted within accredited clinical genetic laboratories.

[H1] Research molecular techniques

WES and WGS are now embedded within clinical genetic laboratories. In comparison, the relative value of integrating long-read and RNA sequencing in the diagnostic setting has been, until recently,

less clear. However, given the considerable benefits they confer when interpreting WES and WGS data, we believe their application is inevitable.

[H2] Long-read sequencing

The third generation of sequencing technologies can produce reads of >10kb (the short reads used in NGS are usually ~150bp in length), enabling the entire mitochondrial genome to be sequenced in one read[66]. Single-molecule real-time sequencing (SMRT, Pacific Biosciences)[85,86], and nanopore sequencing (Oxford Nanopore Technologies) are currently leading the field in long-read sequencing platforms (Fig. 3). In short, SMRT sequencing uses sequencing-by-synthesis of a long circularised DNA molecule[85]. During synthesis, complementary labelled nucleotides are added, one by one, to the circularized DNA by a polymerase. Each nucleotide emits characteristic fluorescence when added, which is recorded in real-time, enabling sequencing. This method can also detect epigenetic modifications, as they alter the timing of the addition of nucleotides. In nanopore sequencing, a single-stranded DNA molecule is fed through a pore base by base, perturbing a current as it passes[87]. Specific bases cause characteristic alterations in the current and epigenetic modifications are also detectable. Compared with NGS, both techniques are less accurate, produce larger volumes of data, and will require further optimisation of bioinformatic tools for data analysis[88]. Consequently, they have not yet been widely introduced into clinical practice but, given the progress made to date, they are likely to be applied in the diagnostic setting in the future.

[H3] Long-read sequencing of mitochondrial DNA

Long-read sequencing of mtDNA is in its infancy but it has the potential to build on NGS as an 'all-in-one' solution to mtDNA genetic testing. It will enable the identification of point mutations and large-scale rearrangements through targeted, PCR-free sequencing and *de novo* assembly **[G]** of the mitochondrial genome.

SMRT sequencing has been used to sequence mtDNA from a variety of preparations, that is, multi-amplicon PCR, 2 amplicon long-range PCR and mtDNA enriched without amplification[64,89,90]. The latter was performed by Pacific Biosciences and has not been published in peer-reviewed form; however, the data were presented at the Advances in Genome Biology and Technology conference in 2018 and a copy of the poster is available on the PacBio website.[90] According to the poster presentation, the authors achieved long reads, spanning the entire molecule, on a small amount of DNA (150 ng) and successfully identified synthesised heteroplasmic variants. However, they acknowledged that further optimization of enrichment and library preparation is required to increase the proportion of long DNA fragments. SMRTseq has also been used in other disease states to parse pseudogenes from their coding counterparts, underlining its potential utility in eliminating nuclear copies of mtDNA during analysis[91,92].

Recently, single-pass SMRT sequencing and NGS have been used in combination to genotype mtDNA mutations[93]. Despite relatively high rates of random small indel and point mutation errors in this study, SMRT sequencing showed a similar accuracy to short-read NGS in identifying mtDNA variants. Unfortunately, the method chosen did not have sufficient read-depth to detect heteroplasmic variants within long-reads. Read-depth is important when distinguishing actual heteroplasmic variants from random errors. Encouragingly, a new technique, known as circular consensus sequencing, has substantially improved the accuracy of SMRTseq[94,95]. Circular consensus sequencing involves multiple passes of the polymerase on optimised DNA, generating high fidelity ('hifi') reads with improved accuracy, even in challenging repetitive areas of the genome, and reducing the computational resources needed to process the data.

A number of protocols have been developed that enable the sequencing of entire mitochondrial genomes with nanopore technology, with or without mtDNA enrichment steps[96,97]. Nanopore approaches promise improved resolution of homopolymeric regions **[G]** of mtDNA and are less work-

intensive than NGS approaches. An initial attempt to use nanopore sequencing in the clinical space (for identification of deletions) yielded promising results[66]. The technology (used on long-range PCR-amplified DNA) successfully identified deletions and achieved superior coverage when compared with NGS methods. Importantly, in two patients long-reads identified deletions that had previously been missed by short-read sequencing. The rate of false positives was too high to identify heteroplasmic SNVs, and bioinformatic optimisation for identifying deletions was difficult for heteroplasmic rearrangements; however, future technological developments might address these issue.

Interestingly, as the entire mtDNA molecule is sequenced in one step, long-read sequencing will enable heteroplasmic variants to be resolved to specific molecules. Additionally, although mitochondrial haplogroups can be constructed from short reads, long reads could have a future role in simplified mitochondrial haplogroup designation.

[H3] Long-read sequencing of nuclear DNA

Third-generation approaches are particularly promising for the sequencing of nDNA. Long-reads from both SMRT and nanopore technologies enable *de novo* assembly and resolve challenging areas of the genome [98–100]. For clinical diagnostics, much of the promise of these technologies lies in their ability to identify repeat **expansions** and SVs, which is computationally demanding with short reads, and to phase variants without parental samples (Fig 4.).

[H3] Phasing difficulties

Understanding the parental origin of variants is important, especially for establishing the pathogenicity of compound heterozygous mutations. Although parental samples can be used to phase variants, it is often difficult to obtain such samples in adults with PMD, given their late presentation to the clinic. Reads detected by NGS are typically too short to link a variant of interest with informative loci and hence designate chromosomal phase; however, long reads are likely to surmount this issue.

Additionally, nanopore long-reads have recently been used to resolve nuclear copies of mtDNA to their nuclear (chromosomal) location and out rule the possibility of patrilineal inheritance of mtDNA[14].

[H3] Structural variants

Breakpoints of SVs are often located within repetitive areas of the genome[101]. Mapping of short-reads within these regions is problematic — when reads fall within an area of non-unique repeats, assembly tools struggle to align them to their specific loci. Where contigs **[G]** of the same repetitive sequence reoccur within a stretch of DNA it is often difficult to align the intervening non-repetitive contigs; hence, alignment or *de novo* assembly will be impaired and pathogenic SVs might be overlooked. In addition to affecting coding DNA, SVs can affect gene expression, underlining the importance of identifying these variants in the clinical setting where they might explain as yet unsolvable monogenic diseases[102]. Although NGS and microarray technologies do identify SVs, they can miss some SVs and NGS CNV analysis generates a large number of false positives. However, use of long-read technologies has been shown to improve the detection of SVs when complemented by NGS[103,104]. SVs have not yet been identified as a cause of PMD, but the introduction of these technologies into the clinical sphere could challenge this finding, particularly given the substantial amount of PMDs that remain genetically undiagnosed following WGS .

[H3] Tandem repeats

Identidying short tandem repeat (STR) expansions with NGS short-read technologies is very challenging. Bioinformatic tools can help address these challenges, but they have not been widely adopted in clinical practice and require confirmatory testing[105,106]. At present, repeat expansions have not been implicated in PMD but, given the rarity of specific PMD entities and the difficulty in identifying these mutations, this is perhaps unsurprising. Long-read technologies that can sequence through repetitive segments are likely to revolutionise diagnostics and gene discovery in STR-related diseases. Several approaches have been developed that use long-read technologies to identify

tandem-repeat variants on a genome-wide basis[107–109], and these long reads have been used to identify known pathogenic STR variants and discover novel tandem repeat causes of disease[110–114].

[H3] Epigenetic changes

DNA epigenetic modifications have been implicated in a number of monogenic conditions, including imprinting disorders and some STR disorders[115,116]. In addition, for some conditions, such as developmental disorders, epigenetic signatures can provide clues to their underlying genetic basis and can be used as functional evidence for clarifying variants of uncertain significance[117]. Epigenetic changes cannot be detected by NGS methods, but long-read sequencing might offer an opportunity to overcome this hurdle. The evidence for methylation of mtDNA has been mixed and influenced by technical issues[118]. However, some studies suggest that mtDNA methylation is substantial, but not localised to the usual CpG sites[119]. Post-transcriptional modifications of mtRNA are complex, but our understanding of these changes is improving[120]. In future, long-reads could provide insights into the epigenetic modification of mtDNA; for example, within the non-coding D-loop, which houses promotor regions and transcription initiation sites[121]. As discussed below, epigenetic modifications might be of particular relevance in studying mitochondrial tRNA molecules, which undergo extensive, disease-relevant post-transcriptional epigenetic changes.

[H2] Transcriptomics

RNA-seq was established in the mid-2000s[122,123] and can be used in both diagnostics and for the study of disease mechanisms. The technique is based on short-read sequencing and essentially involves extraction of RNA, purification of mRNA, fragmentation of molecules, conversion to cDNA, library preparation, sequencing, and alignment to a reference sequence[124]. RNA-seq can also be applied to non-mRNA molecules and we discuss advances in the sequencing of mitochondrial transfer RNAs, which is of particular relevance to PMD.

As with any genomic technology, RNA-seq is not without challenges. For example, as most RNA-seq relies on short-read NGS, the mapping of reads to exons that are common to multiple transcript isoforms is problematic. Thus, identification and quantification of isoforms can be difficult. The use of third-generation technologies can overcome these problems. In particular, nanopore sequencing enables direct RNA sequencing, without the necessity for reverse transcription, and can also identify RNA modifications[125,126]. Unlike DNA-based investigation, relative quantification of RNA levels is important, and these must be normalised against control data. Of note, as gene expression is tissue-specific, the tissue origin of the RNA sample must be considered carefully. Physicians and scientists in mitochondrial medicine are highly attuned to tissue-specificity because of heteroplasmy, which can vary considerably depending on the mtDNA mutation. As solid tissue samples are widely collected in PMD, these disorders are likely to be uniquely placed to benefit from diagnostic RNA-seq in clinical practice. Our routine practice is to undertake RNAseq using RNA extracted from muscle tissue in adults with PMD. If muscle tissue is unavailable, we extract RNA from cultured fibroblasts.

[H3] RNA sequencing can aid diagnostics

RNA-seq can assist diagnostics by uncovering the presence of aberrant transcripts (caused by splicing variants) or altered transcript levels, including differential gene expression and mono-allelic expression (MAE) (Fig. 5). As discussed below, transcriptomic data is particularly important when deciphering the non-coding variants identified in WGS data. Indeed, without this information such variants are extremely difficult to interpret and, unless accompanied by RNAseq, WGS currently has only limited advantages over WES.

Splicing variants are implicated in numerous PMDs (such as mitochondrial encephalopathy, COX deficient Leigh Syndrome, and mitochondrial neurogastrointestinal encephalomyopathy (MNGIE))[127–129]. Importantly, any type of variant, including synonymous changes, can result in aberrant splicing. A high proportion of point mutations have the potential to impact splicing, including at non-canonical

sites[130,131]. At present, DNA-based approaches are limited in their ability to prove, or even identify, disease-causing splicing variants. Although predictive tools are improving, they remain insufficiently sensitive[132]. RNA-seq can identify instances where splicing variants have resulted in aberrant transcripts, for example via exon skipping or intron inclusion.

Differential gene expression analysis is used to identify genes with down-regulated or up-regulated expression compared with healthy physiological levels. This differential expression can be caused either by coding variants or by variants in a non-coding area, for example, enhancer variants, which regulate gene expression as opposed to changing protein structure . Although WGS can uncover variants in non-coding areas, this portion of the genome is extremely large and contains high numbers of variants, which can be challenging to interpret in the absence of functional evidence. Therefore, the identification of differential gene expressioncan be useful for deciphering the effect non-coding variants found during WGS.

In mono-allelic expression, only one allele of a given gene is transcribed, thus transcripts of one parental origin are absent. Although mono-allelic expression is expected for imprinted genes **[G]** , in non-imprinted genes it might be to the result of an unidentified mutation such as an intronic mutation affecting an enhancer. In recessive diseases, identification of mono-allelic expression is useful in suggesting candidate disease-associated genes, which might harbour a second variant in the other allele. For example, RNA-seq analysis in a patient with PMD identified the mono-allelic expression of *ALDH18A1* (a gene involved in mitochondrial proline metabolism) owing to a nonsense variant on one allele[35]. On the other allele, a missense variant of uncertain significance was identified. Follow-up functional studies found levels of the protein product to be significantly reduced as a result of these variants.

Several studies have been undertaken to investigate the role of RNA-seq in the diagnosis of PMD and other rare genetic diseases. Perhaps unsurprisingly, the first major RNA-seq rare disease study was

undertaken in patients with PMD. In 2017 Laura Kremer and colleagues performed RNA-seq on fibroblast cell lines from individuals withsuspected PMD that remained genetically undetermined after WES[35]. The researchers looked for mono-allelic expression of rare variants, aberrant splicing, and abnormal mRNA expression levels. They identified five aberrant splicing events per sample and, by filtering for genes with non-physiological expression levels, reduced this number to one event per sample. This approach enabled the discovery of a novel disease-associated gene and obtained a genetic diagnosis for 5 of the 48 participants, identifying candidate genes for many others. Further successful transcriptomic studies in neuromuscular disease have since followed[133,134]. Interestingly, t-myotubes (engineered from patient fibroblasts) have been found to accurately reflect the transcriptome of myocytes, although with lower expression levels[134]. Despite being labour-intensive, this technique mightrepresent an alternative to muscle biopsy for a highly selected group of patients.

In one study, RNA-seq was performed on lymphocyte RNA from 94 participants with a mixture of rare diseases, achieving a diagnostic rate of 7.5% and providing an additional 16.7% of participants with improved candidate gene resolution[135]. One participant was diagnosed with the PMD mitochondrial enoyl CoA reductase protein associated neurodegeneration, caused by a mutation in *MECR*.

[H3] RNA sequencing for insights into mechanisms

By profiling gene expression levels, transcriptomics can be used to understand the effect of pathological processes on biological pathways, with the potential to identify disease-specific biomarkers and druggable protein targets. Although studies of PMD have previously used array technologies to obtain gene expression data[136], RNA-seq is likely to be increasingly used in future. For example, in one study RNA-seq was performed on muscle samples from healthy controls and participants with myopathy, encephalopathy, lactic acidosis and stroke-like episodes (MELAS) — caused by the m.3243A>G variant in *MT-TL1*. This approach identified a high number of genes (n=224) that showed significantly different expression levels in the group of participats with MELAS compared

with the age-matched controls[137]. These altered gene expression profiles were enriched for immune, metabolic and signal transduction processes. SMRTseq has also been used to study the mitochondrial transcriptome, successfully sequencing full-length transcripts and novel long non-coding RNAs in the D Loop (non-coding region) of the mitochondrial genome[138].

[H2] Mitochondrial transfer RNA molecules

Mt-tRNA diseases were among the first PMDs to be genetically resolved — the most common mtDNA mutations underlying myoclonic epilepsy with ragged-red fibres (m.8344A>G in *MT-TK*) and MELAS (m.3243A>G in *MT-TL1*) were both identified in 1990[139–141]. tRNAs are the most abundant cellular RNA and make up a high proportion (22 in 37) of mitochondrial genes[142]. Therefore, that mutations in mitochondrial tRNA genes cause some of the most common mtDNA-related diseases is unsuprising. The group of known tRNA-associated PMDs continues to expand; these PMDs can be caused by mutations in either mt-tRNA genes or tRNA-modifying genes such as RNase enzymes, which are nuclear-encoded[143,144]. Taking into account gene size (mt-tRNA genes are smaller than mtDNA genes), mutations are more common in mt-tRNA genes than in mitochondrial protein-coding genes[145]. Given this tendency towards a high number of variants, distinguishing pathogenic mutations from benign polymorphisms in mt-tRNA genes can be complex.

The processing of tRNAs is complex. The mtDNA strand is transcribed as a unit before specialised RNase enzymes excise individual tRNA transcripts (Fig. 6)[13]. tRNA transcripts undergo extensive post-transcriptional modifications[13]. These modifications, and the resultant characteristic clover-leaf structure, can disrupt the reverse transcriptase enzyme used in most RNA sequencing, and can affect adaptor ligation **[G]**[146]. Hence, attempts to sequence tRNA can result in artificially truncated fragments[147]. Moreover, the multiplicity of very similar tRNA genes makes mapping transcripts back to a reference sequence difficult[148].

Despite their frequency, the classification of variants in mt-tRNA genes is not straightforward. Important considerations are heteroplasmy; whether the variant affects the anticodon **[G]** or wobble position **[G]** ; nucleotide modifications; and that (unlike in protein coding genes) insertions or deletions do not affect the reading frame, but might affect the structure of the tRNA molecule[44]. A useful article, published in 2020, describes a method for tailoring ACMG criteria to mt-tRNA variant interpretation[44].

[H3] Sequencing tRNA

Techniques that use a bacterial demethylase enzyme (AlkB) — to remove specific 'hard stop' methyl groups — and a robust reverse transcriptase are now facilitating effective sequencing of tRNA[146,149]. Another processing method has been developed that results in shorter tRNA molecules, which are less folded, less modified, and without the aminoacyl-tRNA bond, thus enabling easier 3' adaptor attachment and allowing identification of pre-tRNA transcripts[148]. This method was recently used to create an atlas of human tRNA[148]. Other methods use innovative Y-shaped adaptors to improve efficiency of adaptor ligation[150]. Excitingly, another method, QuantM-tRNA-seq, the details of which were published for the first time in 2020, facilitates high-throughput sequencing and robust quantifcation of mature tRNA[151]. Bioinformatic pipelines that specifically investigate tRNA are also being developed, as described in a preprint article from 2019[152]. Nanopore technologies have been used to study tRNA and can concomitantly identify tRNA sequence modifications and post-transcriptional (epigenetic) modifications[153].

[H3] Insights into tRNA-related PMD

RNA-seq of mt-tRNA molecules can provide functional insights into mt-tRNA-related diseases (Fig. 6). For example, a study of tRNA containing the m.8344A>G variant in *MT-TK* (the most common cause of the MERRF phenotype) identified a missing epigenetic modification in *MT-TK* (N1-

methyladenosine), which appeared to result in defective elongation and decreased stability of the polypeptide chain[154]

tRNA sequencing also offers insights into mutations in nDNA genes involved in tRNA processing. A study identified biallelic *TRMT10C* variants in two patients with a severe neonatal mitochondrial presentation consisting of low tone, poor feeding, lactic acidosis, and deafness[155]. *TRMT10C* encodes a subunit of the mt-RNase P complex which releases the 5' end of tRNAs from the polycistronic transcript[156] . Using RNA-seq, the reads across the mitochondrial transcriptome in the two patients were compared with reads from healthy controls, showing an increase in reads spanning gene boundaries in the patients. This observation suggests that tRNA and mRNA were not being cleaved effectively from the polycistronic transcript.

Functional studies in mt-tRNA-associated diseases have historically been based on the quantification of mutant loads in single muscle fibres (for example, using quantitative polymerase chain reactions) and the correlation of high mutant load with COX deficiency[144]. Levels of tRNA in patients can also be compared with levels in healthy controls, for example, by use of high resolution northern blotting[157,158]. However, RNA-seq is likely to be increasingly applied in this area. For example, RNA-seq has been used to investigate mitochondrial tRNA fragments, a form of small non-coding RNA thought to negatively regulate gene expression, in MELAS[159]. Using a specialised type of RNA-seq researchers found these tRNA fragments to be down-regulated in m.3243A>G cybrids **[G]**from individuals with MELAS and were subsequently able to demonstrate the biological function of one of the fragments.

[H1] Conclusions

The role of genetics in neurology is currently undergoing a transformation. Genomic medicine has evolved from a research-based endeavor to an established and invaluable tool for diagnostic genetic laboratories. Genome sequencing was previously reserved for the most challenging cases, but is increasingly being adopted as a first-line investigation for rare genetic neurological diseases. In order

to achieve the next diagnostic uplift and improve efficiency, new laboratory techniques and sequencing technologies will need to be embraced by clinicians and researchers working in the field of mitochondrial medicine. Carefully selected patients with PMD are likely to benefit substnatially from WGS early in their diagnostic journey. Long-read sequencing has the potential to provide advances in the identification of new genetic causes of PMD, solve phasing issues, and improve RNA and mtDNA investigations through direct sequencing. RNA-seq will increasingly be used for diagnosis and to provide functional support in challenging cases, and the widespread availability of tissue samples from patients with PMD confers major advantages that will support rapid introduction of this technology into diagnostic laboratories. Finally, validation of new tRNA methods will help confirm pathogenicity in this common group of mtDNA-related PMDs.

References

1. Rahman, J. & Rahman, S. Mitochondrial medicine in the omics era. *Lancet* **391**, 2560–2574 (2018).

2. Stenton, S. L. & Prokisch, H. Advancing genomic approaches to the molecular diagnosis of mitochondrial disease. *Essays Biochem.* **62**, 399–408 (2018).

3. Gorman, G. S. *et al.* Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann. Neurol.* **77**, 753–759 (2015).

4. Castro-Gago, M. *et al.* Epidemiology of pediatric mitochondrial respiratory chain disorders in northwest Spain. *Pediatr. Neurol.* **34**, 204–211 (2006).

5. Gorman, G. S. *et al.* Mitochondrial diseases. *Nat. Rev. Dis. Prim.* **2**, 16080 (2016).

6. Carelli, V. & Morgia, C. La. Clinical syndromes associated with mtDNA mutations: Where we stand after 30 years. *Essays Biochem.* **62**, 235–254 (2018).

7. Nesbitt, V. *et al.* The UK MRC Mitochondrial Disease Patient Cohort Study: clinical phenotypes associated with the m.3243A>G mutation--implications for diagnosis and management. *J.*

668 *Neurol. Neurosurg. Psychiatry* **84**, 936–938 (2013).

669 8. Pitceathly, R. D. S. *et al.* Genetic dysfunction of MT-ATP6 causes axonal Charcot-Marie-Tooth
670 disease. *Neurology* **79**, 1145–1154 (2012).

671 9. Pitceathly, R., Keshavan, N., Rahman, J. & Rahman, S. *Moving Towards Clinical Trials for*
672 *Mitochondrial Diseases*. *Journal of Inherited Metabolic Disease* vol. Online (2020).

673 10. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature*
674 **290**, 1–18 (1981).

675 11. Robin, E. D. & Wong, R. Mitochondrial DNA molecules and virtual number of mitochondria
676 per cell in mammalian cells. *J. Cell. Physiol.* **136**, 507–513 (1988).

677 12. Bogenhagen, D. F. Mitochondrial DNA nucleoid structure. *Biochim. Biophys. Acta* **1819**, 914–
678 920 (2012).

679 13. D'Souza, A. R. & Minczuk, M. Mitochondrial transcription and translation: Overview. *Essays*
680 *Biochem.* **62**, 309–320 (2018).

681 14. Wei, W. *et al.* Nuclear-mitochondrial DNA segments resemble paternally inherited
682 mitochondrial DNA in humans. *Nat. Commun.* **11**, 1–11 (2020).

683 15. Rath, S. *et al.* MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle
684 localization and pathway annotations. *Nucleic Acids Res.* 1–7 (2020)
685 doi:10.1093/nar/gkaa1011.

686 16. Frazier, A. E., Thorburn, D. R. & Compton, A. G. Mitochondrial energy generation disorders:
687 Genes, mechanisms, and clues to pathology. *J. Biol. Chem.* **294**, 5386–5395 (2019).

688 17. El-Hattab, A. W., Craigen, W. J. & Scaglia, F. Mitochondrial DNA maintenance defects.
689 *Biochim. Biophys. Acta - Mol. Basis Dis.* **1863**, 1539–1555 (2017).

690 18. Rahman, S., Poulton, J., Marchington, D. & Suomalainen, A. Decrease of 3243 A→G mtDNA

691 mutation from blood in MELAS syndrome: A longitudinal study. *Am. J. Hum. Genet.* **68**, 238–
692 240 (2001).

693 19. Goldstein, A. & Falk, M. J. *Mitochondrial DNA Deletion Syndromes*. *GeneReviews®* (University
694 of Washington, 2019).

695 20. He, Y. *et al.* Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature*
696 **464**, 610–614 (2010).

697 21. Rossignol, R. *et al.* Mitochondrial threshold effects. *Biochem. J.* **370**, 751–762 (2003).

698 22. Elliott, H. R., Samuels, D. C., Eden, J. A., Relton, C. L. & Chinnery, P. F. Pathogenic
699 mitochondrial DNA mutations are common in the general population. *Am. J. Hum. Genet.* **83**,
700 254–260 (2008).

701 23. Maeda, K. *et al.* Clinical phenotype and segregation of mitochondrial 3243A>G mutation in 2
702 pairs of monozygotic twins. *JAMA Neurol.* **73**, 990–993 (2016).

703 24. Lynn, S., Borthwick, G. M., Charnley, R. M., Walker, M. & Turnbull, D. M. Heteroplasmic ratio
704 of the A3243G mitochondrial DNA mutation in single pancreatic beta cells. *Diabetologia* **46**,
705 296–299 (2003).

706 25. Kohda, M. *et al.* A Comprehensive Genomic Analysis Reveals the Genetic Landscape of
707 Mitochondrial Respiratory Chain Complex Deficiencies. *PLoS Genet.* **12**, 1–31 (2016).

708 26. Lieber, D. S. *et al.* Targeted exome sequencing of suspected mitochondrial disorders.
709 *Neurology* **80**, 1762–1770 (2013).

710 27. Ohtake, A. *et al.* Diagnosis and molecular basis of mitochondrial respiratory chain disorders:
711 Exome sequencing for disease gene identification. *Biochim. Biophys. Acta - Gen. Subj.* **1840**,
712 1355–1359 (2014).

713 28. Pronicka, E. *et al.* New perspective in diagnostics of mitochondrial disorders: Two years'

experience with whole-exome sequencing at a national paediatric centre. *J. Transl. Med.* **14**, 1–19 (2016).

29. Taylor, R. W. *et al.* Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. *JAMA - J. Am. Med. Assoc.* **312**, 68–77 (2014).

30. Wortmann, S. B., Koolen, D. A., Smeitink, J. A., van den Heuvel, L. & Rodenburg, R. J. Whole exome sequencing of suspected mitochondrial patients in clinical practice. *J. Inherit. Metab. Dis.* **38**, 437–443 (2015).

31. Cui, H. *et al.* Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. *Genet. Med.* **15**, 388–394 (2013).

32. Zhang, W., Cui, H. & Wong, L. J. C. Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin. Chem.* **58**, 1322–1331 (2012).

33. Seneca, S. *et al.* Analysis of the whole mitochondrial genome : translation of the Ion Torrent Personal Genome Machine system to the diagnostic bench ? *Eur. J. Hum. Genet.* **23**, 41–48 (2015).

34. Riley, L. G. *et al.* The diagnostic utility of genome sequencing in a pediatric cohort with suspected mitochondrial disease. *Genet. Med.* **22**, 1254–1261 (2020).

35. Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* **8**, 1–11 (2017).

36. Ellerby, L. M. Repeat Expansion Disorders: Mechanisms and Therapeutics. *Neurotherapeutics* vol. 16 924–927 (2019).

37. Yu-Wai-Man, P. & Chinnery, P. F. Leber Hereditary Optic Neuropathy. in *GeneReviews®* (eds.

737 Adam, M. P. et al.) 1–19 (University of Washington, Seattle, 2016).

738 38. Wong, L. J. C. *et al.* Utility of oligonucleotide array-based comparative genomic hybridization
739 for detection of target gene deletions. *Clin. Chem.* **54**, 1141–1148 (2008).

740 39. Naini, A. & Shanske, S. Detection of Mutations in mtDNA. *Methods Cell Biol.* **80**, 437–463
741 (2007).

742 40. Chinault, A. C., Shaw, C. A., Brundage, E. K., Tang, L. Y. & Wong, L. J. C. Application of dual-
743 genome oligonucleotidearray-based comparative genomic hybridization to the molecular
744 diagnosis of mitochondrial DNA deletion and depletion syndromes. *Genet. Med.* **11**, 518–526
745 (2009).

746 41. Shanske, S. & Wong, L. J. C. Molecular analysis for mitochondrial DNA disorders.
747 *Mitochondrion* **4**, 403–415 (2004).

748 42. Grier, J., Hirano, M., Karaa, A., Shepard, E. & Thompson, J. L. P. Diagnostic odyssey of patients
749 with mitochondrial disease Results of a survey. *Neurol. Genet.* **4**, (2018).

750 43. Alston, C. L., Rocha, M. C., Lax, N. Z., Turnbull, D. M. & Taylor, R. W. The genetics and
751 pathology of mitochondrial disease. *J. Pathol.* **241**, 236–250 (2017).

752 44. Wong, L. J. C. *et al.* Interpretation of mitochondrial tRNA variants. *Genet. Med.* **22**, 917–926
753 (2020).

754 45. Diroma, M. A. *et al.* Extraction and annotation of human mitochondrial genomes from 1000
755 Genomes Whole Exome Sequencing data. *BMC Genomics* **15**, 1–15 (2014).

756 46. Garret, P. *et al.* Deciphering exome sequencing data: Bringing mitochondrial DNA variants to
757 light. *Hum. Mutat.* **40**, 2430–2443 (2019).

758 47. Griffin, H. R. *et al.* Accurate mitochondrial DNA sequencing using off-target reads provides a
759 single test to identify pathogenic point mutations. *Genet. Med.* **16**, (2014).

760 48. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human
761 mitochondrial DNA variation. *Hum. Mutat.* **30**, 386–394 (2009).

762 49. Ghelli, A. M. *et al.* The background of mitochondrial DNA haplogroup J increases the
763 sensitivity of Leber's hereditary optic neuropathy cells to 2,5-hexanedione toxicity. *PLoS One*
764 **4**, (2009).

765 50. Hudson, G. *et al.* Clinical expression of leber hereditary optic neuropathy is affected by the
766 mitochondrial DNA-haplogroup background. *Am. J. Hum. Genet.* **81**, 228–233 (2007).

767 51. Ye, F., Samuels, D. C., Clark, T. & Guo, Y. High-throughput sequencing in mitochondrial DNA
768 research. *Mitochondrion* **17**, 157–163 (2014).

769 52. Gould, M. P. *et al.* PCR-free enrichment of mitochondrial DNA from human blood and cell
770 lines for high quality next-generation DNA sequencing. *PLoS One* **10**, 1–13 (2015).

771 53. Akbari, M., Hansen, M. D., Halgunset, J., Skorpen, F. & Krokan, H. E. Low Copy Number DNA
772 Template Can Render Polymerase Chain Reaction Error Prone in a Sequence-Dependent
773 Manner. *J Mol Diagn* **7**, 36–39 (2005).

774 54. Santibanez-koref, M. *et al.* Assessing mitochondrial heteroplasmy using next generation
775 sequencing : A note of caution. *Mitochondrion* **46**, 302–306 (2019).

776 55. Hazkani-Covo, E., Zeller, R. M. & Martin, W. Molecular poltergeists: Mitochondrial DNA copies
777 (numts) in sequenced nuclear genomes. *PLoS Genet.* **6**, (2010).

778 56. Tourmen, Y. *et al.* Structure and chromosomal distribution of human mitochondrial
779 pseudogenes. *Genomics* **80**, 71–77 (2002).

780 57. Parr, R. L. *et al.* The pseudo-mitochondrial genome influences mistakes in heteroplasmy
781 interpretation. *BMC Genomics* **13**, 1–13 (2006).

782 58. Marquis, J. *et al.* MitoRS, a method for high throughput, sensitive, and accurate detection of

783 mitochondrial DNA heteroplasmy. *BMC Genomics* **18**, 1–19 (2017).

784 59. Wolff, J. N., Shearman, D. C. A., Brooks, R. C. & Ballard, J. W. O. Selective Enrichment and
785 Sequencing of Whole Mitochondrial Genomes in the Presence of Nuclear Encoded
786 Mitochondrial Pseudogenes (Numts). *PLoS One* **7**, 1–7 (2012).

787 60. Ancora, M. *et al.* Mitochondrial heteroplasmy profiling in single human oocytes by next-
788 generation sequencing. *Mitochondrial DNA Part B Resour.* 543–544 (2017)
789 doi:10.1080/23802359.2017.1365634.

790 61. Yao, Y. *et al.* A simple method for sequencing the whole human mitochondrial genome
791 directly from samples and its application to genetic testing. *Sci. Rep.* **9**, 1–7 (2019).

792 62. Williams, S. L. *et al.* The mtDNA mutation spectrum of the progeroid Polg mutator mouse
793 includes abundant control region multimers. *Cell Metab* **12**, 675–682 (2010).

794 63. Maricic, T., Whitten, M. & Pa, S. Multiplexed DNA Sequence Capture of Mitochondrial
795 Genomes Using PCR Products. *PLoS One* **5**, 9–13 (2010).

796 64. Weerts, M. J. A., Timmermans, E. C., Vossen, R. H. A. M. & Strijp, D. Van. Sensitive detection
797 of mitochondrial DNA variants for analysis of mitochondrial DNA- enriched extracts from
798 frozen tumor tissue. *Sci. Rep.* **8**, 1–12 (2018).

799 65. McDowell, D. G., Burns, N. A. & Parkes, H. C. Localised sequence regions possessing high
800 melting temperatures prevent the amplification of a DNA mimic in competitive PCR. *Nucleic*
801 *Acids Res.* **26**, 3340–3347 (1998).

802 66. Wood, E. *et al.* Clinical long-read sequencing of the human mitochondrial genome for
803 mitochondrial disease diagnostics. *bioRxiv* 597187 (2019) doi:10.1101/597187.

804 67. Strachan, T. & Read, A. *Human Molecular Genetics*. (CRC Press, 2019).

805 68. Almannai, M., El-Hattab, A. W. & Scaglia, F. Mitochondrial DNA replication: Clinical

syndromes. *Essays Biochem.* **62**, 297–308 (2018).

69. Raymond, F. L., Horvath, R. & Chinnery, P. F. First-line genomic diagnosis of mitochondrial disorders. *Nat. Rev. Genet.* **19**, 399–400 (2018).

70. Koenig, M. K. Presentation and diagnosis of mitochondrial disorders in children. *Pediatr. Neurol.* **38**, 305–313 (2008).

71. Posey, J. E. *et al.* Resolution of disease phenotypes resulting from multilocus genomic variation. *N. Engl. J. Med.* **376**, 21–31 (2017).

72. Parikh, S. *et al.* Diagnosis of possible mitochondrial disease: An existential crisis. *J. Med. Genet.* **56**, 123–130 (2019).

73. Uittenbogaard, M. *et al.* The nuclear background influences the penetrance of the near-homoplasmic m.1630 A>G MELAS variant in a symptomatic proband and asymptomatic mother. *Mol. Genet. Metab.* **126**, 429–438 (2019).

74. Boczonadi, V., Bansagi, B. & Horvath, R. Reversible infantile mitochondrial diseases. *J. Inherit. Metab. Dis.* **38**, 427–435 (2015).

75. Horvath, R. *et al.* Molecular basis of infantile reversible cytochrome c oxidase deficiency myopathy. *Brain* **132**, 3165–3174 (2009).

76. Hathazi, D. *et al.* Metabolic shift underlies recovery in reversible infantile respiratory chain deficiency. *EMBO J.* **44**, 1–19 (2020).

77. Rygiel, K. A. *et al.* Complex mitochondrial DNA rearrangements in individual cells from patients with sporadic inclusion body myositis. *Nucleic Acids Res.* **44**, 5313–5329 (2016).

78. Richards, S., Aziz, N., Bale, S., Bick, D. & Das, S. ACMG Standards and Guidelines Standards and guidelines for the interpretation of sequence variants : a joint consensus recommendation of the American College of Medical Genetics and Genomics and the

Association for Molecular Pathology. 1–20 (2015) doi:10.1038/gim.2015.30.

79. Marshall, C. R. *et al.* The Medical Genome Initiative: moving whole-genome sequencing for rare disease diagnosis to the clinic. *Genome Med.* **12**, 48 (2020).

80. Gerner-Smidt, P. *et al.* Whole genome sequencing: Bridging one-health surveillance of foodborne diseases. *Front. Public Heal.* **7**, 1–11 (2019).

81. Giannopoulou, E., Katsila, T., Mitropoulou, C., Tsermpini, E. E. & Patrinos, G. P. Integrating next-generation sequencing in the clinical pharmacogenomics workflow. *Front. Pharmacol.* **10**, 1–6 (2019).

82. Turnbull, C. *et al.* The 100 000 Genomes Project: Bringing whole genome sequencing to the NHS. *BMJ* **361**, 1–7 (2018).

83. Handsaker, R. E. *et al.* Large multiallelic copy number variations in humans. *Nat. Genet.* **47**, 296–303 (2015).

84. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).

85. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science (80-.).* **323**, 133–138 (2009).

86. Travers, K. J., Chin, C. S., Rank, D. R., Eid, J. S. & Turner, S. W. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.* **38**, (2010).

87. van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. & Thermes, C. The Third Revolution in Sequencing Technology. *Trends Genet.* **34**, 666–681 (2018).

88. Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 1–16 (2020).

89. Vossen, R. H. A. M. & Buermans, H. P. J. Chapter 16 Methods for Genotyping-by-Sequencing.

852 in *Genotyping: Methods and Protocols* (eds. White, S. J. & Cantsilieris, S.) 179–184 (Humana
853 Press, 2017). doi:10.1007/978-1-4939-6442-0.

854 90. Chakraborty, S. *et al.* Mitochondrial DNA Sequencing Using PacBio SMRT Technology. in
855 *Advances in Genome Biology and Technology* vol. Poster (2018).

856 91. Borràs, D. M. *et al.* Detecting PKD1 variants in polycystic kidney disease patients by single-
857 molecule long-read sequencing. *Hum. Mutat.* **38**, 870–879 (2017).

858 92. Frans, G. *et al.* Conventional and Single-Molecule Targeted Sequencing Method for Specific
859 Variant Detection in IKBKG while Bypassing the IKBKGP1 Pseudogene. *J. Mol. Diagnostics* **20**,
860 195–202 (2018).

861 93. Alkanaq, A. N. *et al.* Comparison of mitochondrial DNA variants detection using short- and
862 long-read sequencing. *J. Hum. Genet.* **64**, 1107–1116 (2019).

863 94. Wenger, A. M. *et al.* Accurate circular consensus long-read sequencing improves variant
864 detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162 (2019).

865 95. Vollger, M. R. *et al.* Improved assembly and variant detection of a haploid human genome
866 using single-molecule, high-fidelity long reads. *Ann. Hum. Genet.* **84**, 125–140 (2020).

867 96. Zascavage, R. R. *et al.* Approaches to Whole Mitochondrial Genome Sequencing on the
868 Oxford Nanopore MinION. *Curr. Protoc. Hum. Genet.* **104**, e94 (2019).

869 97. Zascavage, R. R., Thorson, K. & Planz, J. V. Nanopore sequencing: An enrichment-free
870 alternative to mitochondrial DNA sequencing. *Electrophoresis* **40**, 272–280 (2019).

871 98. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads.
872 *Nat. Biotechnol.* **36**, 338–345 (2018).

873 99. Chaisson, M. J. P. *et al.* Resolving the complexity of the human genome using single-molecule
874 sequencing. *Nature* **517**, 608–611 (2015).

100. Seo, J. S. *et al.* De novo assembly and phasing of a Korean human genome. *Nature* **538**, 243–247 (2016).

101. Sharp, A. J. *et al.* Segmental Duplications and Copy-Number Variation in the Human Genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).

102. Chiang, C. *et al.* The impact of structural variation on human gene expression. *Nat. Genet.* **49**, 692–699 (2017).

103. Chaisson, M. J. P. *et al.* Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat. Commun.* **10**, 1–16 (2019).

104. Beyter, D. *et al.* Long read sequencing of 1,817 Icelanders provides insight into the role of structural variants in human disease. *bioRxiv* 1–44 (2019) doi:10.1101/848366.

105. Dolzhenko, E. *et al.* ExpansionHunter: A sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* **35**, 4754–4756 (2019).

106. Dolzhenko, E. *et al.* ExpansionHunter Denovo: A computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol.* **21**, 1–14 (2020).

107. Mitsuhashi, S. *et al.* Tandem-genotypes: robust detection of tandem repeat expansions from long DNA reads. *Genome Biol.* **20**, 1–17 (2019).

108. Ummat, A. & Bashir, A. Resolving complex tandem repeats with long reads. *Bioinformatics* **30**, 3491–3498 (2014).

109. Liu, Q., Zhang, P., Wang, D., Gu, W. & Wang, K. Interrogating the 'unsequenceable' genomic trinucleotide repeat disorders by long-read sequencing. *Genome Med.* **9**, 1–16 (2017).

110. Sone, J. *et al.* Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nat. Genet.* **51**, 1215–1221 (2019).

111. Ardui, S. *et al.* Detecting AGG interruptions in females with a FMR1 premutation by long-read

single-molecule sequencing: A 1 year clinical experience. *Front. Genet.* **9**, 1–6 (2018).

112. Cumming, S. A. *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.* **26**, 1635–1647 (2018).

113. Nakamura, H. *et al.* Long-read sequencing identifies the pathogenic nucleotide repeat expansion in RFC1 in a Japanese case of CANVAS. *J. Hum. Genet.* **65**, 475–480 (2020).

114. Mitsuhashi, S. *et al.* Nanopore-based single molecule sequencing of the D4Z4 array responsible for facioscapulohumeral muscular dystrophy. *Sci. Rep.* **7**, 1–8 (2017).

115. Nageshwaran, S. & Festenstein, R. Epigenetics and triplet-repeat neurological diseases. *Front. Neurol.* **6**, 1–9 (2015).

116. Elhamamsy, A. R. Role of DNA methylation in imprinting disorders: an updated review. *J. Assist. Reprod. Genet.* **34**, 549–562 (2017).

117. Sadikovic, B., Aref-Eshghi, E., Levy, M. A. & Rodenhiser, D. DNA methylation signatures in mendelian developmental disorders as a diagnostic bridge between genotype and phenotype. *Epigenomics* **11**, 563–575 (2019).

118. Sharma, N., Pasala, M. S. & Prakash, A. Mitochondrial DNA: Epigenetics and environment. *Environ. Mol. Mutagen.* **60**, 668–682 (2019).

119. Patil, V. *et al.* Human mitochondrial DNA is extensively methylated in a non-CpG context. *Nucleic Acids Res.* **47**, 10072–10085 (2019).

120. Pearce, S. F. *et al.* Regulation of Mammalian Mitochondrial Gene Expression: Recent Advances. *Trends Biochem. Sci.* **42**, 625–639 (2017).

121. Kang, D., Miyako, K., Kai, Y., Irie, T. & Takeshige, K. In vivo determination of replication origins of human mitochondrial DNA by ligation-mediated polymerase chain reaction. *J. Biol. Chem.*

921 **272**, 15275–15279 (1997).

922 122. Emrich, S. J., Barbazuk, W. B., Li, L. & Schnable, P. S. Gene discovery and annotation using
923 LCM-454 transcriptome sequencing. *Genome Res.* **17**, 69–73 (2007).

924 123. Lister, R. *et al.* Highly Integrated Single-Base Resolution Maps of the Epigenome in
925 Arabidopsis. *Cell* **133**, 523–536 (2008).

926 124. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**,
927 631–656 (2019).

928 125. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat.*
929 *Methods* **15**, 201–206 (2018).

930 126. Workman, R. E. *et al.* Nanopore native RNA sequencing of a human poly(A) transcriptome.
931 *Nat. Methods* **16**, 1297–1308 (2019).

932 127. Navarro-Sastre, A. *et al.* A fatal mitochondrial disease is associated with defective NFU1
933 function in the maturation of a subset of mitochondrial Fe-S proteins. *Am. J. Hum. Genet.* **89**,
934 656–667 (2011).

935 128. Pitceathly, R. D. S. *et al.* NDUFA4 Mutations Underlie Dysfunction of a Cytochrome c Oxidase
936 Subunit Linked to Human Neurological Disease. *Cell Rep.* **3**, 1795–1805 (2013).

937 129. Taanman, J. W. *et al.* Characterization of a novel TYMP splice site mutation associated with
938 mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *Neuromuscul. Disord.* **19**,
939 151–154 (2009).

940 130. López-Bigas, N., Audit, B., Ouzounis, C., Parra, G. & Guigó, R. Are splicing mutations the most
941 frequent cause of hereditary disease? *FEBS Lett.* **579**, 1900–1903 (2005).

942 131. Lord, J. *et al.* Pathogenicity and selective constraint on variation near splice sites. *Genome*
943 *Res.* **29**, 159–170 (2019).

944 132. Wai, H. A. *et al.* Blood RNA analysis can increase clinical diagnostic rate and resolve variants
945 of uncertain significance. *Genet. Med.* **0**, 1–10 (2020).

946 133. Cummings, B. B. *et al.* Improving genetic diagnosis in Mendelian disease with transcriptome
947 sequencing. *Sci Transl Med* **12**, 1–25 (2017).

948 134. Gonorazky, H. D. *et al.* Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for
949 Rare Mendelian Disease. *Am. J. Hum. Genet.* **104**, 466–483 (2019).

950 135. Frésard, L. *et al.* Identification of rare-disease genes using blood transcriptome sequencing
951 and large control cohorts. *Nat Med* **25**, 911–919 (2019).

952 136. Tyynismaa, H. *et al.* Mitochondrial myopathy induces a starvation-like response. **19**, 3948–
953 3958 (2010).

954 137. Deng, J. *et al.* RNA-seq profiling, and impaired autophagic process in skeletal muscle of
955 MELAS. *Biochem. Biophys. Res. Commun.* **523**, 91–97 (2020).

956 138. Gao, S. *et al.* Mitochondrion Two novel lncRNAs discovered in human mitochondrial DNA
957 using PacBio full-length transcriptome data. *Mitochondrion* **38**, 41–47 (2018).

958 139. Shoffner, J. M. *et al.* Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated
959 with a mitochondrial DNA tRNALys mutation. *Cell* **61**, 931–937 (1990).

960 140. Goto, Y. I., Nonaka, I. & Horai, S. A mutation in the tRNALeu(UUR) gene associated with the
961 MELAS subgroup of mitochondrial encephalomyopathies. *Nature* **348**, 651–653 (1990).

962 141. DiMauro, S. & Garone, C. Historical perspective on mitochondrial medicine. *Dev. Disabil. Res.*
963 *Rev.* **16**, 106–113 (2010).

964 142. Pan, T. Modifications and functional genomics of human transfer RNA. *Cell Res.* **28**, 395–404
965 (2018).

966 143. Boczonadi, V., Ricci, G. & Horvath, R. Mitochondrial DNA transcription and translation: Clinical

967 syndromes. *Essays Biochem.* **62**, 321–340 (2018).

968 144. Yarham, J. W., Elson, J. L., Blakely, E. L., Mcfarland, R. & Taylor, R. W. Mitochondrial tRNA
969 mutations and disease. *Wiley Interdiscip. Rev. RNA* **1**, 304–324 (2010).

970 145. Wong, L.-J. C. *et al.* Comprehensive scanning of the entire mitochondrial genome for
971 mutations. *Clin. Chem.* **48**, 1901–1912 (2002).

972 146. Zheng, G. *et al.* Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods*
973 **12**, 835–837 (2015).

974 147. Wilusz, J. E. Removing roadblocks to deep sequencing of modified RNAs. *Nat. Methods* **12**,
975 821–822 (2015).

976 148. Gogakos, T. *et al.* Characterizing Expression and Processing of Precursor and Mature Human
977 tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep.* **20**, 1463–1475 (2017).

978 149. Cozen, A. E. *et al.* ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex
979 landscape of modified tRNA fragments. *Nat. Methods* **12**, 879–884 (2015).

980 150. Shigematsu, M. *et al.* YAMAT-seq: an efficient method for high-throughput sequencing of
981 mature transfer RNAs. *Nucleic Acids Res.* **45**, e70 (2017).

982 151. Pinkard, O., McFarland, S., Sweet, T. & Coller, J. Quantitative tRNA-sequencing uncovers
983 metazoan tissue-specific tRNA regulation. *Nat. Commun.* **11**, 1–15 (2020).

984 152. James-bott, A. & Cribbs, A. P. tRNAnalysis: A flexible pre-processing and next-generation
985 sequencing data analysis pipeline for transfer RNA Anna. *bioRxiv* (2019)
986 doi:https://doi.org/10.1101/655829.

987 153. Smith, A. M., Abu-Shumays, R., Akeson, M. & Bernick, D. L. Capture, unfolding, and detection
988 of individual tRNA molecules using a nanopore device. *Front. Bioeng. Biotechnol.* **3**, 1–11
989 (2015).

990 154. Richter, U. *et al.* RNA modification landscape of the human mitochondrial tRNALys regulates
991 protein synthesis. *Nat. Commun.* **9**, 1–11 (2018).

992 155. Metodiev, M. D. *et al.* Recessive Mutations in TRMT10C Cause Defects in Mitochondrial RNA
993 Processing and Multiple Respiratory Chain Deficiencies. *Am. J. Hum. Genet.* **98**, 993–1000
994 (2016).

995 156. Holzmann, J. *et al.* RNase P without RNA: identification and functional reconstitution of the
996 human mitochondrial tRNA processing enzyme. *Cell* **135**, 462–474 (2008).

997 157. Lehmann, D. *et al.* Pathogenic mitochondrial mt-tRNA Ala variants are uniquely associated
998 with isolated myopathy. *Eur. J. Hum. Genet.* **23**, 1735–1738 (2015).

999 158. Taylor, R. W. *et al.* A homoplasmic mitochondrial transfer ribonucleic acid mutation as a
1000 cause of maternally inherited hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **41**, 1786–
1001 1796 (2003).

1002 159. Meseguer, S. *et al.* The MELAS mutation m.3243A>G alters the expression of mitochondrial
1003 tRNA fragments. *Biochim. Biophys. Acta - Mol. Cell Res.* **1866**, 1433–1449 (2019).

1004 160. El-Hattab, A. W., Almannai, M. & Scaglia, F. MELAS. in *Gene reviews* (eds. Adam, M. P. et al.)
1005 (2001).

1006 161. Naing, A. *et al.* Maternally inherited diabetes and deafness (MIDD): diagnosis and
1007 management. *J. Diabetes Complications* **28**, 542–546 (2014).

1008 162. Moraes, C. T. *et al.* Atypical clinical presentations associated with the MELAS mutation at
1009 position 3243 of human mitochondrial DNA. *Neuromuscul. Disord.* **3**, 43–50 (1993).

1010 163. Horga, A. *et al.* Peripheral neuropathy predicts nuclear gene defect in patients with
1011 mitochondrial ophthalmoplegia. *Brain* **137**, 3200–3212 (2014).

1012 164. DiMauro, S. & Hirano, M. MERRF. in (eds. Adam, M. P. et al.) (2003).

165. Rahman, S. *et al.* Leigh syndrome: clinical features and biochemical and DNA abnormalities. *Ann. Neurol.* **39**, 343–351 (1996).

166. Thorburn, D. R., Rahman, J. & Rahman, S. Mitochondrial DNA-Associated Leigh Syndrome and NARP. in (eds. Adam, M. P. et al.) (2003).

Key references

1. Zhang, W., Cui, H. & Wong, L. J. C. Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. Clin. Chem. 58, 1322–1331 (2012).

The key paper that established the role of deep NGS of long-range PCR-amplified mtDNA in identifying point mutations and large deletions, including heteroplasmic variants.

2. Riley, L. G. *et al.* The diagnostic utility of genome sequencing in a pediatric cohort with suspected mitochondrial disease. *Genet. Med.* 22, 1254–1261 (2020).

The first study to use WGS specifically for mitochondrial presentations.

3. Hathazi, D. *et al.* Metabolic shift underlies recovery in reversible infantile respiratory chain deficiency. *EMBO J.* 44, 1–19 (2020).

A very recent paper that elegantly illustrates the complex interplay between the mitochondrial and nuclear genomes.

4. Wong, L. J. C. *et al.* Interpretation of mitochondrial tRNA variants. *Genet. Med.* 22, 917–926 (2020).

An insightful paper that presents the analysis of a large number of mt-tRNA variants in the context of modern variant classification standards and suggests a tailored approach to classifying

mutations in this unique group of molecules.

5. **Wood, E. *et al.* Clinical long-read sequencing of the human mitochondrial genome for mitochondrial disease diagnostics. *bioRxiv* 597187 (2019) doi:10.1101/597187.**

   The first clinical use of nanopore for mtDNA variants; although accuracy problems lead to false positive point mutations, the technology successfully sequenced the entire genome and identified deletions missed by NGS.

6. **Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* 8, 1–11 (2017).**

   This paper showed that RNA-seq could be used to diagnose mitochondrial diseases and was the first large study to demonstrate the clinical utility of the technique in a rare disease.

Acknowledgments

The University College London Hospitals/University College London Queen Square Institute of Neurology sequencing facility receives a proportion of funding from the Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme. The clinical and diagnostic 'Rare Mitochondrial Disorders' Service in London is funded by the UK NHS Highly Specialised Commissioners. The work of R.D.S.P. is supported by a Medical Research Council Clinician Scientist Fellowship (MR/S002065/1). J.V. holds a fellowship from the Health Education England Genomics Education Programme. All authors are supported by an MRC strategic award to establish an International Centre for Genomic Medicine in Neuromuscular Diseases (ICGNMD) (MR/S005021/1).

The authors are grateful for the feedback on the manuscript provided by Dr Lindsay Wilson, research manager for the ICGNMD, and for the expert input from Dr Robyn Labrum and Ms Cathy Woodward, Neurogenetics Unit, The National Hospital for Neurology and Neurosurgery, Queen Square, London.

Author Contributions

All authors contributed equally to the conceptualisation and writing of the paper.

Competing interests

The authors declare no competing interests.

Publisher's note

Related links

Genomics England PanelApp: https://panelapp.genomicsengland.co.uk

MITOMAP: https://www.mitomap.org/MITOMAP

Figure 1| Summary of PMD symptoms

Owing to the central role of mitochondria in cell metabolism across nearly all tissues, primary mitochondrial disease (PMD) can present with a wide range of symptoms and can affect multiple systems, including the CNS and the peripheral nervous system. Figure adapted from ref[5] **[Au: Please make sure you complete a Thrid Party Rights form (see my email) and return it to us so that we can seek permission to reproduce this figure.]**.

Figure 2 |

a | Although targeted testing is feasible with some PMD phenotypes, for many patients with non-syndromic clinical presentations this is not possible. This figure illustrates the approach taken to find

a genetic diagnosis for these individuals. 1. Common point variants are screened for in DNA extracted from blood leukocytes or uroepithelial cells. 2. More extensive study of mitochondrial DNA (mtDNA) is undertaken in DNA extracted from leukocytes that includes both sequencing of the mitochondrial genome and sometimes separate large-scale rearrangement and mtDNA depletion analysis (only in young children), with panel-based testing of relevant mitochondrial nuclear-encoded genes also occurring. This analysis is often followed by whole-exome sequencing if other investigations are negative, especially in children. 3. If no pathogenic variants are detected, further analysis, including repeat sequencing of the mitochondrial genome and mtDNA large-scale rearrangement analysis, is performed using mtDNA extracted from skeletal muscle tissue. In some cases step 2 and 3 proceed concomitantly. LR-PCR, long-range PCR.

Figure 3 | Single-molecule real-time sequencing and Nanopore sequencing.

a | 1) Single-molecule real-time (SMRT) sequencing (Pacific Biosciences) uses hairpin adaptors (orange), which render a double-stranded DNA molecule into a circular template. 2) The template is combined with a polymerase and primer molecule and loaded into a chamber, which also contains fluorescently labelled nucelotides. 3) The nucleotides are added to a complementary strand by the polymerase and fluorescence is emitted when the chamber is illuminated by the light source. 4) The emitted fluorescence is captured in real time by a camera. As the process occurs 'live' the timing between additions of each nucleotide can be monitored, enabling the indirect detection of epigenetic modifications, which extend the intervals between nucleotide additions. **b |** Nanopore sequencing (Oxford Nanopore Technologies) uses a nanopore embedded within a membrane. As a molecule passes through the pore, it perturbs an ionic current. This perturbation can be used to detect the base and any associated epigenetic modifications. In nanopore sequencing, adaptors are ligated to both ends of the DNA fragment and a motor protein is added to the 5' end adapter to help control the passage of the fragment through the nanopore.

Figure 4 | The role of long-read sequencing in mitochondrial medicine.

a| Long-read technologies can sequence the entire mitochondrial genome in a single read. As their accuracy continues to improve, these technologies could increasingly be used in the clinical workflow to identify point mutations and large-scale rearrangements in mitochondrial DNA. **b |** Long-reads can sequence through repetitive segments, which allows the size of repeat expansions to be measured. Resolving repeats is also important for the identification of structural variants, which are more common in repetitive areas. **c |** A simplified example of the use of long reads to link a translocation to its' new locus. **d |** Long-read technologies can detect epigenetic DNA modifications. **e |** Long-reads can resolve biallelic variants to specific alleles.

Figure 5 | The role of RNA sequencing in mitochondrial medicine.

a, b | RNA sequencing (RNA-seq) can provide insights into the effects of mutations on transcript splicing and differential gene expression. **c |** RNA-seq can identify mono-allelic expression, which may help focus analysis towards heterozygous variants in these genes. **d |** Studying the transcriptomes of patients may demonstrate alterations in specific pathways, thus providing information on potential disease pathophysiology.

Figure 6 | Sequencing of transfer RNA in mitochondrial medicine.

a | Given that mtDNA is transcribed as a single polycistronic unit, transfer RNA (tRNA) molecules are connected to other RNA molecules within the initial transcript prior to being released by specialised enzymes. **b–d |** RNA sequencing can provide insights into the effects of mutations on tRNA genes, including quantification of tRNA molecules at the single fibre level in muscle, characterising alterations in epigenetic modifications, and through the detection of tRNAs which have not been correctly excised from neighbouring gene transcripts.

Table 1 | Challenges to the diagnosis of primary mitochondrial diseases.

Challenge	**Implication**
Bigenomic sources of variants	Causative variants can lie in mtDNA or nDNA
Secondary mtDNA mutations	Mutations in mtDNA can occur secondary to pathogenic variants in nDNA genes that control the maintenance of mtDNA
Tissue heteroplasmy	Some variants can be lost from blood cells and only identifiable in stable tissues such as muscle.
Overlapping phenotypes	Although some specific genotype–phenotype correlations exist, many phenotypes overlap substantially, limiting the utility of targeted testing

mtDNA; mitochondrial DNA; nDNA, nuclear DNA

Table 2 Targeted testing for common point mutations in primary mitochondrial diseases.

Genotype	Phenotype	Specificity of genotype-phenotype
m.11778G>A, m.14484T>C, m.3460G>A	Leber Hereditary Optic Neuropathy (LHON)	Together these variants account for >90% of LHON, though variable penetrance occurs. [37]
m.3243A>G	Mitochondrial Encephalopathy, Lactic Acidosis, Stroke- like episodes (MELAS), Chronic Progressive External Ophthalmoplegia (CPEO) Maternally Inherited Deafness and Diabetes (MIDD)	The m.3243A>G mutation accounts for ~ 80% of MELAS cases and >85% of MIDD cases. However it is a rare cause of CPEO (~8%), which is typically associated with additional symptoms[140,160–163]
m.8344A>G	Myoclonic epilepsy with ragged-red fibres (MERRF)	The m.8344A>G mutation underlies ~80% of MERRF[164]
m.8993T>G/C	Maternally-inherited Leigh syndrome (MILS), Neurogenic muscle weakness, ataxia, and retinitis pigmentosa (NARP)	The m.8993T>G/C mutations causes ~10% of MILS and NARP [165,166]

Table 3 | Potential clinical applications of genomic and transcriptomic advances [Au: This was

Technique	Potential clinical applications
Whole genome sequencing	Concomitant analysis of nDNA and mtDNA in blood, when appropriate[a]
Long-read technologies	Sequencing of mtDNA for point mutation and rearrangement analysis
	In nDNA, identification of new variant types and improved phasing
	Identification of epigenetic changes
RNA sequencing	Identification of new diagnoses
	Use in functional work, including sequencing of tRNA

[a] Has already been introduced into clinical practice. mtDNA, mitochondrial DNA; nDNA, nuclear DNA; tRNA, transfer RNA.

Key points

- At present, diagnosis of primary mitochondrial diseases is a multi-step process often involving a number of time-consuming and highly manual molecular techniques.
- In appropriate patients, early whole genome sequencing of blood, analysing both mitochondrial and nuclear DNA, is likely to improve diagnostic efficiency.
- In the future, the application of long-read sequencing to mitochondrial DNA could build on the advances made by next generation sequencing to further improve coverage, and to enable the identification of large-scale rearrangements and point mutations in a single test.
- As with other rare diseases, whole genome long-read sequencing might provide the next diagnostic uplift as, compared with short-read sequencing, it has superior ability to identify structural variants, short tandem-repeat variants, epigenetic modifications and phase compound

heterozygous variants.

- Mitochondrial medicine is poised to benefit substantially from the increasing use of RNA sequencing of tissue samples;advances in pre-processing and sequencing of transfer RNA are enabling new insights into this molecule, which plays an outsized role in these disorders.

Glossary

Polycistronic transcript: A transcript that contains the code for more than one polypeptide.

mtDNA maintenance: continuous re-synthesis of mtDNA by a nuclear-encoded replication apparatus supported by a sustained pool of mitochondrial nucleotides.

Post-mitotic tissue: tissues, such as muscle, that are terminally differentiated and no-longer replicate and therefore are more likely to retain pathogenic mtDNA variants than mitotic tissue.

Deep Sequencing: sequencing a DNA locus many more times than in standard NGS; this allows low levels of alternative alleles (heteroplasmy and mosaicism) to be identified. (Standard WES is 100x, standard WGS is 20-30x.)

Next Generation Sequencing (NGS): process by which DNA is fragmented into short molecules and denatured, millions of sequencing reactions (addition of fluorescence-labelled nucleotides to form a complementary strand) then occur concurrently and the short sequences or “reads” generated are mapped to a reference genome.

Short-reads: the fragments of genetic sequence generated in NGS; typically ~150bp in length.

Structural variants: large genetic variants such as copy number variants (deletions and duplications), inversions, and translocations, typically >1,000bp.

Epigenetic modifications: chemical modifications to DNA or the histone molecules around which DNA is packaged; they do not change the genetic code, but can alter gene expression.

Phase: the homologous chromosome of origin (either maternal or paternal)

Whole-exome trio: sequencing and comparison of the coding DNA of an affected proband and their unaffected parents.

Restriction fragment length polymorphism: differences between individuals in the length of DNA fragments produced by restriction enzymes; the presence of a mutation can create or remove a restriction site.

mtDNA large-scale rearrangements:

rearrangements, typically deletions and/or duplications, of >1,000bp in mitochondrial DNA.

Long-range PCR:

uses specialised polymerase to amplify mtDNA as one or two fragments; traditional PCR amplifies shorter fragments of DNA

Coverage: refers to the adequate sequencing of a locus; targeted sequencing can have poor uniformity of coverage

GC content:

proportion of bases that are Guanine–Cystosine.

Phenocopies:

diseases with clinical presentations that overlap substantially with the disease of interest

De novo assembly:

assembly of reads into a continuous sequence without the need to align them against a reference sequence.

Homopolymeric regions:

sequences of DNA comprising identical repeated units of sequence

Contigs:

Consensus sequences comprising overlapping short sequence reads

Anticodon

the three-nucleotide sequence in tRNA, which is complementary to a codon in mRNA

Wobble position

the third nucleotide in the anticodon; the Watson–Crick base pairing here is less specific than usual and atypical pairing can occur.

Cybrid

cell lines are created by fusing an enucleated cell (only containing mtDNA) with a nucleated cell, which can contain nDNA and mtDNA or be modified to contain only nDNA.

Adaptor ligation: a short synthetic DNA molecule added to the end of the DNA fragment to enable sequencing of that fragment.

Imprinted genes: genes that are expressed from only one parental origin; the silenced parental copy is said to be ‘imprinted’.