Use of Whole-Genome Sequencing for Mitochondrial Disease Diagnosis

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Neurology® Published Ahead of Print articles have been peer reviewed and accepted for publication. This manuscript will be published in its final form after copyediting, page composition, and review of proofs. Errors that could affect the content may be corrected during these processes.
Authors:
Ryan L Davis, PhD1,2; Kishore Raj R Kumar, MBBS, FRACP, PhD1,2,3,4; Clare Puttick, MStat2; Christina Liang, MBBS, FRACP, PhD1,2; Kate E Ahmad, MBBS, FRACP1,3; Fabienne Edema-Hildebrand, Registered Nurse1,3; Jin-Sung Park, PhD3; Andre E Minoche, PhD2; Velimir Gayevskiy, PhD2; Amali C Mallawaarachchi, MBBS, FRACP, PhD2; John Christodoulou, MBBS, FRACP, PhD5,7; Deborah Schofield, PhD5; Marcel E Dinger, PhD2,9; Mark J Cowley, PhD2; Carolyn M. Sue, MBBS, FRACP, PhD1,2,3

Corresponding Author:
Carolyn M. Sue, carolyn.sue@sydney.edu.au

Affiliation Information for All Authors: 1. Department of Neurogenetics, Kolling Institute, Faculty of Medicine and Health, University of Sydney and Royal North Shore Hospital, Northern Sydney Local Health District, St Leonards, NSW, Australia; 2. Translational Genome Informatics Group, Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia; 3. Department of Neurology, Royal North Shore Hospital, Northern Sydney Local Health District, St Leonards, NSW, Australia; 4. Current address: Molecular Medicine Laboratory, Concord Hospital, Concord, NSW, Australia; 5. Current address: Cenyx Biotech, Jongno-gu, Seoul, Republic of Korea; 6. Brain and Mitochondrial Research Group, Murdoch Childrens Research Institute, Parkville, Melbourne, VIC, Australia; 7. Department of Paediatrics, University of Melbourne, Melbourne, VIC, Australia; 8. Current address: GenIMPACT: Centre for Economic Impacts of Genomic Medicine, Macquarie University, Macquarie Park, NSW, Australia; 9. Current address: School of Biotechnology and Biomolecular Sciences, University of New South Wales, Randwick, NSW, Australia; 10. Current address: Computational Biology Group, Children’s Cancer Institute, University of New South Wales, Randwick, NSW, Australia

Equal Author Contribution:
Co-first authors: Ryan L. Davis, Kishore R. Kumar, Clare Puttick. Co-senior authors: Mark J. Cowley, Carolyn M. Sue.

Contributions:
Ryan L Davis: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data
Kishore Raj R Kumar: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data
Clare Puttick: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data
Christina Liang: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Analysis or interpretation of data
Kate E Ahmad: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data
Fabienne Edema-Hildebrand: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Analysis or interpretation of data
Jin-Sung Park: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data
Andre E Minoche: Major role in the acquisition of data; Analysis or interpretation of data
Velimir Gayevskiy: Major role in the acquisition of data; Analysis or interpretation of data
Amali C Mallawaarachchi: Major role in the acquisition of data; Analysis or interpretation of data
John Christodoulou: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data
Deborah Schofield: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data
Marcel E Dinger: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data
Mark J Cowley: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data
Carolyn M. Sue: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology.
Abstract

Objectives: Mitochondrial diseases are the commonest group of heritable metabolic disorders. Phenotypic diversity can make molecular diagnosis challenging and causative genetic mutations may reside in either mitochondrial or nuclear DNA. A single comprehensive genetic diagnostic test would be highly useful and transform the field. We applied whole genome sequencing to evaluate the variant detection rate and diagnostic capacity of this technology with a view to simplifying and improving the mitochondrial disease diagnostic pathway.
Methods: Adult patients presenting to a specialist mitochondrial disease clinic in Sydney, Australia were recruited to the study if they satisfied clinical mitochondrial disease (Nijmegen) criteria. Whole genome sequencing was performed on blood DNA, followed by clinical genetic analysis for known pathogenic mitochondrial disease-associated variants and mitochondrial mimics.

Results: Of the 242 consecutive patients recruited, 62 subjects had ‘definite’, 108 had ‘probable’ and 72 had ‘possible’ mitochondrial disease classification by the Nijmegen criteria. Disease causing variants were identified for 130 subjects, regardless of the location of the causative genetic mutations, giving an overall diagnostic rate of 53.7% (130/242). Identification of causative genetic mutations informed precise treatment, restored reproductive confidence and optimised patient management.

Conclusion: Comprehensive bigenomic sequencing accurately detects causative gene mutations in affected patients and simplifies mitochondrial disease diagnosis, enables early treatment and informs the risk of genetic transmission.

Introduction
Mitochondrial diseases (MDs) are the commonest group of inherited metabolic disorders, and novel therapies in the field are now beginning to emerge. However, targeted treatments and reproductive options rely on a precise molecular diagnosis. Limited genotypic-phenotypic correlation of MDs makes molecular confirmation challenging and many patients remain undiagnosed despite extensive investigation over long periods of time.
MDs are unique because they can be caused by mutations in either the mitochondrial or nuclear genome and can affect both children and adults. Although minimum prevalence studies have estimated that 1 in 4300 live births develop MD, community-based prevalence studies demonstrate that at least 1:250 people carry a pathogenic mitochondrial DNA (mtDNA) variant that puts them at risk of developing a MD. This highlights that a large percentage of at-risk individuals carrying disease causing variants are undiagnosed or remain asymptomatic. Clinical severity of affected patients ranges from mild, oligosymptomatic disease to severe, fatal illness. Presenting features in adults emerge at variable ages of onset, commonly including muscle weakness, fatigue, ptosis, ophthalmoplegia, hearing loss, diabetes, seizures, focal neurological deficits, and visual loss. Standard diagnostic criteria are based on available clinical data and include results of invasive procedures, such as muscle biopsy, but a precise diagnosis requires genetic testing.

Currently, there is no single first-line genetic test integrated into standard mitochondrial disease diagnostic practice. A definitive diagnosis may require sequencing of hundreds of nuclear genes and the majority of the mitochondrial genome, a process that has been impractical to date. To complicate matters further, causative mtDNA variants in blood decline with age, so may be absent or only present at low levels of heteroplasmy (the proportion of variant to wild type mtDNA genomes), necessitating sampling of other tissues, which may be invasive. As a result, affected adult patients often endure prolonged diagnostic odysseys prior to achieving a genetic diagnosis, delaying the benefits of informed family planning and optimal medical
management\textsuperscript{10,15,16}. Undiagnosed, oligosymptomatic adult carriers may unknowingly pass on the disease to their children\textsuperscript{21} or receive inappropriate clinical treatment\textsuperscript{22}.

Now, whole genome sequencing (WGS) provides the capability to comprehensively sequence both nuclear and mitochondrial genomes simultaneously, with the potential to capture the complete spectrum of MD-causing variants in a single blood test, and thereby simplify the diagnostic pathway\textsuperscript{15,23}. Capitalizing on the ability of WGS to provide a high depth of coverage of the mitochondrial genome, we developed a bioinformatic tool\textsuperscript{24} capable of identifying low levels of heteroplasmic mtDNA variants often found in blood. Accordingly, we determined the capacity for WGS to identify known disease-causing variants in both nuclear and mitochondrial genomes and examined the diagnostic utility of WGS as if it were applied as a front-line “genetics first”\textsuperscript{25} blood test for a large cohort of patients with suspected MDs.

**Patients and Methods**

**Standard protocol approvals, registrations and patient consents**

All patients gave written consent to participate in the study, which was approved by the Northern Sydney Local Health District Human Research Ethics Committee (HREC/10/HAWKE/132). All data was de-identified.

**Patient recruitment and samples**

We prospectively recruited 242 consecutive patients reviewed at the Mitochondrial Disease Clinic, Royal North Shore Hospital, Sydney, Australia between 2014-2020. Patients were
eligible for recruitment to the study if they satisfied ‘possible’, ‘probable’ or ‘definite’ Nijmegen mitochondrial disease criteria\textsuperscript{17}. Although the Nijmegen criteria were developed in children with primary mitochondrial disease, they were used here as the clinical manifestations included in these diagnostic criteria reflect the phenotypic variability observed in patients with mitochondrial disease.\textsuperscript{9}

DNA samples from the blood of 41 subjects with known pathogenic variants (30 in nDNA and 11 in mtDNA) and the muscle tissue of a patient with Kearns-Sayre Syndrome (KSS) with a 4.6kb mtDNA deletion were used as ‘positive controls’ for assessing the capacity of WGS to identify known variants in the two different genomes (eTable 1).

**Whole Genome Sequencing and analysis**

Total genomic DNA was isolated from peripheral blood using standard methods. Sequencing libraries were prepared using robotic instrumentation and sequenced on an Illumina HiSeq X platform at the Kinghorn Centre for Clinical Genomics, Sydney, Australia. 2×150bp reads, yielded at least 110Gb of raw sequencing data and a minimum 30× coverage of nDNA per lane. To determine the clinical utility of the test, initial variant analysis was performed blinded and regardless of clinical phenotype, family history, or prior known genetic results.

**Nuclear DNA analysis**

We detected small nDNA variants using a GATK best practices pipeline\textsuperscript{26} and interpreted them using our nuclear variant filtering analysis platform, Seave (https://seave.bio/).\textsuperscript{27} Raw fastq files were aligned to the hs37d5 reference genome using BWA-MEM (v0.17.10-r789), with resulting
BAM file duplicate reads marked using Novosort (default settings) and read alignment improved using GATK Indel Realignment (v3.3)\textsuperscript{26}. SNVs and short indels (<50bp) were identified using GATK HaplotypeCaller, GenotypeVCFs and VQSR (v3.3)\textsuperscript{26}, annotated with VEP (v87), converted into a GEMINI (v0.11.0) database, and imported into Seave\textsuperscript{27} for filtration and prioritization. We detected structural variation (SV) and copy number variation (CNV) from 50bp to whole-chromosome aneuploidy, in the nuclear and mitochondrial genomes, using ClinSV\textsuperscript{28}. Data analysis was performed using R (v3.6.0) via RStudio (v1.2.1335) and plotted using ggplot2.

To restrict the search-space for nDNA variant analysis, we curated a panel of 249 MD genes (eTable 2), 400 neuromuscular disease genes (Orphanet, https://www.orpha.net; accessed May 2015), and for unsolved cases, additional tailored individual searches based on clinical phenotype were made in optic atrophy, metabolic, developmental disorder and other gene panels relevant to the phenotype. Variants were classified using the American College of Medical Genetics and Genomics (ACMG) 2015 guidelines\textsuperscript{29}, with consideration of ‘pathogenic’ or ‘likely pathogenic’ variants. Variants of uncertain significance (VUS) bordering on, but insufficient to be classified as ‘likely pathogenic’, were classed as ‘VUS–favor pathogenic’\textsuperscript{29} (eTable 3) and included in the diagnostic count due to the difficulty in obtaining supporting evidence for a class IV variant. Pathogenic variants were confirmed using Sanger sequencing of an alternate DNA sample on an ABI3100 using the BigDye® Xterminator Kit (Garvan Molecular Genetics, Garvan Institute, Sydney, Australia). Segregation studies were carried out where possible for newly identified variants.
Mitochondrial DNA analysis

To analyze single nucleotide variants (SNVs) and insertion/deletion (indel) variants in mtDNA, we developed an analytical pipeline named ‘mity™’, which runs FreeBayes in an ultra-sensitive mode, and calculates variant quality accurately even for very low heteroplasmic variants. mity™ was developed using 13 replicates of the NA12878 control line, 2,570 healthy controls and one patient from this study with two independent genomic sequences. We optimized the analytical parameters as follows: reads with mapping quality <30 were removed to minimize false-positive variants and spurious signals from nuclear mitochondrial DNA (NUMT), only bases with base quality >=24 were used for variant calling, and we required a variant to have at least 10 supporting reads or a variant allele frequency (VAF; i.e., the fraction of reads carrying the variant versus all reads) >1. We used the VAF as a direct measure of variant heteroplasmy. For variant interpretation, all mitochondrial variants were ordered by decreasing VAF, prioritizing known pathogenic variants and those linked to phenotypes in MITOMAP and the literature.

To determine the capability of ClinSV analysis of WGS to identify and quantitate mtDNA deletions, we studied DNA extracted from muscle taken at autopsy from a patient with KSS (Sample 42, eTable 1).

Pyrosequencing

To evaluate the capability of mity™ in determining mtDNA variant heteroplasmy from WGS, we compared heteroplasmy determined by mity™ and quantitative pyrosequencing from 60 samples (50 blood samples from individual patients and 10 autopsy tissue samples from two patients).
known to have variable levels of the m.3243A>G variant. A custom m.3243A>G pyrosequencing assay was performed by the Australian Genome Research Facility (Perth, Australia). A standard curve was created using wild type or mutant gBlocks® gene fragments of 500bp region around m.3243 (Integrated DNA Technologies, Singapore).

**Long-Range Polymerase Chain Reaction**

To confirm mitochondrial DNA deletions were present in urine sedimentary cell DNA and absent in blood DNA, we amplified mtDNA as one full length fragment using overlapping primers and TaKaRa LA Taq, as previously described.

**Data Availability**

Patients were consented for genomic testing in a clinical setting, and did not consent for the release of raw or processed genomic data. mity™ is available under an open source MIT license, from [https://github.com/KCCG/mity](https://github.com/KCCG/mity)

**Results**

**Patient cohort**

We recruited 242 patients (149 females and 93 males; eFigure 1) with a mean age at DNA sampling of 49.5±16.8 years. According to the Nijmegen mitochondrial disease criteria, 62 subjects were classified as definite, 108 as probable and 72 as possible.
Whole Genome Sequencing capabilities

Coverage of nDNA and mtDNA in blood versus other tissues

WGS provided a high depth of coverage of both nuclear and mitochondrial genomes from blood DNA. A mean nuclear genome coverage of 30–40× was achieved (Figure 1A), with 80% of the genome covered to ≥10×. WGS simultaneously provided 3,000–4,000× mean coverage of the mitochondrial genome (Figure 1B), with >90% of the genome covered to ≥2000×. Using our analytical pipeline with mity™, we were able to detect very low levels (<1%) of heteroplasmic mtDNA variants. Levels of m.3243A>G heteroplasmy quantified by mity™ analysis of WGS strongly correlated with pyrosequencing (n=50, R²=0.994, Figure 1C), with detection of the pathological variants down to a heteroplasmic load of 0.35% (patient E53, eTable 4), well below the reliable limit of detection for pyrosequencing (~5%)33. While these ultra-low levels of heteroplasmy may be difficult to interpret clinically in a de novo situation, the fact they can be detected in blood to this level shows the sensitivity of WGS variant detection in blood but would require validation of the variant in another tissue to confirm the genetic diagnosis.

Analysis of blood and post-mortem tissues (n=12) from two patients who died with m.3243A>G, showed that the sequencing depth of mtDNA ranged from ~3,000× in blood to between ~20,000-90,000× in solid tissues with variable but high levels of heteroplasmy (Figure 2A-D).

Diagnostic yield

Using WGS and applying mity, we identified 57 patients with disease-causing nDNA variants and 73 patients with disease-causing mtDNA variants (Figure 3A, eFigure 1), obtaining an overall diagnostic yield of 53.7% (130/242; 95% confidence interval (95% CI) – 47.2–60.1%).
Data on the clinical features, family history, Nijmegen mitochondrial disease criteria classification and variants identified are summarized in eTables 3-4.

**Detection and impact of pathogenic variant calling of nuclear DNA encoded MDs.**

57 patients were found to have nuclear gene mutations, with pathogenic or likely pathogenic variants located in *AFG3L2, AMACR, MFN2, OPA1, POLG, SPG7, TWNK, TYMP, WFS1,* and *YARS2* (Figure 3B, eFigure 1, eTable 3). In addition, 11 patients had VUS identified, requiring further investigation of pathogenicity, but occurring in known MD-associated genes (eTable 3).

**Nuclear genome copy number variations**

Using *ClinSV*\(^{28}\), we were also able to detect CNVs and SVs, including a novel heterozygous 16.4Mb *de novo* deletion of chromosome 4q26–q28.3 (Figure 3C), in a proband presenting with seizures, ophthalmoparesis, optic atrophy, ataxia, myopathy, diabetes and recurrent pseudo-obstruction of the bowel (patient B3, eTable 3). The deletion encompassed numerous genes including *PRSS12* (Mental retardation, autosomal recessive 1, MIM 249500), *MRT29* (Mental retardation, autosomal recessive 29, MIM 614333), and *SPATA5* (Epilepsy, hearing loss, and mental retardation syndrome, MIM 616577), without a detectable mutation on the alternate allele. This finding was confirmed using comparative genomic hybridization array and confirmed to be absent in the proband’s parents. We also detected a homozygous exon 6 deletion in *SPG7* (Figure 3D) in a proband with spastic paraplegia complicated by cerebellar ataxia, ophthalmoplegia, and sensory neuropathy (patient E75, eTable 3).
Detection and impact of pathogenic mtDNA encoded MDs

We identified 73 patients with disease-causing mtDNA variants (Figure 3A, eFigure 1). Using mity™, we were able to confidently detect a broad range of mtDNA point mutations in blood DNA, (Figure 4A, eTable 4), even if present at low levels of heteroplasmy (Figure 4B).

Using ClinSV²⁸ to identify and quantitate mtDNA deletions, we detected a 4.8kb mitochondrial DNA deletion at 16% heteroplasmy (3,780/24,080 sequencing reads; Figure 4C and D) in the control sample derived from the muscle tissue of a patient with KSS (sample #42, eTable 1). We were also able to detect single 8kb and 5kb mtDNA deletions in blood DNA of two additional patients at extremely low heteroplasmic loads of 0.29% (7/2435 reads) and 0.61% (24/3917 reads) respectively (Figure 4C; patients A10 and C16, eTable 4) and confirmed that they also had mtDNA deletions in other tissues, e.g. muscle or urine (data not shown).

In seven patients with clinical features of CPEO, mtDNA deletions were detectable in muscle or urine using Southern blot or long-range PCR³² (eTable 5). These deletions were not detected by WGS of blood DNA, presumably because they were not present in this tissue (eFigure 2). This is consistent with the known selection against mtDNA deletions in blood²⁰.

Diagnostic rates according to clinical phenotype and age

We found that the diagnostic rate varied depending on the presenting clinical phenotype (Figure 5), rather than disease classification using the Nijmegen criteria (eFigure 3A). The highest
diagnostic rates were achieved when patients with suspected MD presented with clear clinical phenotypes (Figure 5). For individuals with optic atrophy, 23 out of 24 (95.8%; 95% CI = 79.8%-99.3%) were diagnosed using WGS (n=17 nDNA encoded MDs; n=6 mtDNA encoded MDs). In patients presenting with stroke-like episodes, 17 out of 28 (60.1%; 95% CI = 42.4%-76.4%) were diagnosed (n=3 nDNA encoded MDs; n=14 mtDNA encoded MDs) and 35 out of 67 patients with a CPEO phenotype (52.2%; 95% CI = 40.5%-63.8%) also had a molecular cause identified (n=31 nDNA-encoded MDs; n=4 mtDNA-encoded MDs). Thirteen patients with m.3243A>G in our cohort had MIDD. Diagnostic rates for nonsyndromic complex phenotypes (defined as >5 clinical features listed in the Nijmegen criteria) and oligosymptomatic phenotypes (defined as <5 clinical features listed in the Nijmegen criteria), were lower (10/43 complex = 23.3%; 95% CI = 13.2%-37.8% and 26/61 oligosymptomatic = 42.6%; 95% CI = 31.0%-55.1%) (Figure 5). The diagnostic rate using our WGS protocol was higher in patients under 50 years (odds ratio 2.29; 95% CI = 1.36-3.84, p<0.002; eFigure 3B).

Clinical impact of definitive genetic diagnoses confirmed by WGS

Our approach led to genetic diagnoses that change clinical management in patients (e.g. commencement of disease-specific clinical care, avoidance of disease-specific contraindicated care and clarification of reproductive options) with both nuclear and mitochondrial DNA encoded disorders. WGS identified patients with treatable MDs including MNGIE syndrome (n=3; treatable by allogeneic bone marrow transplantation34) and Leber’s Hereditary Optic Neuropathy (n=6; idebenone or potential gene therapy treatment), as well as identifying patients with mutations in POLG, whereby recommendations to avoid contraindicated medications, such as valproic acid that can cause fulminant liver failure or life-threatening status epilepticus, are
important for optimal patient management\textsuperscript{35,36}. Additionally, a 33 year-old female (patient C44, eTable 3) with ptosis, optic atrophy and proximal muscle weakness who was found to have compound heterozygous mutations in \textit{YARS2} causing MLASA2, became pregnant with the confidence from WGS findings that she would be highly unlikely to transmit MD to her child. Thus, we were able to provide certainty for patients making reproductive decisions by obtaining definitive genetic diagnoses.

Regarding MD mimics, WGS was able to detect variants in a larger targeted gene list, thereby allowing diagnosis and differentiation of patients with other treatable conditions, enabling appropriate care and treatment for their respective diseases, whilst ruling out MD. Three patients were diagnosed with \(\alpha\)-methylacyl-CoA-racemase (AMACR) deficiency\textsuperscript{37,38}, inclusive of two sisters (Table 1; patients B45 and E19, eTable 3) who had seizures, encephalopathy and stroke-like episodes suggestive of MD. With the confirmed genetic diagnosis of AMACR deficiency they were treated with dietary restriction of pristanic acid\textsuperscript{37,38} that resulted in symptomatic improvement. Their cerebral MRIs showing T1 high signal intensities in the right parieto-occipital cortical ribbon or bilateral thalami prior to treatment, were reported to be suggestive of the diagnosis of a MD (Figure 6). Furthermore, two siblings (Table 1; patients B42 and B43, eTable 3) who presented with progressive external ophthalmoplegia and proximal muscle weakness were found to have a novel homozygous splicing mutation in the \textit{MUSK} gene (c.358+3G>T; eFigure 4) and were subsequently treated with salbutamol\textsuperscript{39}.
Discussion

WGS comprehensively and simultaneously sequenced both mitochondrial and nuclear genomes to a high depth of coverage from blood DNA, and in combination with mity™ we were able to identify a range of SNVs, indels and CNVs in both genomes to achieve precise genetic diagnoses for a broad spectrum of MDs and MD mimics. When applied as a first-line diagnostic blood test for MD, WGS achieved an overall diagnostic rate of 53.7%, a result that compares favourably to previous genetic disease cohorts sequenced using other next generation sequencing methods, despite their enrichment with more stringent selection criteria. Importantly, our findings demonstrate the simplicity of our comprehensive bi-genomic sequencing diagnostic approach, that for the majority of cases uses DNA from blood and alleviates the need for muscle biopsy or obtaining DNA from other tissues.

WGS provides substantial advantages over targeted mitochondrial sequencing panels, which are less comprehensive, provide lower diagnostic rates, and would require DNA from muscle or urine to achieve the same detection rates seen here. Whilst whole exome sequencing largely provides adequate coverage of nDNA protein coding exons, the average coverage of mtDNA is much lower (~50×) and thus less sensitive when compared to WGS. Whole exome sequencing of nDNA and mtDNA is performed in parallel and is subject to incomplete coverage and target enrichment bias during library preparation, when compared to WGS. Of further benefit, analysis of WGS provides more capability when identifying CNVs and SVs, which can be challenging when using targeted sequencing panels or whole exome sequencing. The superior detection sensitivity of WGS, in combination with mity™, is of particular importance for adults with MD, as our study shows that the majority of diagnosed patients (73/130; 56.2%) had pathogenic
mtDNA variants, rather than nDNA variants (Figure 3A). However, when there is a strong maternal inheritance pattern and typical clinical phenotype indicative of a specific common mtDNA mutation, alternatives, such as RFLP analysis for common mtDNA mutations (e.g. m.3243A>G or m.8344A>G) or full mtDNA sequencing, may be more cost-effective, although careful tissue selection to address the issues of low heteroplasmy needs to be considered. Once a variant has been identified by WGS, cascade testing in relatives can also be approached using diagnostic targeted sequencing, although these methods cannot always provide an estimate of the mutational load of the pathogenic mitochondrial variant involved.

Diagnostic rates were further increased (up to 95%) when patients were stratified by deep clinical phenotyping, underpinning the critical value of clinical expertise and assessment in combination with our WGS pipeline. We demonstrate that a high index of clinical suspicion and knowledge of specific clinical phenotypes, such as CPEO, optic atrophy, and stroke-like episodes (MELAS) that all demonstrate genetic heterogeneity across both genomes, justifies simultaneous dual genomic analysis with WGS (Figure 5, eTables 3-4). WGS diagnostic rates differed depending on the specific phenotype; patients presenting with optic atrophy, MELAS, CPEO or MIDD had a higher diagnostic rate than those who presented with non-syndromic, complex phenotypes, possibly indicating that patients with MD mimics may still satisfy standard clinical diagnostic criteria (Figure 5).

An important limitation highlighted here when using WGS on DNA extracted from blood is that mtDNA deletions may not be present\(^\text{20}\) and therefore are not able to be detected using this readily obtainable and often used tissue (Figure 5, eFigure 2). Thus, to increase detection of deletions in
patients with CPEO or KSS phenotypes, where single or even multiple mtDNA deletions are suspected, WGS or long-range PCR of an alternative tissue, such as muscle, saliva or urine, may be required if initial sequencing of blood fails to identify a causative mutation or if only a low level heteroplasmic mtDNA deletion is identified by WGS. Despite this caveat, our study still diagnosed 52.2% of patients with CPEO using WGS, showing the value of initial testing of DNA sourced from blood.

In our analysis, we also considered variants in neuromuscular disease-associated genes and identified patients presenting to our clinic who had disorders mimicking MD (e.g. congenital myasthenia and neuroacanthocytosis), even though these patients had muscle biopsy abnormalities and clinical symptomology consistent with the diagnosis of a MD (Table 1). The provision of a molecular diagnosis and confirmation of a MD mimic led to changes in medical management (Table 1), as well as informing the risk of transmission of their disorder to their offspring. In addition, three patients with neurological presentations (focal neurological deficits, seizures associated with abnormalities on cerebral MRI) suggestive of MELAS or Leigh-syndrome were identified as having AMACR deficiency\textsuperscript{e1} (Figure 6, eTable 3), demonstrating that WGS was able to change clinical management by identifying disorders that are treatable by simple dietary restriction\textsuperscript{38}, as well as providing a diagnosis for MD phenocopies that have treatment options (Table 1).

Identification of the precise genetic cause of MD is clinically important and clarifies reproductive options for affected patients and their families. For instance, patients with nDNA mutations can undergo prenatal genetic diagnosis, whereas those with mtDNA mutations are
now able to consider the novel in-vitro fertilization option, mitochondrial donation\textsuperscript{21,48}. This advantage of WGS is further underpinned by its ability to quantify mtDNA heteroplasmy in blood, as this provides predictive information regarding disease transmission in mtDNA disorders\textsuperscript{21}.

A limitation of this study was that we conservatively restricted our variant calling to known pathogenic mutations and those that fulfilled stringent pathogenicity classification criteria. Whilst the diagnostic rate was high, a number of patients still remain undiagnosed. These patients may have mutations in novel disease genes that are yet to be discovered or associated with MD. Furthermore, subsequent analysis may reveal disease-causing mutations in non-coding regions, as atypical splice variants, or as tissue-specific mtDNA mutations. Other patients may have VUS in known disease genes that require confirmation of pathogenicity with functional studies (eTable 3). Moreover, mtDNA mutations (particularly deletions) that typically disappear in blood with advancing age\textsuperscript{19,20}, may be more difficult to detect. Re-analysis with updated variant and gene lists and further functional genomic investigation of novel variants has the potential to identify additional diagnoses in the future. Where parents can be recruited, trio analysis of WGS may further increase the diagnostic yield through improved filtering of autosomal recessive or de novo nuclear disorders\textsuperscript{49}. However, this is often challenging with adult mitochondrial disease patients and familial segregation studies may be the only option, although these can also be challenging for various reasons (disease penetrance, insurance considerations, and deniability by family members).
In this study, we did not compare the diagnostic utility of WGS from different tissues, such as muscle or urine, or to other diagnostic methods such as limited gene panels, WES and long-range PCR and whole mtDNA genome sequencing. Rather, we applied the standard clinical situation to determine the capability of WGS when used as a single diagnostic test that could be standardized, scaled and widely implemented. Although direct gene testing or whole mitochondrial genome sequencing may lead to confirmation of a genetic diagnosis in some patients who have typical clinical presentations, given the limited genotype-phenotype correlation of this disorder, WGS could prove to be more cost effective and may become preferential as costs decrease, especially when considering the high diagnostic rate observed in this study. At present, the cost of WGS is falling and will become more accessible, requiring evidence and analytical pipelines, as presented here, to support its routine clinical uptake for mitochondrial disease diagnosis in the future. Further evaluation of the cost effectiveness of WGS compared to conventional genetic testing methods (e.g. targeted mtDNA mutation, single gene or gene panel analysis) will be important as the benefits of a single diagnostic blood test to inform directed genetic testing in an extensive number of family members are considerable in terms of time to diagnosis and the costs of testing.

Conclusions

Comprehensive simultaneous sequencing of both mitochondrial and nuclear genomes by WGS is an accurate and non-invasive test to diagnose patients with MDs and has the capability to transform the MD diagnostic pathway. Improvements in health outcomes from early genetic diagnosis, appropriate intervention and treatment, avoidance of adverse events, reduced costs of inappropriate therapy, and potential to prevent disease inheritance are all advantages that could
be enabled by the introduction of our WGS analysis pipeline and emphasizes the benefits for integrating WGS into future clinical practice.

**Author information**

**RLD**, Conceptualization, Methodology, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization, Supervision, Project Administration, Funding Acquisition; **KRK**, Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization, Supervision, Project Administration, Funding Acquisition; **CP**, Software Development, Formal Analysis, Investigation, Resources, Data Curation, Writing – Review and Editing, Visualization; **CL**, Investigation, Resources, Writing – Review and Editing; **KEA**, Investigation, Resources; **FE-H**, Investigation, Resources, Writing – Review and Editing; **J-SP**, Formal Analysis, Writing – Review and Editing; **AEM**, Software, Formal Analysis, Investigation, Resources, Writing – Review and Editing; **VG**, Software, Investigation, Resources, Writing – Review and Editing; **ACM**, Formal Analysis, Writing – Review and Editing; **JC**, Writing – Review and Editing, Funding Acquisition; **DS**, Writing – Review and Editing, Funding Acquisition; **MED**, Resources, Writing – Review and Editing, Funding Acquisition; **MJC**, Conceptualization, Methodology, Software Development, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization, Supervision, Project Administration; **CMS**, Conceptualization, Methodology, Formal Analysis, Investigation, Resources, Writing – Original Draft, Writing – Review and Editing, Supervision, Project Administration, Funding Acquisition.
References


**Figure Legends**

**Figure 1:** Whole genome sequencing (WGS) provides deep, uniform, coverage of both the mitochondrial and nuclear genome. The average depth of sequencing coverage across the (A) nuclear and (B) mitochondrial genomes from a cohort of adult subjects with suspected mitochondrial disease (n=242). (C) m.3243A>G variant heteroplasmy assessed by WGS and pyrosequencing for DNA extracted from blood (black) from n=50 patients diagnosed with the m.3243A>G variant and multiple autopsy tissues (n=10; coloured points) from subjects E9 and E59. Heteroplasmy was highly correlated between WGS and pyrosequencing (R^2=0.994).

Colours in (C) are consistent with those in Figure 2 for the different tissues.
Figure 2: Different tissues provide high depth of coverage of the mitochondrial genome and high levels of heteroplasmy. (A and C) Depth of sequencing coverage across the mitochondrial genome and (B and D) m.3243A>G heteroplasmy varied between autopsy tissues from two subjects with m.3243; being considerably higher in solid tissues compared to blood. Coloured lines in A and C are consistent with coloured bars in B and D. Colours in A-D are consistent with those in Figure 1C for the different tissues.
Figure 3: The diagnostic performance of whole genome sequencing (WGS) for nuclear DNA based mitochondrial diseases. (A) The number of diagnoses made by WGS in the nuclear or mitochondrial genomes, from n= 242 patients with suspected mitochondrial disease. (B) The frequency of diagnoses made from mitochondrial disease genes in the nuclear genome. (C) WGS identified a hemizygous 16.4 megabase deletion on chromosome 4 in a patient with seizures, mild ophthalmoparesis, optic atrophy, cerebellar ataxia, myopathy, diabetes, recurrent pseudo-obstruction of the bowel, and liver dysfunction. (D) WGS identified a homozygous 946 base pair deletion in SPG7 in a patient with hereditary spastic paraplegia complicated by cerebellar ataxia, ophthalmoplegia, and sensory neuropathy.
Figure 4: The diagnostic performance of whole genome sequencing (WGS) for mitochondrial DNA based mitochondrial diseases. (A) The number of mtDNA mutations detected by WGS, and (B) their corresponding variant heteroplasmy. (C) Two mitochondrial DNA single deletions were identified in blood, both at heteroplasmy below 1%. One mitochondrial deletion was found in autopsy muscle at a much higher heteroplasmy of 15.7%, shown in more detail in (D), which highlights the sequencing coverage and reads spanning the deletion.
**Figure 5: Molecular and diagnostic heterogeneity within different MD phenotypes.** Patients were grouped into six clinical phenotypes (CPEO, Stroke-like episodes, Optic atrophy, Maternally inherited deafness and diabetes (MIDD), Complex and Oligosymptomatic) to highlight the frequency of pathogenic nuclear gene and mitochondrial DNA mutations identified by WGS. Precise molecular diagnoses are shown and the number of undiagnosed cases in each subgroup is shown in grey.
Figure 6: Cerebral Magnetic Resonance Imaging in α-methylacyl-CoA racemase deficient patients. (A) Representative axial T1 weighted images showing focal areas of abnormal signal intensity in the right parieto-occipital region following episodes of encephalopathy in patient E19. Note there are also small high signal intensities in the scattered throughout the subcortical white matter (B) Magenta arrows indicate high signal areas in the thalami consistent with adult-onset Leigh syndrome in patient B54.
Table 1: Case studies of mitochondrial disease or mitochondrial mimics for which a diagnosis led to a change in management

<table>
<thead>
<tr>
<th>ID</th>
<th>Diagnosis</th>
<th>Variants identified by WGS</th>
<th>Clinical features</th>
<th>Change in management</th>
</tr>
</thead>
<tbody>
<tr>
<td>E19</td>
<td>α-methylacyl-CoA-racemase deficiency</td>
<td>Homozygous variants in AMACR (NM_014324.5:c.154T&gt;C, NP_055139.4:p.Ser52Pro) (eTable 3)</td>
<td>49 year-old female who presented with seizures, encephalopathy, migraineous auras without headache (e.g. visual scotomas, right face and hand numbness, right hand weakness) and multiple focal neurological deficits (e.g. transient visual loss, transient hearing loss, numbness of the upper limbs, left homonymous hemianopia) associated with multiple lesions on cerebral MRI (Figure 6A). Sister (sample B45, eTable 3) presented with seizures and a stroke-like episode (aphasia) associated with bilateral thalamic lesions of cerebral MRI (Figure 6B). Pristanic acid levels were markedly elevated for both subjects (196-255µM; Normal &lt;2.5µM).</td>
<td>Commenced pristanic acid dietary restriction (Refsum diet), which resulted in a reduction of seizures and improvements in immediate attention and working memory, information processing speed, letter and semantic fluency, as documented on formal neuropsychometric testing.</td>
</tr>
<tr>
<td>B43</td>
<td>Congenital myasthenia</td>
<td>Homozygous variant in MUSK (NM_005592.3:c.358+3G&gt;T) (eTable 3). Mutations in this gene are associated with congenital myasthenic syndrome 9 (MIM 616325).</td>
<td>77 year-old male with proximal myopathy, ophthalmoplegia, and respiratory difficulties since his 40s. Eldest of 11 siblings from a consanguineous marriage. Sister had a similar presentation. His muscle biopsy showed ragged-red fibers. The homozygous variant affected a donor splice site and impacted on splicing, as indicated by transcript analysis from patient-derived fibroblasts (eFigure 4).</td>
<td>Genetic diagnosis ended a 20-year diagnostic odyssey, prompting a trial of salbutamol (albuterol) and recommendations to avoid agents that block the neuromuscular junction.</td>
</tr>
<tr>
<td>B39</td>
<td>McLeod neuroacanthocytosis</td>
<td>Nonsense variant in the XK gene (NM_021083.2:c.268delT, NP_066569.1:p.Tyr90ThrfsX40) (eTable 3). Mutations in this gene are associated with McLeod neuroacanthocytosis syndrome (MIM 300842).</td>
<td>A 58 year-old male with bilateral ptosis, proximal myopathy, seizures, tics and COX-negative fibers on muscle biopsy. Blood film examination identified acanthocytes and confirmed the diagnosis of neuroacanthocytosis.</td>
<td>Surveillance for cardiac abnormalities and seizures was instituted and recommendations to use Kx-negative blood or banked autologous blood for transfusions when required.</td>
</tr>
</tbody>
</table>

ID = patient identification number for this study found in eTable 3; WGS = Whole Genome Sequencing; MIM = Mendelian Inheritance in Man; MRI = Magnetic Resonance Imaging; µM = Micromolar; COX = Cytochrome c oxidase