Use of whole genome sequencing to determine the genetic basis of suspected mitochondrial disorders: a cohort study

Katherine R. Schon[1,2,3] (clinical research fellow), Rita Horvath[1] (clinical director of research), Wei Wei[1,2] (senior bioinformatician), Claudia Calabrese[1,2] (research associate), Arianna Tucci[4] (clinical research fellow), Kristina Ibañez[4] (senior research associate), Thiloka Ratnaike[1,2,5](clinical research fellow), Robert D.S. Pitceathly[6] (principal research fellow), Enrico Bugiardini[6] (clinical research associate), Rosaline Quinlivan[6] (professor), Michael G. Hanna[6] (professor), Emma Clement[7] (consultant clinical geneticist), Emma Ashton[8] (clinical scientist), John A. Sayer[9] (professor), Paul Brennan[10] (consultant clinical geneticist), Dragana Josifova[11] (consultant clinical geneticist), Louise Izatt[11] (consultant clinical geneticist), Carl Fratter[12] (consultant clinical scientist), Victoria Nesbitt[12] (consultant paediatrician), Timothy Barrett[13] (professor), Dominic J McMullen[13] (consultant clinical scientist), Audrey Smith[14] (consultant clinical geneticist), Charulata Deshpande[14,11] (consultant clinical geneticist), Sarah F. Smithson[15] (professor), Richard Festenstein[16] (professor), Natalie Canham[17] (consultant clinical geneticist), Genomics England Research Consortium[18], Mark Caulfield[18] (professor), Henry Houlden[6,19] (professor) and Shamima Rahman[19,20,21] (professor) on behalf of the Neurology and Mitochondrial Disorders Genomics England Clinical Interpretation Partnership[18], Patrick F. Chinnery[1,2,19]* (professor).

[1] Department of Clinical Neurosciences, School of Clinical Medicine, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK.

[2] Medical Research Council Mitochondrial Biology Unit, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK.

[3] East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK

[4] William Harvey Research Institute, Queen Mary University of London, London, UK.

[5] Department of Paediatrics, University of Cambridge, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK

[6] Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology and The National Hospital for Neurology and Neurosurgery, London, UK

[7] Department of Clinical Genetics and Genomic Medicine, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

[8] NHS North Thames Genomic Laboratory Hub, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK
[9] Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK
[10] Northern Genetics Service, Newcastle Hospitals NHS Foundation Trust, International Centre for Life, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK
[11] Department of Clinical Genetics, Guy's Hospital, Great Maze Pond, London, SE1 9RT, UK
[12] NHS Highly Specialised Services for Rare Mitochondrial Disorders – Oxford Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[13] West Midlands, Oxford and Wessex Genomics Laboratory Hub, Birmingham Women's and Children's NHS Foundation Trust, Edgbaston, Birmingham, B15 2TG, UK
[14] Manchester Centre for Genomic Medicine, 6th Floor St Mary's Hospital, Oxford Road, Manchester, M13 9WL, UK
[15] Department of Clinical Genetics, University Hospitals Bristol, Marlborough Street, Bristol, BS1 3NU, UK
[16] Department of Brain Sciences, London Institute of Medical Sciences, Mansfield Centre for Innovation, Imperial College, Hammersmith Hospital, Du Cane Road, London, W12 0NN
[17] Liverpool Centre for Genomic Medicine, Liverpool Women's Hospital , Crown Street, Liverpool, L8 7SS
[18] Genomics England, William Harvey Research Institute, Queen Mary University of London, London, UK.

[19] Neurology and Mitochondrial Disorders Genomics Clinical Interpretation Partnership

[20] Metabolic Unit, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK.

[21] Mitochondrial Research Group, Department of Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, London, UK.

* correspondence to:
Professor Patrick Chinnery
Department of Clinical Neurosciences, University Neurology Unit, Level 5, A Block, Box 165, Cambridge Biomedical Campus, Cambridge, CB2 0QQ
ORCID ID 0000-0002-7065-6617

Twitter @cam_mito

pfc25@cam.ac.uk

Genomics England Research Consortium: Ambrose, J. C.[1] ; Arumugam, P.[1] ; Bevers, R.[1] ; Bleda, M.[1] ; Boardman-Pretty, F.[1,2] ; Boustred, C. R.[1] ; Brittain, H.[1] ; Caulfield, M. J.[1,2] ; Chan, G. C.[1] ; Fowler, T.[1] ; Giess A.[1] ; Hamblin, A.[1] ; Henderson, S.[1,2] ; Hubbard, T. J. P.[1] ; Jackson, R.[1] ; Jones, L. J.[1,2] ; Kasperaviciute, D.[1,2] ; Kayikci, M.[1] ; Kousathanas, A.[1] ; Lahnstein, L.[1] ; Leigh, S. E. A.[1] ; Leong, I. U. S.[1] ; Lopez, F. J. [1]; Maleady-Crowe, F.[1] ; McEntegart M.[1] ;Minneci F.[1] ; Moutsianas, L.[1,2] ; Mueller, M.[1,2] ; Murugaesu, N.[1] ; Need, A. C.[1,2] ; O'Donovan P.[1] ; Odhams, C. A.[1] ; Patch, C.[1,2] ; Perez-Gil, D.[1] ; Pereira, M. B.[1] ; Pullinger, J.[1] ; Rahim, T.[1] ; Rendon, A.[1] ; Rogers, T.[1] ; Savage, K.[1] ; Sawant, K.[1] ; Scott, R. H.[1] ; Siddiq, A.[1] ; Sieghart, A.[1] ; Smith, S. C.[1] ; Sosinsky, A.[1,2] ; Stuckey, A.[1] ; Tanguy M.[1] ; Taylor Tavares A. L.[1] ; Thomas, E. R. A.[1,2] ; Thompson, S. R.[1] ; Tucci, A.[1,2] ; Welland, M. J.[1] ; Williams, E.[1] ; A. Witkowska, K.[1,2] ; Wood, S. M. [1,2]

1. Genomics England, London, UK
2. William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK.

Abstract

Objective To determine whether whole genome sequencing (WGS) can be used to define the molecular basis of suspected mitochondrial disease.

Design Patients with suspected mitochondrial disorders were recruited to the 100,000 Genomes Project in England between 2015 and 2018.

Setting A multicentre study in the National Health Service in England including secondary and tertiary care.

Participants 319 families with suspected mitochondrial disease were referred for WGS after excluding common genetic causes. 345 affected individuals were recruited (186 female, 159 male), with median age 25 years (range 0 to 92). Trio or quad analysis with both parents was performed in 148 families (46%). Clinical information was collected systematically using Human Phenotype Ontology (HPO) terms with a median of 7 HPO terms per participant. Commonest HPO terms were delayed gross motor development, intellectual disability, myopathy, seizures and ptosis.

Intervention: Short-read whole genome sequencing was performed. Nuclear variants were prioritised based on gene panels chosen according to phenotypes, ClinVar pathogenic/likely pathogenic variants, and the top ten prioritised variants from Exomiser. mtDNA variants were called using an in-house pipeline and compared to a list of pathogenic variants. Copy number variants and short tandem repeats for thirteen neurological disorders were also analysed. ACMG guidelines were followed for the classification of variants.

Main Outcome Measures: Finding a definite or probable genetic diagnosis.

Results: A definite or probable genetic diagnosis was identified in 98 families (31%), with an additional 6 possible diagnoses (2%). Fourteen of the diagnoses (4% of the 98 families) explained only part of the clinical features. A total of 95 different genes were implicated. 37.5% of diagnosed families had a mitochondrial diagnosis and 62.5% had a non-mitochondrial diagnosis.

Conclusion WGS is a useful diagnostic test in patients with suspected mitochondrial disorders yielding a diagnosis in a further 31% after excluding common causes. The majority of diagnoses were non-mitochondrial disorders and included developmental disorders with intellectual disability, epileptic encephalopathies, other metabolic disorders, cardiomyopathies and leukodystrophies. These would have been missed if a targeted approach was taken, with some having specific treatments.

Summary Box

What is already know on this topic

- Mitochondrial disorders are amongst the most common inherited diseases, but a genetic diagnosis is not possible in ~40% of patients, limiting genetic counselling and prevention.
- Whole genome sequencing (WGS) has the potential to shorten the "diagnostic odyssey" for patients with suspected mitochondrial disorders, but its use in a national healthcare system has not been previously investigated.

What this study adds

- After excluding common genetic causes, WGS identified a definite or probable genetic diagnosis in an additional 31% of patients (including 4% with partial diagnoses), and an additional 2% had possible diagnoses.
- 62.5% of the newly diagnosed families had a non-mitochondrial diagnosis showing that a wide genomic approach is more useful than a targeted panel testing approach. The new diagnoses included treatable disorders.

Introduction

Mitochondrial disorders have emerged as a common cause of inherited metabolic disease affecting ~1 in 5000 people[1] and are caused by mutations in genes which primarily affect oxidative phosphorylation and ATP synthesis[2]. Mitochondria are intra-cellular organelles which play a pivotal role in cellular energy metabolism. This is achieved by a series of complex enzymes located in the inner mitochondrial membrane which perform oxidative phosphorylation and synthesise adenosine triphosphate (ATP). ATP is a chemical source of energy required for all active cellular processes. The impairment of mitochondrial function tends to affect tissues with high energy demand such as the brain, the peripheral nerves, the eye, the heart and the peripheral muscles. The clinical diagnosis of mitochondrial disorders is challenging because they can affect a single organ such as the eye in Leber Hereditary Optic Neuropathy[3], or many different systems, and they can present at any age. Although some patients present with a classical mitochondrial syndrome, such as mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS), many present with only one or a few of the clinical features (oligosymptomatic cases)[4] or with an ill-defined multi-system disorder. Mitochondria contain their own DNA in the form of a small 16.5kb circle of double stranded DNA (mtDNA), which encodes for 13 peptides, two ribosomal RNAs and 22 transfer RNAs which are essential for synthesising proteins within the organelle. However, the vast majority of proteins making up the mitochondria are encoded by genes in the nucleus and are synthesised in the cytosol before being imported through a bespoke import machinery. Mitochondrial disorders can be caused by pathogenic variants in either the mtDNA or the nuclear genome, can follow any inheritance pattern (autosomal dominant, autosomal recessive, X-linked, *de novo* or maternal) and are highly genetically heterogeneous[5]. For example, Leigh syndrome, the commonest childhood presentation of mitochondrial disease which usually presents in the first year of life with stepwise loss of skills, is caused by mutations in ~100 nuclear and mtDNA genes[6]. These challenges often result in a prolonged patient journey from symptom onset to reaching a diagnosis, referred to as a “diagnostic odyssey”[7,8]. In patients with rare diseases, this typically involves multiple appointments - first in primary care, then with different specialist services - and many investigations, sometimes over many years. One survey[8] found that patients saw an average of 8 physicians before being diagnosed with a mitochondrial disease, and 70% had a muscle biopsy.

The diagnosis of mitochondrial disease has traditionally relied on an invasive tissue biopsy for biochemical and histochemical analysis[9], which can be normal even in patients with a defined genetic diagnosis. Sequencing a pre-defined list of genes known to cause a specific disorder (multi-gene panels), and sequencing of the protein-coding regions (exons) of all genes (exome sequencing), have been effective for diagnosing mitochondrial disorders[10–16] and in discovering new mitochondrial disease genes. However, it is still not possible to make a genetic diagnosis in ~40%[10] even in highly selected cohorts, hence the need for new approaches.

A definitive genetic diagnosis benefits patients and families[17], allowing tailored information about prognosis and treatment, genetic counselling and access to reproductive options such as prenatal diagnosis (genetic testing during pregnancy – usually by chorionic villus sampling or amniocentesis), pre-implantation genetic diagnosis (the use of assisted reproductive technology and genetic testing of embryos), and mitochondrial transfer (replacing the mother's mitochondria in an ovum or early embryo with healthy mitochondria from another woman's donor egg or embryo, used for disorders caused by mtDNA mutations)[18]. Whole genome sequencing (WGS) is a next-generation sequencing technology that is used to sequence the entire genome of an individual. WGS has the added benefit of being able to diagnose pathogenic mutations affecting the mtDNA and the nuclear genome[19], and thus has the potential to make a diagnosis in more families and shorten the time to diagnosis[20]. The objective of this study was to see if whole genome sequencing could be used to define the molecular basis of suspected mitochondrial disorders in a national health care system in patients assessed in mainstream secondary care as well as tertiary centres. The 100,000 Genomes Project was set up to introduce and embed genomic testing into the mainstream National Health Service (NHS), discover new disease genes and make genetic diagnosis available for more patients[21]. Following an initial pilot phase[22], patients in the 100,000 Genomes Project were recruited from NHS Genomic Medicine Centres (GMCs) across England. Here, we report the results for 345 patients with a suspected mitochondrial disorder recruited into the main programme between 2015 and 2018.

Materials and Methods

Participants

Participants with suspected mitochondrial disorders were recruited to the 100,000 Genomes Project (main programme) between 2015 and 2018 with an unexplained multi-system progressive disorder usually involving the central nervous system and/or neuromuscular system. All participants provided written informed consent and the study was approved by the HRA Committee East of England Cambridge South (REC Ref 14/EE/1112). Eligibility criteria stated that mtDNA and common nuclear genetic causes (e.g. *POLG* mutations) should have been excluded (full inclusion criteria in Supplementary Methods). All participants recruited under the category 'Suspected Mitochondrial Disorder' who had Tiering data available in data release v8_2019-11-28 were included in the study. In 2015-2018, genetic testing in the UK was arranged through 20 regional genetics laboratories, and there were three NHS highly specialised services for rare mitochondrial disorders. Testing of *POLG* and common mtDNA mutations (m.3243A>G associated with MELAS and maternally inherited diabetes and deafness [MIDD], m.8344A>G associated with myoclonic epilepsy with ragged red fibres [MERRF], and m.8993T>G/C associated with Leigh syndrome and neurogenic muscle weakness ataxia and retinitis pigmentosa [NARP]) in DNA extracted from blood was available through the highly specialised service laboratories. Further testing (such as gene panel testing) was available after discussion with the highly specialised services depending on the patient's clinical features (Supplementary Table 1).

Clinical Information

The referring clinician provided clinical information against a standardised list of clinical features as yes/no answers (Supplementary Table 2), or using a standardised vocabulary for describing clinical features encountered in human disease (the Human Phenotype Ontology, HPO[23]). HPO classifies clinical features by organ system using a branching tree incorporating increasing levels of detail. We used modified HPO terms to score participants using the Nijmegen Mitochondrial Disease Criteria (NMDC)[24,25](Supplementary Table 3), which uses clinical features (muscular, central nervous system and multi-system), MRI and biochemical features, and muscle biopsy results to give a score out of 12. Patients with a total of 0-1 were classified as being unlikely to have a mitochondrial disorder; score 2-4 a possible mitochondrial disorder; score 5-7 a probable mitochondrial disorder; and score 8 or more as definite mitochondrial disorder. One investigator scored all the participants. 100 participants were also scored independently by a second investigator to develop the modifications for using the NMDC with HPO terms. For each HPO term, the HPO system

was determined by following the branching tree back to the subtype of 'Phenotypic abnormality'.

Whole Genome Sequencing

DNA extraction from peripheral blood, quantification, and sequencing was performed according to a national specification (Illumina TruSeq, HiSeq 2500 and HiSeq X)[26,27] with reads aligned to the Genome Reference Consortium human genome build 37 (GRCh37) for the earlier participants recruited, and GRCh38 for later participants using Isaac Genome alignment software. Family based variant calling of single nucleotide variants and insertion deletions for chromosomes 1-22, X and the mtDNA was performed using the Platypus variant caller[28] allowing joint variant calling for all family members and considering the sequence alignments from all family members together.

The analysis workflow is shown in Figure 1.

Nuclear variant analysis – single nucleotide variants and small insertion-deletion variants (indels)

Genomes were analysed in families and variants were classified into four 'Tier' groups according to the probability of the variant being causative[21]. Tier 1 included loss of function variants (nonsense variants, essential splice donor and essential splice acceptor variants) and *de novo* missense or splice region variants in genes on the panels applied; Tier 2 included missense and splice region variants in genes on the panels applied; Tier 3 included other rare variants; and a final group of unclassified variants had higher population frequency or the segregation pattern in the family was not consistent with the family history. Virtual gene panels were chosen according to each participant's phenotypes, using curated 'PanelApp' gene lists, which include causative genes for each disorder generated through crowdsourcing[29]. This allowed the prioritisation of variants likely to be causative and minimised the reporting of abnormal genetic results which are unrelated to the reason for testing, for example cancer predisposition genes (referred to as incidental findings). All participants had the mitochondrial panel applied, and further panels depended on the phenotypes. Tier 1-3 variants were accessed from the Main Programme v8_2019-11-28. All Tiered variants had passed in-house Genomics England quality control[30]. Variants were also prioritised by Exomiser[31], an application which prioritises variants in exome or genome data using protein interaction networks, clinical relevance and cross-species phenotype

comparisons as well as computational filters for variant frequency, predicted pathogenicity and pedigree information. Variants classified as Pathogenic, Likely Pathogenic or Pathogenic/Likely Pathogenic were extracted from Clinvar[32] (3.3.2020) for GRCh38 and GRCh37 and compared against Tier 1-3 variants using bedtools intersect (https://bedtools.readthedocs.io) in order to identify previously reported pathogenic and likely pathogenic variants.

Gene panels applied

The Mitochondrial Disorders panel was applied in all participants. Other panels applied included Undiagnosed Metabolic Disorders (148 participants), Intellectual Disability (139), Congenital Myopathy (77), Hereditary Ataxia (60)(see Supplementary Table 4). The mean number of panels applied per participant was 4.7 (range 1 to 18). A total of 93 different panels were applied to between 1 and 345 participants.

Nuclear variant analysis – copy number variants

Copy number variant (CNV) calls were detected using Canvas software[33] based on sequence coverage and both nucleotide and Indel variant calling. We only included CNV calls with PASS filter status assigned by Canvas and overlapped any gene in PanelApp. CNV calls were annotated using gencode v29. CNVs which interrupted exons of a PanelApp gene were evaluated. All the CNVs reported as a diagnosis were manually confirmed on Integrative Genomics Viewer (IGV).

Nuclear variant analysis – short tandem repeat expansions

Short tandem repeat expansion genotyping was performed using the ExpansionHunter version 3.2.2 software package[34]. Thirteen loci were assessed (*HTT, AR, ATN1, ATXN1, ATXN2, ATXN3, ATXN7, CACNA1A, TBP, C9orf72, FXN*, *FMR1* and *DMPK*) using the coordinates listed in Supplementary Table 5. Potentially causative repeat expansions above an established threshold for each locus (Supplementary Table 6) were visually inspected using pileup plots and re-classified based on the quality of the reads[34].

mtDNA variant analysis

An in-house pipeline was used to call mtDNA single nucleotide variants above an established detection threshold of >1% variant allele frequency (VAF, or percentage heteroplasmy

level)[35] after excluding likely errors[27]. These were compared to a manually curated list of pathogenic mutations with functional evidence supporting pathogenicity[36]. Our pipeline does not detect large scale mtDNA rearrangements which usually require targeted mtDNA analysis in DNA extracted from skeletal muscle.

Clinical Review – Single nucleotide variants and small insertion-deletion variants (indels)
Tier 1 and Tier 2 variants, the top ten prioritised variants by Exomiser, ClinVar pathogenic/likely pathogenic variants and the mtDNA variants (tier 1-3 and from the in-house pipeline) were reviewed by a Clinical Geneticist and classified using internationally accepted criteria for pathogenic variants, likely pathogenic variants, variants of uncertain clinical significance, and likely benign or benign variants (American College of Medical Genetics [ACMG] criteria[37]), incorporating information from gnomAD[38], ensembl[39], VarSome[40], OMIM and a review of the literature (a description of the online resources used is in the Supplementary Methods). Variant quality and variant allele frequency were checked using IGV[41].

Clinical Review – Copy Number Variants
CNVs which overlapped at least one exon in a PanelApp gene were reviewed in DECIPHER. Information about the type of variant (copy number loss, copy number gain or loss of heterozygosity), the size of the variant and the gene content (in particular, haploinsufficient genes and OMIM morbid genes) was reviewed. A comparison was also made with previously observed CNVs in the general population or in affected individuals.

Clinical Review – Short tandem repeat expansions
Participants with repeat expansions were reviewed to see whether they were causative, considering the size of the expansion (whether it was a 'pre-mutation' or 'full mutation') and the participant's age and phenotypic features.

Feedback from Genomic Medicine Centre laboratories and clinical teams
The Genomic Medicine Centre laboratories were commissioned to analyse the Tier 1 and Tier 2 variants. They filled out an 'exit questionnaire' for each family through an online questionnaire. For each family, they fed back whether genetic cause had been identified with options of 'solved', 'partial', 'uncertain' or 'no' and whether the result was already known through clinical testing. For variants which the clinical scientists had identified as

potentially causative, they fed back their variant classification according the ACMG criteria and whether or not the variant had been confirmed by Sanger sequencing and fed back to the clinician. Feedback from the GMC laboratories was available in 277/319 (87%) families (data release v12_21_05_06). We successfully contacted clinicians for further information in 55 (17%) families.

Overall clinical assessment

The molecular diagnosis was described as 'definite', 'probable', or 'possible' based on our overall clinical assessment of whether the variant(s) explained the clinical features, taking into account the ACMG classification of the variant(s), the inheritance pattern and the clinical fit between the patient's HPO terms and the reported clinical phenotypes for the gene or variant. For example, we used 'probable' for compound heterozygotes where the phenotype fitted with the disorder but only one variant was classified as pathogenic/likely pathogenic and the other was a variant of uncertain clinical significance. The contribution was described as 'full' or 'partial' depending on whether the whole phenotype or only one aspect could be explained by the variant(s). Nuclear genes were classified as 'mitochondrial' based on the curated PanelApp list[29], plus recently discovered genes known to have a direct effect on oxidative phosphorylation.

Statistical Methods

Statistical analyses were performed in R. Fisher's exact test was used to compare the number of participants with a genetic diagnosis in children and adults, and in singletons compared to trios/quads. The mean number of HPO terms in diagnosed and undiagnosed individuals was compared using students t-test, and similarly for the number of HPO systems affected. Fisher's exact test was used to compare the proportion of families with each inheritance pattern between nuclear-mitochondrial diagnoses and non-mitochondrial disorders. The proportion of participants with each HPO system affected was compared between participants with definite mitochondrial diagnoses and definite non-mitochondrial diagnoses, using Fisher's exact test. The mean Mitochondrial Diagnostic Criteria score was compared between participants with mitochondrial diagnoses, non-mitochondrial diagnoses and undiagnosed participants using a One-way ANOVA and post-hoc Tukey testing.

Patient and Public Involvement

The 100, 000 Genomes Project has a Participant Panel made up of participants and parents or carers of people involved in the project which was established in 2016. The Panel meets with senior staff from Genomics England and NHS England four times a year. They are asked about project design and help ensure that participants' health data is looked after with respect and used in the best interests of the participants. Panel members sit on other committees including the Access Review Committee and the Ethics Advisory Committee. Patients and carers were involved in developing the consent literature and the infographics and information for the public.

Results

Demographics

345 affected individuals (186 females, 159 males) were referred with a suspected mitochondrial disorder after excluding common causes. These were from 319 families of different reported ethnicities across England (Figure 2). Genomic data were available for more than one affected member in 25 families (15 sibling pairs, eight mother and child pairs, one father and child pair, and one mother with two affected children). The median age at enrolment for probands was 25 years (IQR 10-54, range 0 to 92 years). 143 (41%) were aged <=18 at enrolment. No participants withdrew or were lost to follow up.

Phenotype data

Phenotypic data was available for 341 participants (missing in 4 participants). 3095 HPO terms were recorded, with a median of 7 HPO terms per participant (range 1 to 39). 806 different HPO terms were used. The commonest clinical terms, investigation result terms and HPO systems affected are shown in Figure 2d-f. A median of 4 HPO systems were affected per participant (range 1 to 13). Application of the HPO-modified Nijmegen score gave a mean total score of 4.30 (range 0-10), with 24 participants (7%) classified as being unlikely to have mitochondrial disease (score 0 or 1), 193 (56%) 'Nijmegen possible' (score 2-4), 95 (27.5%) 'Nijmegen probable' (score 5-7) and 33 (9.5%) 'Nijmegen definite' with a score of >=8.

Diagnostic Yield

The definite or probable genetic diagnosis was identified in 31% of families with a possible diagnosis in an additional 2% (Table 1, Figure 1, Figure 3a). A definite genetic diagnosis was reached in 28% (89/319) including 14 genetic diagnosis (4%) which provided only a partial explanation for their clinical features. Nine participants (3%) had a probable diagnosis and six (2%) had a possible diagnosis. One participant was diagnosed with two disorders which together fully explained their combination of phenotypes. The majority of diagnoses 69/104 (66%) came from single nucleotide variants (SNVs) or INDELs on the panels applied (Tier 1 or Tier 2 variants). An additional 19/104 (18%) diagnoses were made through other analyses of the SNVs and INDELS (from the GMC laboratory reports, comparison to ClinVar and Exomiser). The copy number variant analysis added 8 diagnoses (8%) and the short tandem repeat analysis added three diagnoses (3%). Five diagnoses were made through the mtDNA analysis (5%). Further details of the variant analysis are show in Supplementary Table 7.

Factors affecting diagnostic yield

The overall diagnostic rate was 59/186 (32%) in females and 51/159 (32%) in males. 64/143 (45%) participants recruited aged 18 or under received a diagnosis (any type), compared to 46/202 (23%) participants over the age of 18 ($p<0.001$). Considering definite diagnoses only, 50/143 (35%) participants under 18 received a definite molecular diagnosis compared to 30/202 (15%) adults over the age of 18 ($p<0.001$). Figure 3b shows the age profile for diagnosis. Although diagnoses were more frequent in younger people, diagnoses were still being made in participants who were in their 70s or 80s at enrolment.

The diagnostic yield per family (all diagnoses) was 23% for singletons (23/102), 29% (12/42) for duos with a parent, 42% (62/148) for trios/quads and 26% (7/27) for other family structures. The diagnostic rate in trios/quads was higher than in singletons ($p=0.005$). The mean age of singletons was 56 years, compared to a mean age of 12 years for probands in trios/quads.

The mean number of HPOs in participants with any diagnosis (10.0) was slightly higher than in undiagnosed patients (8.48) ($p=0.030$). The mean number of HPO systems affected in diagnosed patients (4.67) was not significantly different from the mean number of HPO systems affected in undiagnosed patients (4.61) ($p=0.82$).

Types of Diagnosis

Of the families with a definite diagnosis which fully explains the phenotype, 28 were in genes known to cause primary mitochondrial disease (37%) including four mtDNA variants and 24 diagnoses in nuclear-mitochondrial genes, whereas 47 were in non-mitochondrial genes (63%). For families with any genetic diagnosis (including probable or possible diagnoses and partial explanations), 39 were in genes known to cause primary mitochondrial disease (37.5%) including six mtDNA variants and 30 nuclear-mitochondrial diagnoses, and 65 were in non-mitochondrial genes (62.5%). These non-mitochondrial disorders included a wide variety of disorders, including developmental disorders with intellectual disability, metabolic disorders, epileptic encephalopathies, Bardet-Biedl syndrome, cardiomyopathies, *MYH2*-related myopathy and amyloidosis.

Potentially Treatable Disorders

There were potentially treatable disorders identified in six participants with a mitochondrial disorder and nine participants with a non-mitochondrial disorder (shown in Supplementary Table 8). However, it should be noted that the evidence base is weak for several of these rare disorders[42].

Mitochondrial Diagnoses compared to Non-Mitochondrial Diagnoses

Inheritance Patterns: 70% (23/33) of families with a nuclear-mitochondrial diagnosis showed an autosomal recessive inheritance pattern, compared to 35% (23/65) of families with non-mitochondrial disorders. Five families (15%) had *de novo* diagnostic variants in nuclear-mitochondrial genes (*AIFM1*, *LONP1* and *PDHA1* and two with de novo duplications in the *ATAD3* gene cluster*)* compared to 29% (19/65) of families with non-mitochondrial disorders having a *de novo* diagnosis (Figure 3c), particularly those with intellectual disability or epileptic encephalopathy. The *HK1* variant was *de novo* in two siblings with presumed germline mosaicism (somatic mosaicism was not detected in either parent). The proportions with each inheritance pattern were significantly different between nuclear-mitochondrial diagnoses and non-mitochondrial disorders ($p=0.007$).

HPO Systems affected: In our cohort, participants with a mitochondrial diagnosis were more likely to have HPO-terms in the Metabolism/Homeostasis system ($p<0.001$) such as increased lactate or decreased mitochondrial complex activities, compared to those with non-mitochondrial diagnoses (Figure 4).

Modified Nijmegen Mitochondrial Disease Criteria scores: The majority (22/30, 73%) of participants with genetically confirmed mitochondrial diagnoses (nuclear-mitochondrial and mtDNA) had scores in the Nijmegen probable (5-7) and Nijmegen definite (8-12) modified Nijmegen MDC category (Figure 3d), whereas 60% (30/50) of participants with confirmed non-mitochondrial disorders had scores in the Nijmegen possible range (2-4). The mean MDC score differed between diagnostic groups ($p<0.001$), including between the mitochondrial diagnoses and the non-mitochondrial diagnoses ($p<0.001$), and between the mitochondrial diagnoses and patients with no diagnosis ($p<0.001$), but not for the non-mitochondrial diagnoses and undiagnosed patients ($p=0.36$). However, 16 of the 50 (32%) participants with confirmed non-mitochondrial disorders had scores in the 'Nijmegen probable' range, and 3 (6%) in the 'Nijmegen definite' range. The sensitivity and specificity of the MDC score in our cohort is shown in Supplementary Table 9. Using 'Nijmegen definite', the MDC score was highly specific but sensitivity was only 40%, whereas using a 'Nijmegen probable and definite' scores, the score was moderately sensitive and moderately specific.

Muscle biopsy findings and mitochondrial complex activities

117 participants (34%) had HPO terms relating to muscle biopsy abnormalities, and 73 (21%) had abnormal mitochondrial complex activities. In patients with definite mitochondrial diagnoses, 15/28 (54%) had muscle biopsy abnormalities, 11/28 (39%) had abnormal respiratory chain complex activities and two had abnormal PDH complex activity. In patients with definite non-mitochondrial diagnoses, 14/47 (30%) had muscle biopsy abnormalities and 12/47 (26%) had abnormal respiratory chain complex activities. The patients with non-mitochondrial diagnoses and abnormal respiratory chain complex activities had pathogenic variants in *ASXL3*, *CACNA1E*, *CTBP1*, *EXOSC3*, *HK1*, *KCNT1*, *NPHP1*, *P4HTM*, *PPP2R5D* and *SCN2A*, *SOS1* and *TANGO2*.

Discussion

In patients referred for WGS with a suspected mitochondrial disorder, a definite or probable diagnosis was identified in 31% (including partial diagnosis in 4%) and a further 2% had a possible diagnosis. The mitochondrial diagnoses were nearly all unique to one family, reflecting the high level of genetic heterogeneity in mitochondrial disorders. Non-mitochondrial disorders were more common than mitochondrial disorders and had features

resembling mitochondrial diseases (often referred to as 'phenocopies'). These could be broadly classified as developmental disorders with intellectual disability, metabolic disorders, myopathies, cardiomyopathies, epileptic encephalopathies, leukodystrophies, ciliopathies, amyloidosis, and other neurogenetic disorders, including basal ganglia calcification and neurodegeneration with iron accumulation. 29% of the non-mitochondrial disorders were caused by *de novo* pathogenic variants.

The diagnostic yield was significantly greater in children than adults. There are several possible explanations for this. First, children were more likely to be recruited as trios with both parents. This makes analysis more straightforward[43] because *de novo* variants can be identified, rare familial variants can be filtered out if the parents are clinically unaffected for recessive disorders, and it can be checked that the variants were inherited from both parents. Second, the most severe phenotypes are seen in children because affected individuals do not survive until adulthood[44]. Severe phenotypes are the most likely to be caused by single gene disorders. Milder phenotypes overlap with acquired disorders, are less likely to be caused by a single gene defect, and are more common in adults. Third, most adults with genetically proven mitochondrial disorders have a mutation in the mtDNA which were excluded by the clinical laboratories before inclusion in this study. Studies of whole exome sequencing in rare disease have also noted a decrease in diagnostic yield with an increasing age of the probands[45]. Despite this, in our study, new genetic diagnoses were made across the whole age spectrum. The oldest patient to receive a genetic diagnosis was recruited at age 86.

There was also variability in the number of people recruited in different age groups. The highest numbers were recruited in the paediatric age group, with a second peak of patients aged over 60 years. This could be because some adult-onset mitochondrial disorders, such as progressive external ophthalmoplegia (CPEO), take time to develop and only become obvious in later life (CPEO was present in 17/71(24%) in over 60s compared to 16/274 (6%) in under 60s). Alternatively, this could reflect the fact that genetic testing had not have been offered previously to the older individuals due to the financial cost and perceived lack of immediate management implications. The 30-39 years age group were the least likely to be recruited and had a low diagnostic rate. A common reason for seeking a clinical genetics referral is for reproductive advice, so we speculate that this age group are more likely to have been reviewed by a clinical geneticist and offered up to date genetic testing (such as panel

testing) in a clinical setting, meaning that there were fewer undiagnosed patients and that the remaining ones were more difficult to diagnose.

It is possible that the relatively high frequency of partial diagnoses in this study (4%) reflects our current knowledge of the phenotypic spectrum of ultra-rare genetic diseases, and that some of the features we have not ascribed to the causal variants are actually due to the underlying mutations. This is likely to be a particular problem for mitochondrial disorders because of their diverse phenotypes, some of which are only just being recognised[36], but will become easier as our knowledge base increases.

Strengths and weaknesses of the study in relation to other studies

Our study had several strengths, including the large sample size, and the fact that patients were recruited nationally from both secondary and tertiary care, meaning that the findings have a wider relevance than studies focussed on highly selected groups identified by specialist centres. In addition, the use of HPO terms allowed phenotype data to be analysed in a systematic way, and we contacted clinicians for detailed phenotypic information in selected participants. Our analysis of the nuclear genes included copy number variants and short tandem repeat expansions in addition to single nucleotide variants and small indels; and we also studied mtDNA variants with a heteroplasmy level >1%. The advantage of using WGS rather than WES was that it is easier to pick up CNVs, repeat expansions and lower level heteroplasmies. Weaknesses are that we have not explored novel disease genes or variants in the non-coding regions, other than previously published variants, and it was not possible to trace all family members.

The previously published study using whole genome sequencing in patients with suspected mitochondrial disease[19] looked at a cohort of 40 children recruited from four paediatric genetic metabolic centres in Australia. 34/40 had a ‘probable’ or ‘definite’ mitochondrial disorder using the modified Nijmegen mitochondrial disease criteria score (between 5 and 12) and 28 had abnormal respiratory chain enzyme activities. Analysing nuclear and mtDNA enabled a definitive genetic diagnosis for 55% of patients, and a likely molecular diagnosis in 67%, with 17.5% having a non-mitochondrial disorder. The higher diagnostic yield in this study[19] likely reflects the section criteria which focussed on children within trios identified by national specialist clinics, compared to our more inclusive recruitment criteria where ~2/3 of our probands were adults, and only 46% of families were recruited as trios.

A number of studies using whole exome sequencing in patients with suspected mitochondrial disorders have been reviewed recently[46]. Most studies of these studies were conducted in highly selected patients seen in specialist centres[10,12,16], focussing on childhood-onset disorders (apart from Ref 14 which included patients up to 27 years of age). Their stringent recruitment criteria led to higher diagnostic yields (57-68%) and lower number of non-mitochondrial diagnoses in comparison with our study. Previous exome studies of larger cohorts[14,15] have also tended to have a lower diagnostic yield (34.5- 39%) which is thought to be more reflective of everyday clinical practice[46]. Non-mitochondrial disorders have been found in some of the previous studies, but in much smaller numbers compared to this study. The recruitment criteria in our study were broad, which is reflective of mainstreaming of genomic medicine and we studied both adults and children. Most participants (64%) did not have muscle biopsy results available at the time of referral, whereas most participants in previous studies had evidence of mitochondrial dysfunction. The more inclusive eligibility criteria have led to finding a wide range of different genetic diagnoses. The finding that a large number of patients had non-mitochondrial disorders is important because these diagnoses would have been missed if the participants had been investigated only for mitochondrial disorders through muscle biopsy and/or a specific mitochondrial gene panel. This would have led to missed opportunities for treatment, surveillance and reproductive management. Our findings highlight the difficulty of diagnosing these rare multisystem disorders clinically and the need to keep an open mind about the differential diagnosis.

Considering other studies of integration of whole genome sequencing into healthcare, Stranneheim *et al*[47] describe the results for 3219 rare disease patients recruited in Stockholm, Sweden between 2015 and 2019. Their study involved much more direct collaboration between academia, healthcare and their SciLifeLab (which provides WGS), which were all located in the same city. This meant that potential diagnoses were discussed and specialist advice was fed back quickly into clinical practice. Specialist clinicians could also access and analyse the data together with clinical scientists. The 100,000 genomes project used a different model with the bioinformatics managed centrally and the clinical interpretation of variants done by Genomic Medicine Centre laboratories. Researchers were able to access pseudonymised data. The limited ability researchers and clinicians to discuss potential diagnoses together and the long turn-around time were disadvantages compared to the Swedish model.

Unanswered questions and future research

The participants in this study had standard of care NHS genetic testing prior to enrolment. The use of WGS as a first genetic test for suspected mitochondrial disorders has not been directly explored. The only mitochondrial diseases which cannot be diagnosed by WGS of DNA extracted from blood are extremely rare muscle specific mtDNA mutations and large mtDNA deletions which usually have usually a very specific clinical phenotype. Based on known epidemiology[1,50], these muscle specific mtDNA mutations and deletions account for ~11.5% of genetically confirmed mitochondrial disorder patients. Therefore nearly 90% of mitochondrial disorder patients can be diagnosed by WGS of DNA extracted from blood[51]. Additionally, WGS can diagnose other monogenic disorders which have similar clinical features, so we would expect a very high diagnostic yield. Finally, by carrying out this study we have identified a cohort of patients with suspected mitochondrial disease where the likely diagnosis is contained within the WGS sequence data, but cannot be distinguished from background sequence variation. This will form a useful resource for the discovery of future mitochondrial disease genes that are not in the coding space, and thus are not detectable by WES. Large scale mtDNA rearrangements are not currently reliably detected in DNA extracted from blood in adults, meaning that patients with suspected CPEO and Kearns-Sayre syndrome are investigated by muscle biopsy[20]. Future research will determine whether deep mtDNA sequencing will detect these large scale mtDNA rearrangements in DNA extracted from blood or another accessible tissue such as urinary epithelium.

Policy implications

Our findings indicate that whole genome sequencing is a useful diagnostic test in patients with suspected mitochondrial disorders recruited from secondary and tertiary care settings. We recommend that whole genome sequencing should be offered early in the diagnostic pathway in a patient's local secondary or tertiary care centre, and before invasive tests such as a muscle biopsy. Exceptions to this would be patients whose clinical features are highly suggestive of a specific cause which can be confirmed by a single gene test or common mtDNA mutation testing, and also patients with progressive external ophthalmoplegia which is currently diagnosed in most patients by testing mtDNA from a muscle biopsy sample for a large-scale rearrangements. Referral to a specialised mitochondrial clinic should be arranged if a mitochondrial diagnosis is confirmed, or for further investigations if WGS sequencing is uninformative. Further investigations likely to increase the diagnostic yield beyond WGS

include laboratory studies of mitochondrial function and other ‘omics’ approaches including transcriptomics, which provided an additional diagnosis in 10% of suspected mitochondrial disease patients in one study[48], proteomics, and metabolomics.

There are also wider policy implications for the integration of whole genome sequencing into healthcare. The relatively high number of patients with probable or possible diagnoses partly reflects the current lack of capacity for the functional evaluation of variants of uncertain clinical significance, for example through splicing assays or Western blotting. Resources should also be made available for the regular re-analysis of WGS data, either at specified time intervals in undiagnosed patients or on clinician request. The re-analysis of exome data has been shown to significantly increase diagnostic yield, mainly due to newly discovered disease genes[49]. Finally, rapid trio whole genome sequencing[17] should be offered in acutely unwell individuals with suspected mitochondrial disorders, so that the results can help guide clinical management. In the UK, rapid trio exome sequencing is currently available for acutely unwell children only.

Author Statements

Contributors

KRS, RH and PFC designed the study. WW, CC, AT and KA did the bioinformatic analysis. KRS, RH, AT, RDSP, RQ, EB, MH, EC, JS, PB, DJ, LI, CN, TB, AS, CD, SFS, RF, HH, SR and PFC recruited patients. KRS, RH, AT, TR EA, CF, DM and PFC did clinical interpretation of genetic variants. MC was the chief scientific officer for Genomics England and oversaw the 100,000 genomes project. KRS and PFC wrote the manuscript. All authors reviewed and revised the manuscript. PFC is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding

This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support. PFC is a Wellcome Trust Principal Research Fellow (212219/Z/18/Z), and a UK NIHR Senior Investigator, who receives support from the Medical Research Council Mitochondrial Biology Unit (MC_UU_00015/9), the Medical Research Council (MRC) International Centre for Genomic Medicine in Neuromuscular Disease (MR/S005021/1), the Leverhulme Trust (RPG-2018-408), an MRC research grant (MR/S035699/1), an Alzheimer's Society Project Grant (AS-PG-18b-022) and the National Institute for Health Research (NIHR) Biomedical Research Centre based at Cambridge University Hospitals NHS Foundation Trust and the University of Cambridge. KRS is supported by Addenbrooke’s Charitable Trust and the Medical Research Council (MRC) International Centre for Genomic Medicine in Neuromuscular Disease (MR/S005021/1). RH is a Wellcome Trust Investigator (109915/Z/15/Z), who receives support from the Medical Research Council (UK) (MR/N025431/1), the European Research Council, the Newton Fund (UK/Turkey,

MR/N027302/1), the Evelyn Trust and the Lily Foundation. 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. R.D.S.P. is supported by a Medical Research Council Clinician Scientist Fellowship (MR/S002065/1). R.D.S.P. and M.G.H. are funded by a Medical Research Council strategic award to establish an International Centre for Genomic Medicine in Neuromuscular Diseases (ICGNMD) (MR/S005021/1). RF is funded by Imperial NIHR BRC Imperial College Healthcare Trust. All research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Competing Interests

All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/disclosure-of-interest/and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval

Ethical approval for the 100,000 Genomes Project was received from the HRA Committee East of England Cambridge South (REC Ref 14/EE/1112).

Data sharing

100,000 genomes project data is available to researchers and clinicians through joining a Genomics England Clinical Interpretation Partnership (GeCIP) htttps://www.genomicsengland.co.uk/about-gecip/joining-research-community/

Transparency

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities

We will distribute the article to clinicians and advocacy groups including the Lily Foundation, Muscular Dystrophy UK, and the United Mitochondrial Disease Foundation. We will distribute on social media and a plain-language summary on the Cambridge Clinical Mitochondrial Research Group website https://www-neurosciences.medschl.cam.ac.uk/mitocamb/. We will issue a press release through the University of Cambridge and Genomics England.

Provenance and peer review

Not commissioned, externally peer reviewed.

References

1. Gorman GS, Schaefer AM, Ng Y, Gomez N, Blakely EL, Alston CL, et al. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. Ann Neurol. 2015; 77:753–9.

2. Vafai SB, Mootha VK. A Common Pathway for a Rare Disease? Science. 2013;342:1453–4.

3. Jurkute N, Yu-Wai-Man P. Leber hereditary optic neuropathy: bridging the translational gap. Curr Opin Ophthalmol. 2017;28:403–9.

4. Gorman GS, Chinnery PF, DiMauro S, Hirano M, Koga Y, McFarland R, et al. Mitochondrial diseases. Nat Rev Dis Primer. 2016;2:16080.

5. Rahman J, Rahman S. Mitochondrial medicine in the omics era. Lancet Lond Engl. 2018;391:2560–74.

6. Rahman J, Noronha A, Thiele I, Rahman S. Leigh map: A novel computational diagnostic resource for mitochondrial disease. Ann Neurol. 2017;81:9–16.

7. Black N, Martineau F, Manacorda T. Diagnostic odyssey for rare diseases: exploration of potential indicators. PIRU publication. 2015 https://piru.ac.uk/assets/files/Rare%20diseases%20Final%20report.pdf

8. Grier J, Hirano M, Karaa A, Shepard E, Thompson JLP. Diagnostic odyssey of patients with mitochondrial disease: Results of a survey. Neurol Genet. 2018;4:e230.

9. McFarland R, Taylor RW, Turnbull DM. A neurological perspective on mitochondrial disease. Lancet Neurol. 2010;9:829–40.

10. Taylor RW, Pyle A, Griffin H, Blakely EL, Duff J, He L, et al. Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. JAMA. 2014;312:68–77.

11. Pronicka E, Piekutowska-Abramczuk D, Ciara E, Trubicka J, Rokicki D, Karkucińska-Więckowska A, et al. New perspective in diagnostics of mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric centre. J Transl Med. 2016;14:174.

12. Puusepp S, Reinson K, Pajusalu S, Murumets Ü, Õiglane-Shlik E, Rein R, et al. Effectiveness of whole exome sequencing in unsolved patients with a clinical suspicion of a mitochondrial disorder in Estonia. Mol Genet Metab Rep. 2018;15:80–9.

13. Lieber DS, Calvo SE, Shanahan K, Slate NG, Liu S, Hershman SG, et al. Targeted exome sequencing of suspected mitochondrial disorders. Neurology. 2013;80:1762–70.

14. Wortmann SB, Koolen DA, Smeitink JA, van den Heuvel L, Rodenburg RJ. Whole exome sequencing of suspected mitochondrial patients in clinical practice. J Inherit Metab Dis. 2015;38:437–43.

15. Kohda M, Tokuzawa Y, Kishita Y, Nyuzuki H, Moriyama Y, Mizuno Y, et al. A Comprehensive Genomic Analysis Reveals the Genetic Landscape of Mitochondrial Respiratory Chain Complex Deficiencies. PLoS Genet. 2016;12:e1005679.

16. Theunissen TEJ, Nguyen M, Kamps R, Hendrickx AT, Sallevelt SCEH, Gottschalk RWH, et al. Whole Exome Sequencing Is the Preferred Strategy to Identify the Genetic Defect in Patients With a Probable or Possible Mitochondrial Cause. Front Genet. 2018;9:400.

17. French CE, Delon I, Dolling H, Sanchis-Juan A, Shamardina O, Mégy K, et al. Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. Intensive Care Med. 2019;45:627–36.

18. Poulton J, Steffann J, Burgstaller J, McFarland R, workshop participants. 243rd ENMC international workshop: Developing guidelines for management of reproductive options for families with maternally inherited mtDNA disease, Amsterdam, the Netherlands, 22-24 March 2019. Neuromuscul Disord NMD. 2019;29:725–33.

19. Riley LG, Cowley MJ, Gayevskiy V, Minoche AE, Puttick C, Thorburn DR, et al. The diagnostic utility of genome sequencing in a pediatric cohort with suspected mitochondrial disease. Genet Med Off J Am Coll Med Genet. 2020

20. Schon KR, Ratnaike T, van den Ameele J, Horvath R, Chinnery PF. Mitochondrial Diseases: A Diagnostic Revolution. Trends Genet TIG. 2020;36:702–17.

21. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ. 2018;361

22. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, et al. Whole-genome sequencing of patients with rare diseases in a national health system. Nature. 2020;583:96–102.

23. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014;42:D966-974.

24. Bernier FP, Boneh A, Dennett X, Chow CW, Cleary MA, Thorburn DR. Diagnostic criteria for respiratory chain disorders in adults and children. Neurology. 2002;59:1406–11.

25. Morava E, van den Heuvel L, Hol F, de Vries MC, Hogeveen M, Rodenburg RJ, et al. Mitochondrial disease criteria: diagnostic applications in children. Neurology. 2006;67:1823–6.

26. Caulfield M, Davies J, Dennys M, Elbahy L, Fowler T. The National Genomics Research and Healthcare Knowledgebase v5. 2019 https://www.genomicsengland.co.uk/wp-content/uploads/2019/08/The-National-Genomics-Research-and-Healthcare-Knowledgebase-v5-1.pdf

27. Wei W, Tuna S, Keogh MJ, Smith KR, Aitman TJ, Beales PL, et al. Germline selection shapes human mitochondrial DNA diversity. Science. 2019;364

28. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, WGS500 Consortium, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014;46:912–8.

29. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet. 2019;51:1560–5.

30. Baple E. Genomics England Rare Disease Results Guide. 2020; Document Key/Version Number:BIO-GUI-018/V5.0.

31. Robinson PN, Köhler S, Oellrich A, Sanger Mouse Genetics Project, Wang K, Mungall CJ, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res. 2014;24:340–8.

32. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46:D1062–7.

33. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. Bioinformatics. 2016;32:2375–7.

34. Ibanez K, Polke J, Hagelstrom T, Dolzhenko E, Pasko D, Thomas E, et al. Whole genome sequencing for diagnosis of neurological repeat expansion disorders. bioRxiv. 2020;2020.11.06.371716.

35. Calabrese C, Simone D, Diroma MA, Santorsola M, Guttà C, Gasparre G, et al. MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. Bioinforma Oxf Engl. 2014;30:3115–7.

36. Ratnaike TE, Greene D, Wei W, Sanchis-Juan A, Schon KR, van den Ameele J, et al. MitoPhen database: a human phenotype ontology-based approach to identify mitochondrial DNA diseases. Nucleic Acids Res. 2021;49:9686-9695.

37. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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. Genet Med Off J Am Coll Med Genet. 2015;17:405–24.

38. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.

39. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020;48:D682–8.

40. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. Bioinformatics. 2019;35:1978–80.

41. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.

42. Yung Tiet M, Lin Z, Gao F, Jennings MJ, Horvath R. Targeted Therapies for Leigh Syndrome: Systematic Review and Steps Towards a 'Treatabolome'. J Neuromuscul Dis. 2021; doi: 10.3233/JND-210715. Online ahead of print.

43. Tan TY, Lunke S, Chong B, Phelan D, Fanjul-Fernandez M, Marum JE, et al. A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis. Eur J Hum Genet EJHG. 2019;27:1791–9.

44. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nat Rev Genet. 2018;19:253–68.

45. Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, Wang X, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. Genet Med Off J Am Coll Med Genet. 2016;18:678–85.

46. Stenton SL, Prokisch H. Genetics of mitochondrial diseases: Identifying mutations to help diagnosis. EBioMedicine. 2020;56:102784.

47. Stranneheim H, Lagerstedt-Robinson K, Magnusson M, Kvarnung M, Nilsson D, Lesko N, et al. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. Genome Med. 2021;13:40.

48. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Commun. 2017;8:15824.

49. Liu P, Meng L, Normand EA, Xia F, Song X, Ghazi A, et al. Reanalysis of Clinical Exome Sequencing Data. N Engl J Med. 2019;380:2478–80.

50. Castro-Gago M, Blanco-Barca MO, Campos-González Y, Arenas-Barbero J, Pintos-Martínez E, Eirís-Puñal J. Epidemiology of pediatric mitochondrial respiratory chain disorders in northwest Spain. Pediatr Neurol. 2006;34:204–11.

51. Raymond FL, Horvath R, Chinnery PF. First-line genomic diagnosis of mitochondrial disorders. Nat Rev Genet. 2018;19:399–400.

Tables & Figures

Table 1. Variants identified in patients with definite, probable and possible diagnoses. Genetic diagnoses made in 104 families, including age and sex of participants, contribution to the phenotype from this gene (full or partial), gene, variants, ACMG variant classification (P=Pathogenic, LP=Likely Pathogenic, VUS= Variant of Uncertain Clinical Significance) and inheritance pattern. (A) Variants in mitochondrial genes are shown in the upper panel and (B) those in non-mitochondrial genes are shown in the lower panel table. Novel variants are indicated with # and five previously published families are also indicated with $. For

family 84, the full phenotype is explained by the two genetic diagnoses. In family 77, the NPHP1 homozygous deletion explains the phenotype in the sibling with a severe phenotype (with childhood onset end stage renal disease) and is heterozygous in the other sibling.

Table 1A Mitochondrial Diagnoses

	Age (y)	Sex	Contribution	Gene	Variants	ACMG	Inheritance
Definite							
1	10	F	Full	*AARS2*	NM_020745.4:c.302G>A p.(Arg101His) Homozygous	LP	Biallelic
2	1	M	Full	*AIFM1*	NM_004208.4: c.603_605del p.(Arg201del)	P	XLR *de novo*
3	0	F	Full	*ATAD3*	Duplication in ATAD3 gene cluster	P	Monoallelic *de novo*
4	0	M	Full	*ATAD3*	Duplication in ATAD3 gene cluster	P	Monoallelic *de novo*
5	13	M	Full	*C12orf65*	NM_152269.5:c.210del p.(Gly72fs) NM_152269.5:c.258_270dupCATCCCCTCAGGC p.(Ile91fs)#	P P	Biallelic
6	6	M	Full	*EARS2*	NM_001083614.2:c.184A>T p.(Ile62Phe) # Homozygous	LP	Biallelic
7	2	F	Full	*FBXL4*	NM_001278716.2:c.1641_1642delTG p.(Cys547Ter) NM_001278716.2:c.141delC p.(Asn48fs) #	P P	Biallelic
8	18	F	Full	*HIBCH*	NM_014362.4:c.1126T>G p.(Phe376Val) # Homozygous	LP	Biallelic
9	22	F	Full	*KARS1*	NM_001130089.1:c.683C>T p.(Pro228Leu) NM_001130089.1:c.774A>T p.(Arg258Ser)	P LP	Biallelic
10	1	F	Full	*MRPL44*	NM_022915.4:c.467T>G p.(Leu156Arg) Homozygous	P	Biallelic
11	23	M	Full	*MRPS25* $	NM_022497.5:c.215C>T p.(Pro72Leu) Homozygous	LP	Biallelic
12	42	F	Full	*MT-ATP6*	m.8618dupT 14% heteroplasmy	P	mtDNA very low level in mother
13	24	M	Full	*MT-ATP6*	m.8969G>A 84% Heteroplasmy	P	mtDNA *de novo*
14	26	F	Full	*MT-ND3*	m.10158T>C 23% Heteroplasmy	P	mtDNA *de novo*
15	67	M	Partial	*MT-ND6*	m.14484T>C Homoplasmic	P	mtDNA unknown
16	18	M	Full	*MT-TE*	m.14674T>C Homoplasmic	P	mtDNA maternally inherited
17	18	F	Partial	*MT-RNR1*	m.1555A>G Homoplasmic	P	mtDNA maternally inherited
18	10	F	Full	*MTO1*	NM_012123.4:c.1232C>T p.(Thr411Ile) Homozygous	P	Biallelic
19	13	F	Full	*NDUFAF5*	NM_024120.5:c.480-3T>G# NM_024120.5:c.827G>A p.(Arg276Gln) #	LP LP	Biallelic
20	1	M	Full	*NDUFAF8* $	NM_001086521.2:c.45_52dup p.(Phe18fs) NM_001086521.2:c.195+271C>T	P LP	Biallelic

21	7	M	Full	*OPA1*	NM_015560.2:c.2708_2711delTTAG (splice acceptor variant) NM_015560.2:c.1146A>G p.(Ile382Met)	P P	Biallelic	
22	11	F	Full	*PDHA1*	NM_000284.4:c.434G>A p.(Cys145Tyr) #	LP	XLD *de novo*	
23	17 13	F F	Full	*PDP1*	NM_018444.4:c.571C>T p.(Gln191Ter) # Homozygous	LP	Biallelic	
24	56	M	Full	*POLG*	NM_002693.3:c.1399G>A p.(Ala467Thr) Homozygous	P	Biallelic	
25	48 71	F F	Full	*RRM2B*	NM_001172477.1:c.242G>A p.(Asp142Asn)	P	Monoallelic (AD)	
26	0	M	Full	*SCO2*	NM_001169109.1:c.418G>A p.(Glu140Lys) NM_001169109.1:c.625_627delTAC p.(Tyr209del) #	P LP	Biallelic	
27	0	F	Full	*SCO2*	NM_001169109.1:c.323A>G p.(Asp108Gly) # NM_001169109.1:c.281T>C p.(Leu94Pro) #	LP LP	Biallelic	
28	74	F	Full	*SLC25A4*	NM_002252.4:c.311A>G p.(Asp104Gly)	P	Monoallelic (unknown)	
29	20	F	Full	*TTC19*	NM_017775.4:c.184+1G>A# NM_017775.4:c.275_278delCCGA p.(Ala92fs) #	P LP	Biallelic	
30	66	M	Full	*TWNK*	NM_021830.5:c.1374G>T p.(Gln458His)	P	Monoallelic (unknown)	
Probable								
31	61	M	Full	*DNM1L*	NM_012062.5:c.239_241delGAG p.(Gly80del) #	LP	Monoallelic Unknown	
32	4	F	Full	*ELAC2*	NM_018127.7:c.2009delG p.(Cys670fs) NM_018127.7:c.2245C>T p.(His749Tyr) #	P VUS	Biallelic	
33	7	F	Full	*GFER*	NM_005262.3:c.199delC p.(Arg67fs) NM_005262.3:c.259-28C>G#	P VUS	Biallelic	
34	19	M	Full	*MTFMT*	NM_139242.4:c.626C>T p.(Ser209Leu) NM_139242.4:c721+5G>A#	P VUS	Biallelic	
35	4	F	Full	*RRM2B*	NM_001172477.1:c.578G>A p.(Arg193His) NM_001172477.1:c1253C>A p.(Thr418Asn)#	LP VUS	Biallelic	
36	3	M	Full	*SDHA*	NM_004168.4:c.290G>C p.(Arg97Thr) # NM_004168.4:c.424A>G p.(Met142Val) #	VUS VUS	Biallelic	
Possible								
37	18	M	Full	*LONP1*	NM_004793.4: c.1694A>G p.(Tyr565Cys)	VUS	De novo	
38	0	M	Full	*PDHA1*	NM_000284.4: c.759+5G>T	VUS	XL (from unaffected mother)	
39	56 54	M F	Full	*TOP3A*	NM_004618.5:c.284C>T p.(Ala95Val) # NM_004618.5:c.109C>G p.(Leu37Val) #	VUS VUS	Presumed Biallelic (parents not tested)	

Table 1B Non-Mitochondrial Diagnoses

	Age (y)	Sex	Contribution	Gene	Variants	ACMG	Inheritance
Definite							
40	0	F	Full	*ACTA2*	NM_001614.4:c.536G>A p.(Arg179His)	P	Monoallelic *de novo*
41	42	F	Full	*AMACR*	NM_014324.6:c.857delT p.(Ile286fs) # NM_014324.6:c.437C>T p.(Pro146Leu) #	P LP	Biallelic
42	57	M	Full	*AMACR*	NM_014324.6:c.154T>C p.(Ser52Pro) Homozygous	P	Biallelic
43	71	F	Full	*AMACR*	NM_014324.6:c.154T>C p.(Ser52Pro) Homozygous	P	Biallelic
44	9	F	Full	*AMPD2*	NM_001368809.2: c.2228T>C p.(Leu743Pro)# Homozygous	LP	Biallelic
45	29 58	M F	Full	*APP*	NM_000484.4:c.2075C>G p.(Ala692Gly)	P	Monoallelic AD
46	18	M	Partial	*ASL*	NM_000048.4:c.1153C>T p.(Arg385Cys) Homozygous	P	Biallelic
47	15	M	Full	*ASXL3*	NM_030632.3:c.3464c>A p.(Ser1155Ter)	P	Monoallelic *de novo*
48	15	M	Full	*ATP1A3*	NM_152296.5:c.2452G>A p.(Glu818Lys)	P	Monoallelic *de novo*
49	86	F	Full	*ATP1A3*	NM_152296.5:c.2452G>A p.(Glu818Lys)	P	Monoallelic Unknown
50	5	M	Full	*ATP6V1A*	NM_001690.4:c.845A>T p.(Asn282Ile) #	LP	Monoallelic *de novo*
51	1	F	Full	*ATXN7* $	Very large CAG repeat expansion	P	Monoallelic
52	44	F	Full	*BBS1*	NM_024649.5:c.1169T>G p.(Met390Arg) Homozygous	P	Biallelic
53	2	M	Full	*BCAP31*	NM_001256447.2:c.565C>T p.(Gly189Ter) #	P	XLR *de novo*
54	26	F	Full	*C19orf12*	NM_001256047.1:c.245dupC p.(Ala83fs) #	LP	Monoallelic unknown
55	12	F	Full	*CACNA1A*	NM_001127221.1:c.4177G>A p.(Val1393Met)	LP	Monoallelic *de novo*
56	2	M	Full	*CACNA1E*	NM_001205293.3:c.683T>C p.(Leu228Pro)	LP	Monoallelic *de novo*
57	12	M	Full	*CTBP1*	NM_001328.3:c.1024C>T p.(Arg342Trp)	P	Monoallelic *de novo*
58	11 10	M F	Full	*DOCK6*	NM_020812.4:c.4106+5G>T NM_020812.4:c.1902_1905delGTTC p.(Phe635fs)	LP P	Biallelic
59	12	F	Full	*DSP*	NM_004415.4:c.1799T>C p.(Phe600Ser) #	LP	Monoallelic *de novo*
60	8	F	Full	*EXOSC3*	NM_016042.4:c.395A>C p.(Asp132Ala) Homozygous	P	Biallelic
61	61	M	Partial	*EYA4*	NM_004100.5:c.1741A>T p.(Lys581Ter) #	LP	Monoallelic Unknown
62	2	M	Full	*FIG4*	NM_014845.6:c.447-2A>G# NM_014845.6:c.827C>T p.(Ser276Phe) #	P LP	Biallelic
63	6	F	Full	*GCDH*	NM_000159.4:c.1204C>T p.(Arg402Trp) NM_000159.4:c.1304C>T p.(Thr435Met) #	P LP	Biallelic
64	3	F	Full	*HADHA*	NM_000182.5:c.1528G>C p.(Glu510Gln) NM_000182.5:c.1664T>G p.(Met555Arg) #	P LP	Biallelic
65	10	F	Full	*HK1*	NM_000188.3:c.1334C>T p.(Ser445Leu)	P	Monoallelic

	8	F					*de novo*
66	9	M	Full	*HSD17B4*	NM_000414.4:c.590_597dupGATCACGG p.(Met200fs)* NM_000414.4:c.743G>A p.(Arg248His) #	P LP	Biallelic
67	71	M	Partial	*HTT* $	~40 CAG repeats	P	Monoallelic
68	7	F	Full	*HTT* $	Very large CAG repeat expansion	P	Monoallelic
69	41	M	Partial	*KCNQ4*	NM_004700.4:c.961G>A p.(Gly321Ser)	P	Monoallelic unknown
70	6	F	Full	*KCNT1*	NM_020822.3:c.1885A>c p.(Lys629Gln)	LP	Monoallelic *de novo*
71	46	F	Full	*KIF11*	NM_004523.4:c.78-2A>G#	P	Monoallelic AD
72	29	M	Partial	*KMT2C*	NM_170606.3:c.11669delA p.(Gln3890fs) #	P	Monoallelic *de novo*
73	0	F	Full	*MBD5*	NM_001378120.1 deletion of exon 2	LP	Monoallelic unknown
74	72	M	Full	*MYH2*	NM_017534.6:c.2116G>A p.(Glu706Lys)	P	Monallelic unknown
75	57	M	Partial	*MYH7*	NM_000257.4:c.1357C>T p.(Arg453Cys)	P	Monoallelic unknown
76	13	M	Full	*NARS1*	NM_004539.4:c.1600C>T p.(Arg534Ter)	P	Monoallelic *de novo*
77	19, 17	F, F	Full	*NPHP1*	Gene deletion Homozygous	P	Biallelic
78	3	M	Full	*NPHP1*	Gene deletion Homozygous	P	Biallelic
79	31	F	Partial	*OPTN*	Gene deletion	P	Monoallelic unknown
80	1	M	Full	*P4HTM*	NM_177939.3:c.659G>A p.(Trp220Ter) # NM_177939.3:c.569_579del p.(Gln190fs) #	P P	Biallelic
81	36	F	Full	*PDGFB*	Gene deletion	P	Monoallelic AD
82	11	F	Partial	*PHKB*	NM_000293.3:c.2109delT p.(Ser704fs) # NM_000293.3:c.2427+977C>T #	P P	Biallelic
83	3	M	Partial	*PKD2*	NM_000297.4:c.1390C>T p.(Arg464Ter)	P	Monoallelic AD
84	45	F	Full	*PMM2,* *AQP2*	NM_000303.3:c.442G>A p.(Asp148Asn) NM_000303.3:c.305A>G p.(Tyr102Cys) # NM_000486.6:c.707_720dupTGCTGAAGG GCCTG p.(Glu241fs) # NM_000486.6:c.34G>A p.(Ala12Thr) #	P LP LP LP	Biallelic Biallelic
85	2	F	Full	*POGZ*	NM_05100.4:c.2571-2delA#	P	Monoallelic *de novo*
86	54	F	Full	*POLR3A*	NM_007055.4:c.2119C>T p.(Gln707Ter) NM_007055.4:c.1909+22G>A	P P	Biallelic
87	2	M	Full	*PPP2R5D*	NM_006245.4:c.592G>A p.(Glu198Lys)	P	Monoallelic *de novo*
88	7	F	Full	*SAMD9*	NM_001193307.1: c.2053C>T p.(Arg685Ter)	LP	Monoallelic *de novo*
89	24	M	Full	*SCN2A*	NM_021007.3:c.4480C>A p.(Gln1494Lys) #	LP	Monoallelic *de novo*
90	49	F	Full	*SHOC2*	NM_007373.4:c.519G>A p.(Met173Ile)	LP	Monoallelic unknown
91	56	M	Partial	*SLC20A2*	NM_001257180.2:c.852delC p.(Ile1285fs) #	LP	Monoallelic unknown
92	25	M	Full	*SLC52A2*	NM_001363118.2:c.368T>C p.(Leu123Pro) NM_001363118.2:c.916G>A p.(Gly306Arg)	P P	Biallelic

93	11	F	Full	*SOS1*	NM_005633.3.3:c.1294T>C p.(Trp432Arg)	P	Monoallelic *de novo*
94	12	M	Partial	*TAB2*	NM_001292034.3: c.-90+1G>C	LP	Monoallelic *de novo*
95	7	F	Full	*TANGO2*	Deletion exons 3-9 Homozygous	P	Biallelic
96	58	M	Partial	*TTN*	NM_001267550.2:c.59926+1G>A	P	Monoallelic unknown
97	73	M	Full	*TTR*	NM_000371.3:c.407A>C p.(Tyr136Ser)	P	Monoallelic unknown
98	22	F	Full	*ZBTB20*	NM_001164342.2: c.1916G>A p.(Cys639Tyr) #	LP	Monoallelic unknown
Probable							
99	59	F	Full	*CTNNB1*	NM_001904.4:c.2315delA p.(Asn772fs) #	LP	Monoallelic Unknown
100	0	M	Full	*MYBPC3*	NM_000256.3:c.1357_158delCC p.(Pro453fs) NM_000256.3:c.1576G>C p.(Ala526Pro) #	P LP	Biallelic
101	22	F	Full	*PEX16*	NM_057174.2:c.851A>C p.(Tyr284Ser) # Homozygous	VUS	Biallelic
Possible							
102	60	F	Full	*MARS1*	NM_004990.4:c.493_495delGAG p.(Glu165del) #	LP	Monoallelic Unknown
103	52	M	Full	*MYH2*	NM_017534.6:c.2387C>A p.(Ala796Asp) #	VUS	Monoallelic unknown
104	13	M	Full	*MYO9A*	NM_006901.4:c.6796A>T p.(Asn2266Tyr) # NM_006901.4:c.1574A>T p.(Glu525Val) #	VUS VUS	Biallelic

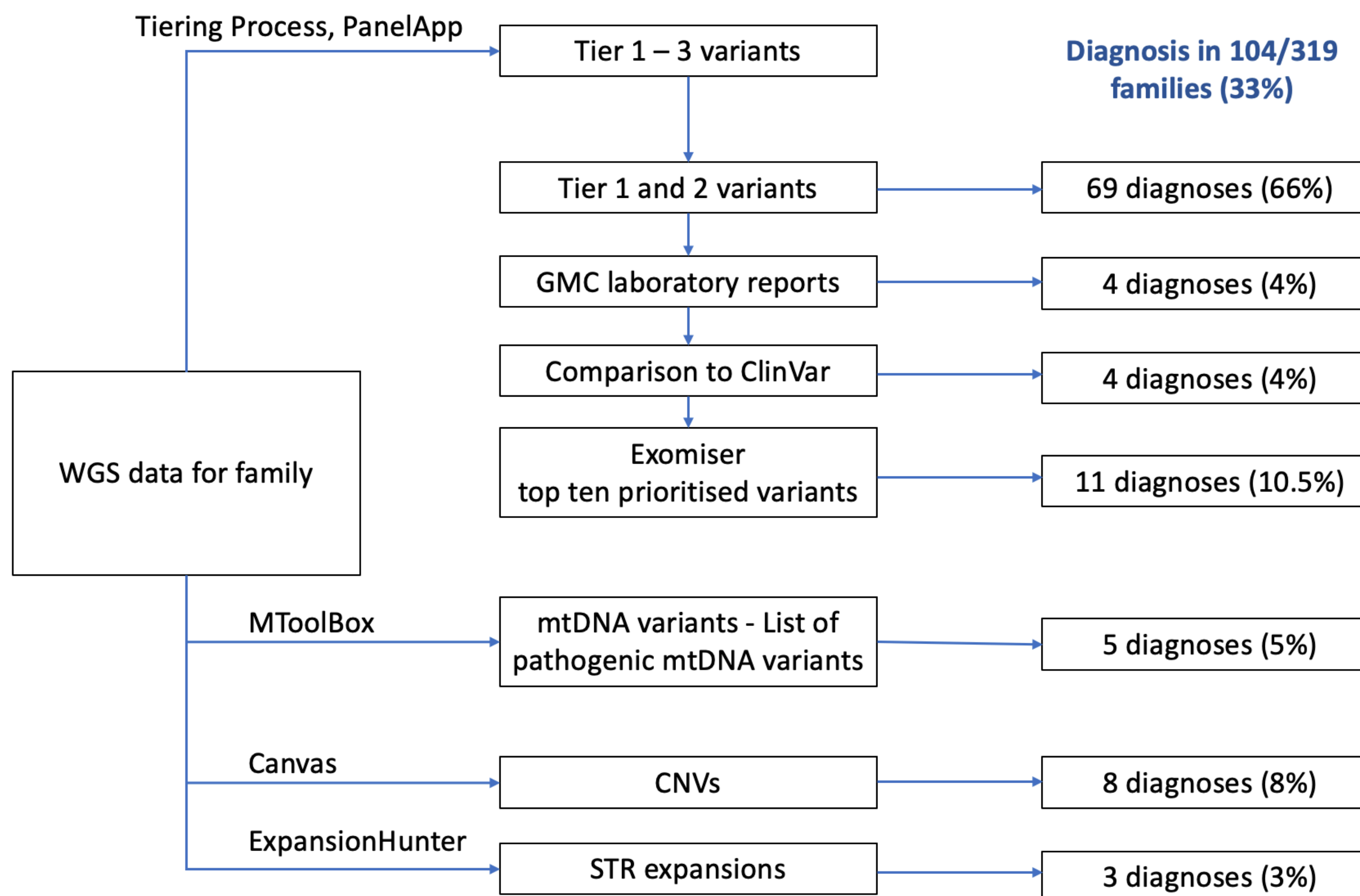

Figure 1. Overview of the analyses performed and sources of diagnoses. Variants in nuclear genes were analysed using the Genomics England tiering system. All tier 1 and tier 2 variants were reviewed and these provided 66% of the diagnoses. A further 19.5% of diagnoses were based on feedback from the Genomic Medicine Centre (GMC) labs, comparison to Clinvar pathogenic and likely pathogenic variants, and a review of the top ten Exomiser prioritised variants. mtDNA variants were analysed separately using an in-house pipeline and comparison against a list of 89 pathogenic variants, yielding and extra 5 diagnoses (6%). Copy number variants accounted for 8% of diagnoses, and short tandem repeat expansions for 3%.

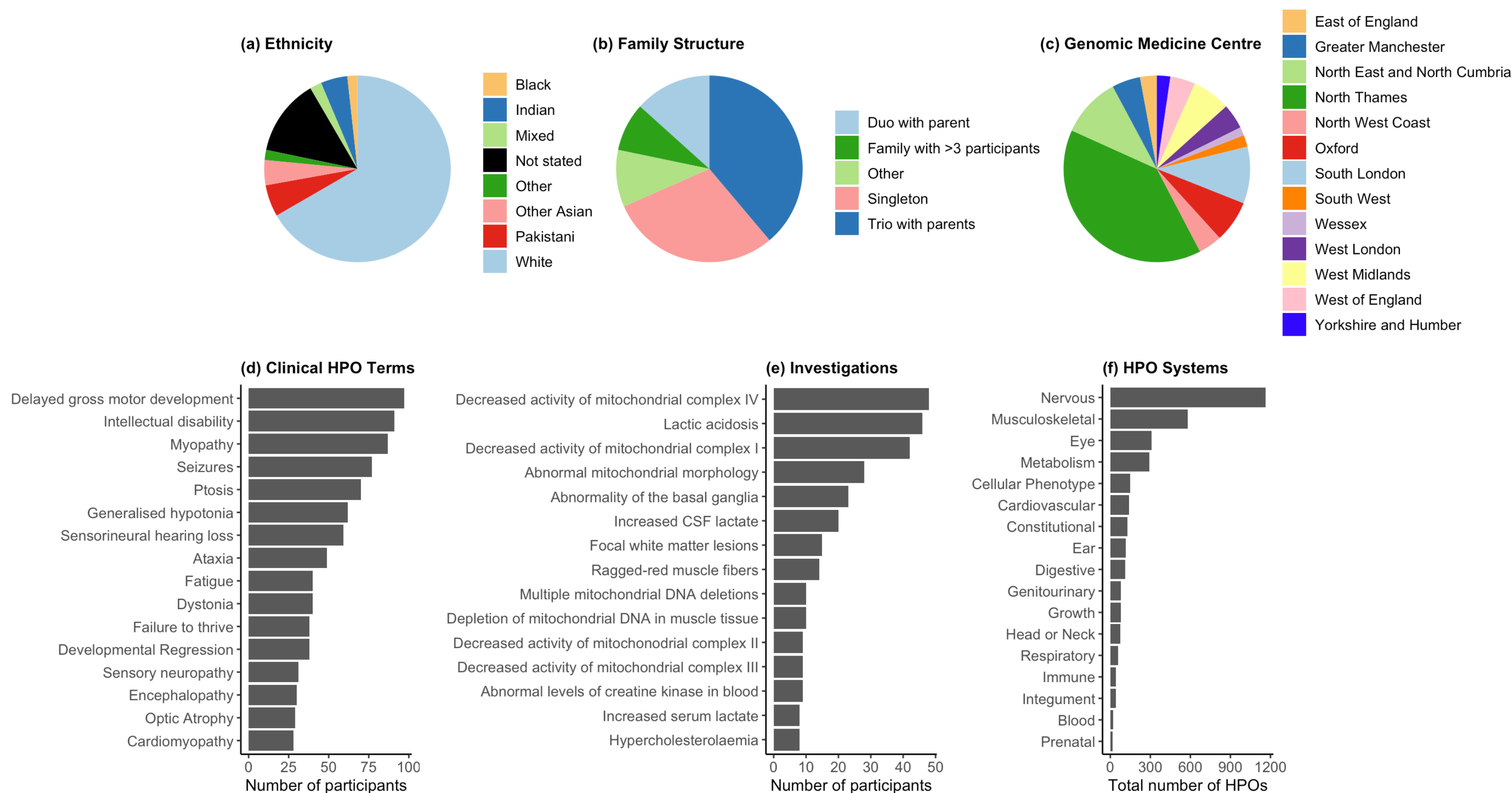

Figure 2. Demographics and Human Phenotype Ontology terms for the participants recruited. (a) Ethnicities recorded in the participants reflected the ethnicity of the overall population in England. (b) The most commonly recruited family structures were trios with both parents and singletons. (c) Genomic Medicine Centres. Participants were recruited from Genomic Medicine Centres across England. (d) Clinical HPO terms. The most commonly recorded clinical HPO terms including delayed gross motor development, intellectual disability and myopathy. (e) Investigation HPO terms. The most commonly recorded investigation results HPO terms including decreased activity of mitochondrial complex IV, lactic acidosis and decreased activity of mitochondrial complex I. (f) HPO Systems. The total number of HPO terms recorded for the 345 participants according to the ancestor HPO system. (Note that some HPO terms have more than one ancestor HPO system.)

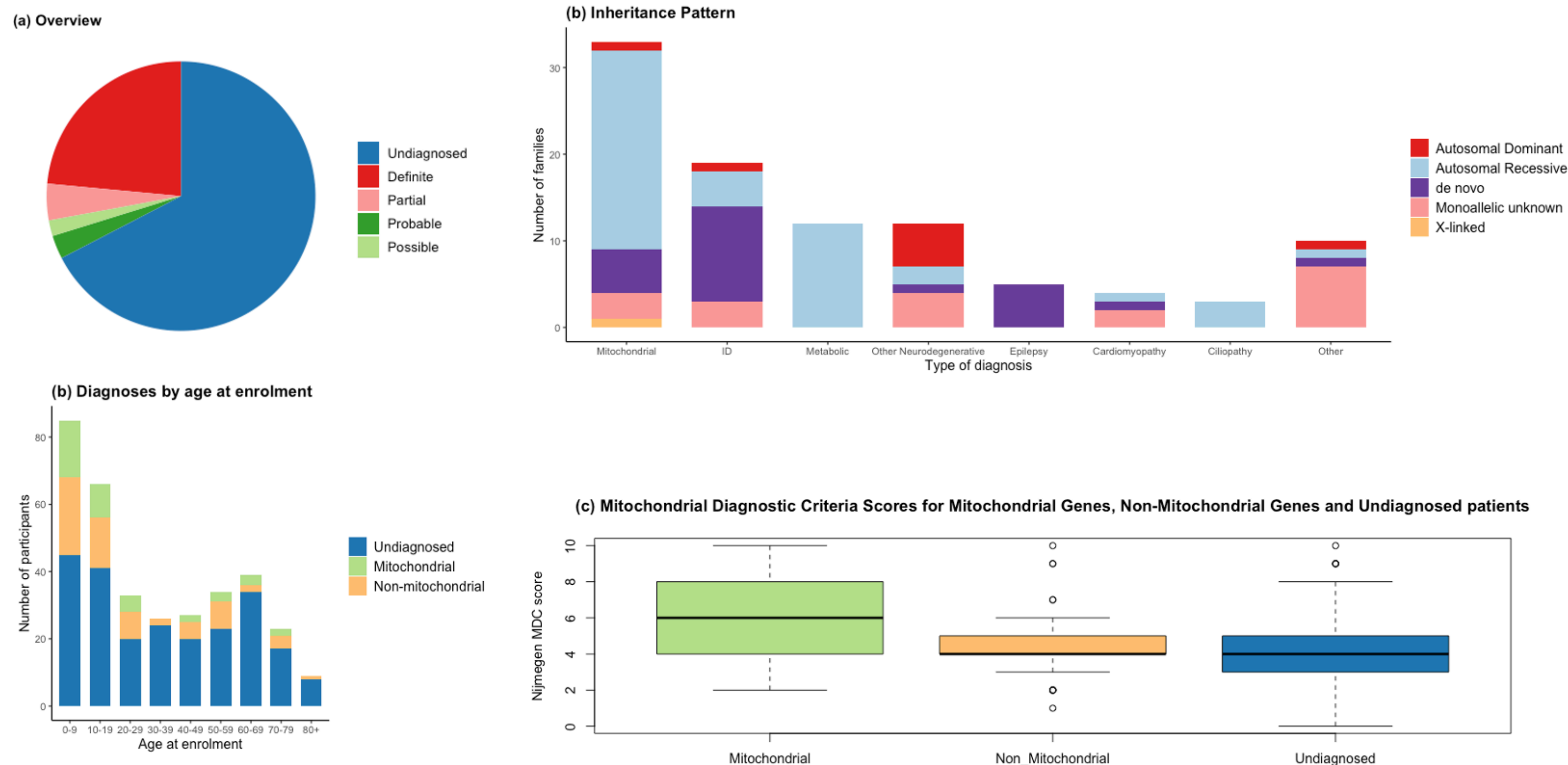

Figure 3. Clinical features in patients with different types of diagnoses. (a) Overview of proportion of patients diagnosed. (b) Age distribution of the participants at the time of enrolment and the type of diagnoses made. The diagnostic yield was higher in younger participants, but diagnoses were still being made in patients enrolled in their 70s and 80s. (c) Inheritance patterns in nuclear mitochondrial disorders and different types of non-mitochondrial disorders. Most families with nuclear mitochondrial disorders showed autosomal recessive inheritance. *De novo* dominant pathogenic variants were common in families with developmental disorders causing intellectual disability and in epileptic encephalopathies. (d) HPO-Modified Nijmegen Mitochondrial Diagnostic Scores in participants with confirmed genetic diagnoses of mitochondrial and non-mitochondrial disorders and in undiagnosed participants. Participants with 'probable', 'possible' and 'partial' diagnoses were excluded from this analysis. MDC scores were higher in patients with mitochondrial diagnoses than non-mitochondrial or undiagnosed ($p<0.05$)

Figure 4. Comparison of the Human Phenotype Ontology terms in patients with mitochondrial and non-mitochondrial diagnoses. HPO ancestor systems for the HPO terms recorded in participants with definite mitochondrial diagnoses and definite non-mitochondrial diagnoses (excluding partial diagnoses). Each column represents one participant (family number in brackets), and those with mitochondrial diagnoses (nuclear or mtDNA) are shown on the left side, and those with non-mitochondrial diagnoses are shown on the right side. Each row represents a different HPO ancestor system, listed in order of how frequently they were affected, with the nervous system at the top. The numbers indicate how many of the participant's HPO terms related to the HPO ancestor system (for example the nervous system or the musculoskeletal system). The colours go from green through to red as the number of terms related to the HPO ancestor system increases. Overall, the pattern of HPO systems involved is similar between the two groups, but the Metabolism was more commonly affected in patients with mitochondrial disorders (p=0.0002).

Supplementary Information

Supplementary Methods

I. Inclusion criteria for recruitment to the 100,000 genomes project under the category 'Suspected mitochondrial disorder'

II. Web based resources used for variant interpretation

Supplementary Results – Diagnostic yield by reference genome used and by recorded ethnicity

Supplementary Table 1 – Mitochondrial Genetic Testing in UK National Genomic Test Directory

Supplementary Table 2 – Proforma for recording phenotypic information

Supplementary Table 3. Modified Nijmegen Mitochondrial Diagnostic Score to use with Human Phenotype Ontology terms

Supplementary Table 4. Gene panels applied in the 345 participants with suspected mitochondrial disorders

Supplementary Table 5. Genomic coordinates for short tandem repeat regions for the 13 short tandem repeat genes analysed

Supplementary Table 6. Repeat-size threshold for premutation for the 13 short tandem repeat genes analysed

Supplementary Table 7. Further information about analysis of nuclear variants in the diagnosed families – population frequency, computational predictions, functional studies, tiering data and method of making the diagnosis

Supplementary Table 8. Potentially treatable disorders identified in this study

Supplementary Table 9. Sensitivity and specificity of HPO-modified Mitochondrial Disease Criteria Score to differentiate between mitochondrial and non-mitochondrial disorders