Panel-Based Nuclear and Mitochondrial Next-Generation Sequencing Outcomes of an Ethnically Diverse Pediatric Patient Cohort with Mitochondrial Disease

Maryke Schoonen,1 Izelle Smuts,2 Roan Louw, 1 Joanna L. Elson, 1,3 Etresia van Dyk, 1
Lindi-Maryn Jonck,1 Richard J.T. Rodenburg,4 and Francois H. van der Westhuizen1,*

1From Human Metabolomics, Faculty of Natural and Agricultural Sciences, North-West University, Potchefstroom, South Africa
2The Department of Paediatrics and Child Health, Steve Biko Academic Hospital, University of Pretoria, Pretoria, South Africa
3The Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom;
4The Department of Pediatrics, Radboudumc Amalia Childrens Hospital, Radboud Center for Mitochondrial Medicine, Nijmegen, the Netherlands

*Address correspondence to Francois H. van der Westhuizen, Ph.D., Human Metabolomics, North-West University, Potchefstroom Campus, Private Bag X6001, Potchefstroom 2520, South Africa.
E-mail: francois. vanderwesthuizen@nwu.ac.za.

Abstract

Mitochondrial disease (MD) is a group of rare inherited disorders with clinical heterogeneous phenotypes. Recent advances in next-generation sequencing (NGS) allow for rapid genetic diagnostics in patients who experience MD, resulting in significant strides in determining its etiology. This, however, has not been the case in many patient populations. We report on a molecular diagnostic study using mitochondrial DNA and targeted nuclear DNA (nDNA) NGS of an extensive cohort of predominantly sub-Saharan African pediatric patients with clinical and biochemically defined MD. Patients in this novel cohort presented mostly with muscle involvement (73%). Of the original 212 patients, a muscle respiratory chain deficiency was identified in 127 cases. Genetic analyses were conducted for these 127 cases based on biochemical deficiencies, for both mitochondrial (n = 123) and nDNA using panel-based NGS (n = 86). As a pilot investigation, whole-exome sequencing was performed in a subset of African patients (n = 8). These analyses resulted in the identification of a previously reported pathogenic mitochondrial DNA variant and seven pathogenic or likely pathogenic nDNA variants (ETFDH, SURF1, COQ6, RYR1, STAC3, ALAS2, and TRIOBP), most of which were identified via whole-exome sequencing. This study contributes to knowledge of MD etiology in an understudied, ethnically diverse population; highlights inconsistencies in genotype-phenotype correlations; and proposes future directions for diagnostic approaches in such patient populations.
Mitochondria are ubiquitous in the human body and serve mainly as the energy-producing organelle via oxidative phosphorylation (OXPHOS). This metabolic pathway comprises five protein complexes (CI to CV), consisting of a total of 92 structural subunits encoded by both mitochondrial DNA (mtDNA) and nuclear DNA genes. Underlying genetic mutations generate disruptions within this system, which manifest clinically, often affecting multiple highly energy-demanding organs simultaneously. The heterogeneous class of clinical phenotypes associated with such mutations is referred to as mitochondrial disease (MD). Many genes (at least 289), extending beyond the structural OXPHOS genes, have been identified as being involved in MD.

Traditionally, MDs are diagnosed through extensive clinical evaluation, including biochemical tissue analysis, followed by genetic screening for selected mutations. The current suite of next-generation sequencing (NGS) options available for MD diagnosis, and for heterogeneous disease diagnosis in general, includes targeted panel sequencing, unbiased whole-exome sequencing (WES), and whole-genome sequencing; each has particular advantages, disadvantages, and considerations. Recent publications advocate for a genetics first diagnostic approach, with the promise of eliminating the need for functional and biochemical analyses in most diagnoses. This raises some concerns for understudied ethnically diverse populations in developing countries, where relatively little progress has been made toward understanding the genetic etiology of MD and where the genotype-phenotype correlations are poorly understood and are inconsistent with those for non-African populations (as an example). To address these limitations, we and others have undertaken various clinical, biochemical, and genetic studies on MDs in the South African population, one of the few developing countries to do so (Supplemental Figure S1).

To date, a traditional diagnostic trajectory of extensive clinical evaluation and functional biochemical diagnosis of referred patients, followed by screening for known, common mutations, has been followed. Currently, published patient data, and public genetic and disease databases from predominantly non-African populations, are used as reference because of the limited (specific) information available on African MD etiology.

To address these diagnostic challenges for MD in an understudied ethnically diverse population, we report on the outcome of an NGS approach when targeting reported nuclear and mtDNA-encoded genes involved in MD. This approach was investigated in a predominantly African cohort of 212 South African pediatric patients selected on the basis of clinical and muscle respiratory chain (RC) enzymology data; this approach would be considered prudent in a diagnostic setting. We highlight the contrasting outcome of a WES approach in a small subset of this patient cohort and, with specific consideration of the genotype-phenotype correlation suggested by selected cases, propose future diagnostic directions that should be considered for similar understudied population groups.

Materials and Methods

Patient Cohort

Since 2006, >6000 patients with neurologic symptoms have been referred to the Steve Biko Academic Hospital (a state-funded institution in Pretoria, South Africa) and clinically evaluated, according to an MD criteria scoring system first set forth by Wolf and Smeitink (2002) and subsequently refined by Smuts et al. Currently, this cohort consists of 212 pediatric patients who manifested clinically with MD signs or symptoms from as early as the neonatal period. This cohort originated from the northern provinces of South Africa and is predominantly African (64%), with an
equal number of males and females. Urine and muscle (vastus lateralis) samples were collected from the entire cohort for subsequent biochemical and molecular genetic investigations. Limited availability of samples from parents and patients prevented segregation analysis. This study was approved by the Ethics Committees of the University of Pretoria (Pretoria, South Africa; number 91/98 and amendments) and the North-West University (Potchefstroom, South Africa; number NWU-00170-13-A1).

Biochemical Analysis

Muscle RC enzyme analyses for CI to CIV [Enzyme Commission number (EC) 1.6.5.3, EC 1.3.5.1, EC 1.10.2.2, and EC 1.9.3.1, respectively] and CII + CIII were performed and normalized against citrate synthase (EC, 2.3.3.1) activity for all 212 patients. Enzymes were analyzed in 600 × g homogenates prepared from frozen muscle samples, as previously described. Other biochemical analyses performed included urine metabolic analysis and muscle coenzyme Q10 (CoQ10) analysis in muscle samples. In total, 127 patients were identified to have an RC deficiency.

Genetic Analysis

mtDNA Sequencing

Genomic DNA preparation from muscle homogenate was performed using a standard protocol, as previously described. The Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA) was used for quantification of genomic DNA. The complete mitochondrial genome was sequenced in 123 patients, all of whom had a known combined clinical and biochemical MD profile. Two chemistries were used: 71 patient samples were sequenced using the 454 GS-FLX platform (Roche, Basel, Switzerland), and 52 patient samples were sequenced using the Ion PGM platform (Thermo Fisher Scientific). The 454 GS-FLX sequencing, including library preparation and enrichment, was done according to procedures described by van der Walt et al. The Ion PGM sequencing, including library preparation and enrichment, was done according to the manufacturer’s protocol for Ion Torrent platform. Four samples could not be sequenced because of insufficient sample quantity and quality.

Nuclear Panel Sequencing

For nuclear gene investigations, Ion Torrent amplicon panel sequencing was performed on 86 patients, all of whom had a known combined clinical and biochemical MD profile. Patients were included for sequencing based on their MD criteria scores for clinical presentations and the severity of their biochemical deficiencies [ie, patients who had enzyme activity lower than the reference range when expressed against at least two of three enzyme markers (citrate synthase, CII, or CIV) were considered]. Genes to be included in sequencing panels were selected on the basis of their mitochondrial RC involvement, either direct or indirect (Supplemental Tables S1, S2, and S3). Three custom gene panels consisting of 136 genes in total were designed and are briefly described. Panel 1 consisted of 78 targeted genes associated with CI deficiency (HaloPlex Target Enrichment System; Agilent Technologies, Santa Clara, CA) (Supplemental Table S1). This customized panel had a target region size of 360,091 kbp and 99.6% targeted coverage. In total, 30 patients were sequenced using panel 1.

Panel 2, from Ion AmpliSeq Custom DNA Panels (Thermo Fisher Scientific), had 78 targeted genes associated with CI to IV deficiency (CI, 38 genes; CII, 6 genes; CIII, 10 genes; and CIV, 24 genes), with a total target region size of 157,834 kbp and 98.2% targeted coverage (Supplemental Table S2). In
total, 48 patients were sequenced using panel 2, of which 5 patients overlapped with panel 1 and 10 patients overlapped with panel 3.

Panel 3, from Ion AmpliSeq, targeted 18 genes known to be involved with primary and secondary CoQ10 deficiency (Supplemental Table S3). The design size was 61 kbp with a targeted coverage of 98%. In total, 26 patients were sequenced in panel 3 (6 patients overlapping with panel 1). The entire coding region of each gene, including flanking regions of introns-exons, was sequenced using the Ion PGM platform, as per manufacturers' protocol (HaloPlex, catalog number G9912C, and AmpliSeq, catalog number 4480441). The selected genes and panels were not African population specific, as the underlying genetic cause for MD is mostly unclear in African populations.

Whole-Exome Sequencing

As an initial comparison on the outcomes of a panel versus the WES approach, WES was performed on a subset of eight randomly selected African (haplogroup L) cases in whom no strong candidate disease-causing variants had been identified by initial mtDNA and/or nuclear panel sequencing. WES was performed at the Central Analytical Facilities, Stellenbosch University (Stellenbosch, South Africa), using the Ion Proton sequencer (Thermo Fisher Scientific), according to the manufacturer’s protocol for the Ion Torrent platform. An approximate 95% on-target coverage was achieved, with an average depth coverage of approximately 140×.

Bioinformatics for mtDNA Variants

Mitochondrial DNA sequences were aligned against the human mitochondrial revised Cambridge Reference Sequence (NC_012920 gi:251831106). Haplogroup assignment, and variant identification and annotation, was performed using mtDNA-Server version 1.20.0, MitoMap, and MitoMaster. Homoplasmic and heteroplasmic (levels >30%) nonsynonymous variants were further evaluated on the basis of their allele frequency reported in GenBank and appearance on Phylotree, and those with a population allele frequency <0.1% were considered significant. Variant pathogenicity was evaluated according to the criteria proposed for mtDNA variants. For example, the MutPred scoring system was used to classify nonsynonymous variants in structural subunits of OXPHOS (http://mutpred.mutdb.org/help.html, last accessed April 23, 2018). A MutPred score >0.5 suggests a probable damaging impact on protein function, with scores between 0.75 and 1.0 indicating such functional damage on a protein/amino acid with high confidence. Mitochondrial-tRNA variants were individually evaluated using MitoTIP and classified according to a scoring system first set forth by McFarland et al and subsequently refined by Yarham et al. According to Yarham et al, a low score (<10) weighs more toward benign or neutral classification, whereas a score >10, with substantial evidence from functional tests, weighs more toward a pathogenic classification. Variants were also evaluated using the guidelines put forth by the American College of Medical Genetics and Genomics, where possible (notably for mtDNA variants).

Bioinformatics for Nuclear DNA Variants

Raw sequencing files, obtained from the Ion PGM, were analyzed with the use of Torrent Suite version 5.0.2 (Thermo Fisher Scientific). The sequence files were aligned against Genome Reference Consortium Human Build 37 (hg19), followed by coverage analysis and variant calling using the coverageAnalysis and variantCaller plugins version 5.0 from the Torrent Suite, respectively. The variant calling format files were further annotated using the offline Variant Effect Predictor version
81 from Ensembl (https://uswest.ensembl.org/info/docs/tools/vep/index.html, last accessed July 10, 2018), followed by variant mining using GEMINI version 20. The output text files generated using GEMINI contained information on both novel and reported variants. Most notably, they detailed whether a variant had previously been reported as pathogenic. Further variant filtering was done using population databases, such as Exome Aggregation Consortium and gnomAD (specifically for African allele frequencies); disease-specific databases, such as ClinVar and Online Mendelian Inheritance in Man (OMIM); and sequence databases, such as the National Center for Biotechnology Information Genome and RefSeqGene. As supporting evidence, the missense variants of interest were cautiously evaluated using various in silico predictive algorithms (SIFT, Polyphen-2, and CADD). These algorithms, however, have been shown to have low sensitivity, specificity, and accuracy. Candidate variants of interest were evaluated using American College of Medical Genetics and Genomics guidelines, and are classified as pathogenic, likely pathogenic, variants of uncertain significance, likely benign, or benign.

Results
Clinical Profiles

From the >6000 neurologic patients referred for clinical assessment, 212 patients (113 males and 99 females) presented with mitochondrial phenotypes, forming the cohort that is described herein. These clinically defined MD patients were comprehensively evaluated and were predominantly of African ancestry (n = 130) (Figure 1A). Age of onset was as early as the first year of life, including the neonatal period (n = 139) (Figure 1B). The most common clinical finding was muscle involvement, which manifested in 73% (n = 155) of the cohort. Cardiac involvement and deafness were the least observed, at 5% (n = 10) and 8% (n = 17), respectively (Figure 1C).
Figure 1. Demographic information for 212 pediatric patients. A: Summary of the ethnicity distribution. White, Asian, and colored patients are collectively referred to as non-African patients in text. B: Summary of the age categories representing the age at symptom onset for each of the 212 patients. C: Summary of the clinical features and findings in the patient cohort. CNS, central nervous system; ENT, ears, nose, and throat; GIT, gastrointestinal tract; PNS, peripheral nervous system.

Biochemical Profiles

An RC enzyme complex deficiency was identified in 127 cases (64 males and 63 females); 64 cases had isolated deficiency, and the remaining 63 had a combined deficiency (Figure 2A). CI deficiency, either isolated (n = 43) or combined (n = 32), was most prevalent, as is frequently reported, followed by CIII deficiency (n = 54; isolated, n = 15; and combined, n = 39) (Figure 2B).
Figure 2. Summary for the biochemical respiratory chain (RC) enzyme deficiency distribution. **A:** Combined versus isolated enzyme deficiency for 127 patients. **B:** Summary for complex I to complex IV enzyme deficiency in 64 patients with isolated RC deficiency.

Genetic Analysis

The findings for whole mtDNA, panel nuclear DNA, and whole-exome NGS are summarized in Table 1. Whole mtDNA sequencing was successfully applied in 123 cases (79
African and 44 non-African), and a variant of interest was identified in each of 12 African cases, none of which could be classified as disease causing. An example of such a variant of interest is the well-known Leber hereditary optic neuropathy (OMIM number 535000) mutation, m.14484T>C, identified in an African male (S014) who did not present with typical Leber hereditary optic neuropathy symptoms and was later clinically diagnosed with fucosidosis. An additional 11 mtDNA variants of interest, identified in 10 cases, may, however, still prove to have a functional effect on disease initiation or progression (Supplemental Table S4). Notably, this may still be the case for variants with a comparatively high population allele frequency (>0.1%) because further investigation may reveal that these variants appear at a higher frequency in these understudied population lineages.

Panel nuclear DNA sequencing of 86 cases (61 African and 25 non-African) revealed four variants of interest in two cases. Two pathogenic compound heterozygous electron-transfer flavoprotein dehydrogenase (ETFDH) variants were identified in one case (S057), and two possible pathogenic compound heterozygous surfeit 1 (SURF1) variants were identified in a second case (S085).

WES sequencing of eight African cases revealed nine variants of interest (in six cases), eight of which (variants in the genes coenzyme Q₁₀ monooxygenase 6 (COQ6), ryanodine 1 (RYR1), SRC homology 3 and cysteine-rich domains 3 (STAC3), and 5′-aminolevulinate synthase 2 (ALAS2)) are considered possibly pathogenic and one of which (in the gene TRIOBP) is considered as a variant of uncertain significance under the American College of Medical Genetics and Genomics classification.

**Pathogenic Variants**

The ETFDH compound heterozygous variants (c.1448C>T and c.1067G>A; p.Pro483Leu and p.Gly356Glu; ENST00000307738) were identified in a non-African female (S057) who presented with
features of multiple acyl-dehydrogenase deficiency (OMIM number 231680). She had severe muscle weakness, exercise intolerance, chronic fatigue, and hepatomegaly. Metabolic markers of multiple acyl-dehydrogenase deficiency, such as dicarboxylic acids, ethylmalonic acid, and glutaric acid, as well as acylcarnitines (butyrylcarnitine, isovalerylcaritinie, and glutarylcaritinie), acylglycines (hexanoyl-isobutyryl glycine, isovalerylglycine, and suberylglycine), and conjugates were observed in the patient’s urine. Furthermore, she had CI, CIII, and CII + CIII RC deficiency, with severely reduced CoQ10 levels in muscle. These mutations cause structural instability in electron-transferring-flavoprotein dehydrogenase (ETFDH) and were confirmed with functional analysis.

Likely Pathogenic Variants

An African female (S085), with clinically confirmed Leigh disease (OMIM number 256000) and confirmed mitochondrial CIV deficiency, harbored compound heterozygosity for missense and frameshift variants (c.575G>A and c.754_755delAG; p.Arg192Gln and p.Ser252HisfsTer39; ENST00000371974) in exons 6 and 8, respectively, in the gene SURF1. The patient clinically presented early in life with several central nervous system involvements (nystagmus and extrapyramidal and pyramidal symptoms) and muscle involvements (myopathy, hypotonia, and weakness). Furthermore, changes in the basal ganglia/thalami were observed. Metabolic investigation revealed elevated lactic acid.

An African female (S002) with confirmed primary CoQ10 deficiency was found to be carrying compound heterozygous variants (c.41G>A and c.859G>T; p.Trp14Ter and p.Ala49Ser; ENST00000394026) in the gene COQ6. One of the variants identified in a highly conserved region, c.41G>A, results in a premature truncation of the protein with high loss-of-function (LoF) probability. Clinically, she presented at 4 years 4 months of age with a history of severe weakness since birth. Other features were noted on clinical examination and included macrocephaly, severe hypotonia and head lag, pseudohypertrophy of calves and triceps, limitation of extension at knees and elbows, and reduced reflexes. No other candidate COQ6 gene variants were identified in this case.

Two sets of compound heterozygous variants, associated with centronuclear myopathy and minicore myopathy and external ophthalmoplegia (OMIM number 255320), were identified in the gene RYR1 in two African females (S032 and S033) using WES. For S032, a previously reported pathogenic variant (c.8342_8343delTA; p.Ile2781ArgfsX49; ENST00000359596) with high LoF confidence and a missense variant (c.11926C>T; p.His3976Tyr; ENST00000359596), both in highly conserved regions, were identified. She presented with severe hypotonia and myopathy with myopathic facial features, and she initially did not show signs of external ophthalmoplegia. She also had an affected sibling with myopathic facial features and external ophthalmoplegia. A muscle biopsy specimen was collected from the sibling who revealed thickened connective tissue between muscle fibers and evidence of fat infiltration. The female described herein had CIV RC deficiency, and her brother had CI + CIII deficiency. Both siblings presented with an additional CII + CIII deficiency.

For S033, a previously reported pathogenic missense variant (c.14524G>A; p.Val482Met; ENST00000359596) and a frameshift variant (c.11193+1G>A; ENST00000359596) were identified. Both variants are in highly conserved regions. She presented with severe hypotonia, mild myopathic facial features, and dense external ophthalmoplegia.

A homozygous mutation (c.851G>C; p.Trp284Ser; ENST00000332782) was detected by WES in an African female (S011) in the gene STAC3, and was first described by Horstick et al in five families.
with Native American myopathy (OMIM number 255995). The phenotype observed in the female described herein is similar to that for a case described elsewhere, and includes severe myopathy, failure to thrive, developmental delay, relative macrocephaly, and ptosis with no external ophthalmoplegia. She had minor dysmorphic features, including a low nasal bridge. She was born prematurely and had intrauterine growth restriction. Biochemically, she presented with isolated CIII deficiency in muscle.

An African male (S117), with skin and muscle involvement, carried a homozygous gain-of-function variant (c.1757A>T; p.Tyr586Phe; ENST00000330807) in exon 11 of the gene ALAS2. This variant was first described in a Spanish patient with erythropoietic porphyria by To-Figuera et al., who state that ALAS2 acts as a modifier gene in patients with erythropoietic porphyria (X-linked protoporphyria; OMIM number 300752). Another manifestation is sideroblastic anemia type 1 (OMIM number 300751), which is also as a result of ALAS2 mutations. However, our patient did not present with symptoms for the latter. He instead presented with swelling of his face, hands, and feet and experienced nonspecific body pain, symptoms that are more similar to X-linked protoporphyria. He had depigmented skin lesions in his face, on the extensor areas of the fore and upper arms, and over the knees and lateral right thigh. Furthermore, he experienced severe muscle weakness with decreased muscle bulk in all four limbs, with muscle histology revealing atypical dermatomyositis. Biochemically, he had confirmed CI, CIII, and CIV deficiency in muscle.

**Variants of Uncertain Significance**

A trio- and filamentous-actin–binding protein (TRIOBP) homozygous missense variant (c.3232C>T; p.Arg1078Cys; ENST00000406386) was identified using WES in an African male (S059) who presented with developmental delay, visual impairment, muscle weakness and hypotonia, and clinodactyly. Furthermore, he presented with mild facial dysmorphisms, which included an epicanthic fold and low-set ears. Most notably, the patient had hearing impairment, which was confirmed by an abnormal auditory brainstem response. Mutations in this gene are known to cause autosomal recessive deafness (OMIM number 609823), a feature found in the case presented herein. Metabolic profiles revealed significant ketosis associated with dicarboxylic aciduria involving C6 to C10 acids (ie, adipic, suberic, and sebacic acids), with a normal amino acid profile.

Several cases (n = 13) presented with variants of uncertain significance, and are listed in Supplemental Table S5. These nuclear variants are classified as likely pathogenic or pathogenic, according to ClinVar, but did not adhere to the American College of Medical Genetics and Genomics standards and guidelines. For example, previously reported pathogenic variants identified in the genes NDUFA9, SDHA, SDHB, and POLG were all detected as heterozygous, whereas homozygous variants in the genes TRMU, ACADVL, and GLUD2 had high African allele frequencies. These variants, however, remain of interest for South African population groups.

**Discussion**

The genetic diagnosis of MD, and identification of the number of genes involved, has rapidly improved since the first mutations were reported in the late 1980s. Although the use of clinical scoring systems, such as MD criteria scoring, in addition to biochemical evaluation of RC/OXPHOS function in tissue, is still the hallmark of MD diagnostics, it has been increasingly complimented by NGS in recent years, notably exome sequencing. The main advantage of using NGS in MD diagnoses is that it allows for the discrimination between primary MD (having a direct genetic etiology) and
secondary MD (caused by nongenetic factors, such as environmental toxins). A retrospective investigation into genetic causes of MD was conducted for 127 patients with clinically suspected and biochemically confirmed RC deficiency in an understudied population (predominantly African). Herein, we report on two high-throughput NGS techniques (whole mtDNA sequencing and panel sequencing) that were used to find common, previously reported pathogenic or likely pathogenic variants in reported genes involved in MD in this understudied ethnically diverse cohort with mostly unknown MD genotype-phenotype correlations.

The mtDNA sequencing data for a subsection of this cohort have been extensively investigated elsewhere. Most variants found could not be classified as pathogenic as they fail to meet several mtDNA criteria. Most notably, among the observed variants was a well-known Leber hereditary optic neuropathy–associated mutation (m.14484T>C at 53% heteroplasmy in muscle), identified in an African male (S014). Although this patient had clinical features of eye involvement, his clinical phenotype did not match that expected for Leber hereditary optic neuropathy (visual failure and optic atrophy). This inconsistency could be ascribed to varied penetrance of the disease, highlighting the importance of investigating the penetrance of diseases associated with this and other pathogenic variants when detected in understudied populations.

Panel sequencing revealed pathogenic and likely pathogenic variants in two cases: S057 and S085. For S057, the compound heterozygous variants that were identified in ETFDH (c.1448T>C and c.1067G>A) have previously been investigated extensively, leading to the classification of these variants as pathogenic with substantial clinical, biochemical, and in vivo supporting experimental evidence (from structural and functional analyses). For S085, compound heterozygous variants were identified in SURF1 (c.754_755delAG and c.575G>A). Surfeit locus protein 1 (SURF1) is directly involved with cytochrome c oxidase maintenance and assembly. LoF mutations in SURF1 cause major structural instability in cytochrome c oxidase and are responsible for the phenotype of Leigh disease as clinically diagnosed in the case reported herein. The frameshift variant has previously been reported as pathogenic for a different (Japanese) patient, with substantial clinical and experimental supporting evidence and a confirmed LoF mechanism. Consequently, this variant, which has a low African allele frequency, was classified as likely pathogenic in this case. The missense variant has not previously been associated with a clinical phenotype, and the extremely low allele frequency suggests moderate likelihood to be pathogenic.

Because the panel sequencing of genes known to be involved in MD revealed a pathogenic or likely pathogenic variant in only two cases (S057 and S085) of the 86 patients investigated, it was evident that this is not an effective approach to follow in this patient population. Although targeted/gene-panel NGS is considered a prudent NGS approach in many diagnostic settings, while being aware that this is a heterogeneous, understudied patient population and that an expanded gene panel may not necessarily increase diagnostic yield, genetic investigations were expanded to include WES to probe its outcome on a small subset of eight African patients in whom no initial NGS results were obtained. A variant of interest in the genes COQ6, RYR1, STAC3, ALAS2, and TRIOBP was identified in six of the eight cases.

The compound heterozygous COQ6 variants (c.41G>A and c.859G>T) identified in S002 have been extensively investigated elsewhere, where functional and structural analyses showed significantly decreased levels of coenzyme Q6 monooxygenase (COQ6). This protein is directly involved with CoQ10 biosynthesis, and mutations in this gene result in primary CoQ10 deficiency. This correlates well with the clinical and biochemical profiles observed in the case reported herein. On the basis of
these experimental findings, the correlation between the observed and reported phenotypes, and their allele frequencies, these variants are classified as likely pathogenic.

Two previously reported pathogenic and two reported RYR1 variants were identified in two compound heterozygotes: S032 (c.8342_8343delTA and c.11926C>T) and S033 (c.14524G>A and c.11193+1G>A). The c.14524G>A and c.8342_8343delTA variants are classified as founder mutations for South African patient populations with centronuclear myopathy and minicore myopathy and external ophthalmoplegia. RYR1 encodes a homotetrameric calcium channel in skeletal muscle and regulates cytosolic Ca$^{2+}$ levels. Dysfunctional RYR1 disrupts the Ca$^{2+}$ balance, directly affecting different mitochondrial functions, such as ATP synthesis regulation and reactive oxygen species generation, consequently contributing to myopathy, external ophthalmoplegia, and ptosis. The two cases reported herein had clinical phenotypes consistent with reported cases. For S032 and her brother, the specific type of congenital myopathy is still unclear. Both had neurogenic features that are absent in multiminicore myopathy. The siblings, whose features correlated more strongly with centronuclear myopathy, were similar in their manifestation (albeit with slight, but significant, differences). For example, the female initially did not have external ophthalmoplegia, whereas the brother did. More important, no homozygous variants have been reported in patients with centronuclear myopathy. S033 had correlating multiminicore myopathic features, as seen in previously reported cases, which included mild myopathic facial features with dense external ophthalmoplegia. Both sets of compound heterozygous variants are classified as likely pathogenic in these cases because of substantial supporting clinical and experimental evidence confirming the LoF mechanism, the confirmed founder effect for South African populations with a low or absent allele frequency in several population databases, and multiple in silico algorithms supporting a deleterious effect on the gene product for the two frameshift variants.

The previously reported pathogenic STAC3 variant (c.851G>C) was identified as homozygous in S011. STAC3 encodes a putative muscle-specific adaptor protein, which takes part in excitation-contraction in muscle. Disruptions within this protein are thought to cause a reduction in mitochondrial Ca$^{2+}$, either intramitochondrial or extramitochondrial, which has a direct effect on OXPHOS. Clinical features, as a result, include congenital myopathy with facial dysmorphic features, including ptosis, which is consistent with the case reported herein. This variant was previously sequenced as part of the 1000 Genomes Project and was undetected in 113 white controls. Consequently, we classify this variant as likely pathogenic in this case because of its previously confirmed pathogenicity (with substantial supporting evidence) and low African allele frequency.

The homozygous variants identified in ALAS2 (c.1757A>T) and TRIOBP (c.3232C>T) in S117 and S059, respectively, are classified as variants of uncertain significance, and benign or likely benign, respectively, according to several disease databases. The two cases reported herein, S117 and S059, had overlapping, but also slightly inconstant, phenotypical manifestations compared with reported cases.

Specifically, S117 did not present with typical X-linked protoporphyria, the primary phenotype associated with ALAS2 mutations. His primary cause of disease is interplay between two deficiencies: dermatomyositis and X-linked porphyria. The muscle histology from this patient was suggestive of an inflammatory myopathy, but the inflammatory cell infiltrate was perivascular, and the typical pattern of peripheral muscle cell atrophy in dermatomyositis was not demonstrated. As the patient had definitive skin involvement, which included swelling and redness, it is suggested that the patient...
has porphyria rather than inflamed myopathy. His mitochondrial dysfunction could arise from impaired electron transport as a direct result of bone marrow haem synthetic dysfunction. The clinical significance is still unclear and, therefore, this variant is classified as a variant of uncertain significance. A case can be made, however, that this variant should be classified as likely pathogenic because it has a low African allele frequency, and there is substantial supporting evidence for its pathogenicity.

S059 had a severe mitochondrial phenotype, where deafness was most notable. The latter is consistent with reported cases harboring TRIOPB mutations\(^69, 70\); however, the former has not yet been associated with mutations in this gene. The underlying genetic cause for this patient’s mitochondrial phenotype is still unclear. This variant, c.3232C>T, has not yet been associated with a clinical phenotype and, therefore, lacks substantial supporting pathogenicity evidence. Consequently, it was classified as a variant of uncertain significance.

To conclude, in this unique and predominantly African cohort, looking first at nuclear and mitochondrial genes known to be involved in MD (which is in line with current diagnostic practices) resulted in a poor diagnostic outcome, with only a relatively small number of pathogenic or likely pathogenic variants being successfully identified. Initial indications from limited WES data are much more promising as nine likely pathogenic variants were identified in six cases using WES compared with five variants identified in three cases using panel NGS and mtDNA sequencing.

Considering all of the cases in this study, a strong genotype-phenotype correlation could be established in only five cases (S057, S085, S032, S033, and S011), and a moderately strong correlation could be established in two cases (S002 and S117). For the remaining two cases, S014 and S059, a non-specific correlation was observed. Observations like these serve as a strong motivation that a genetics-first/only approach, without supporting clinical and biochemical investigation, is not suitable in such understudied, ethnically diverse populations. Furthermore, when following a genetic approach, we conclude that panel sequencing could be an efficient approach in populations where the genotype-phenotype correlations are well established for specific monogenic diseases. For heterogeneous diseases such as MD, however, even in such populations, WES/whole-genome sequencing performs significantly better compared with a targeted gene-panel approach.\(^4, 5, 71, 72, 73\) Our results are, thus, in line with proposals that WES should be considered as the primary option for genetic investigations in heterogeneous inherited diseases such as MD and, in fact, may be particularly ideal in understudied, ethnically diverse populations where there is evidence of inconsistencies with documented MD phenotypes. However, in such populations, the value of extensive clinical and biochemical (structural and functional) investigations to support molecular genetic data outcomes should not be neglected, and at this time, they are, in fact, more crucial than in well-studied populations.
References

20. Van Oven, M. PhyloTree Build 17: growing the human mitochondrial DNA tree. Forensic Sci Int. 2015; 5: e392–e394


