



Published in final edited form as:

Sci Transl Med. 2012 January 25; 4(118): 118ra10. doi:10.1126/scitranslmed.3003310.

Molecular Diagnosis of Infantile Mitochondrial Disease with Targeted Next-Generation Sequencing

Sarah E. Calvo^{1,2,3}, Alison G. Compton⁴, Steven G. Hershman^{1,2,3}, Sze Chern Lim^{4,5}, Daniel S. Lieber^{1,2,3}, Elena J. Tucker^{4,5}, Adrienne Laskowski⁴, Caterina Garone^{6,7}, Shangtao Liu¹, David B. Jaffe³, John Christodoulou^{8,9}, Janice M. Fletcher^{10,11}, Damien L Bruno^{4,12}, Jack Goldblatt¹³, Salvatore DiMauro⁶, David R. Thorburn^{4,5,12}, and Vamsi K. Mootha^{1,2,3}

¹Center for Human Genetic Research and Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street Sixth Floor, Boston, MA 02114, USA

²Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA

³Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02141, USA

⁴Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville Victoria 3052, Australia

⁵Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia

⁶Department of Neurology, Columbia University College of Physicians and Surgeons, 710 W 168th St, New York, NY 10032, USA

⁷Human Genetics Joint PhD program, Universities of Turin and Bologna, Italy

⁸Genetic Metabolic Disorders Research Unit, Children's Hospital at Westmead, 212 Hawkesbury Road Westmead, NSW 2145, Australia

⁹Discipline of Paediatrics & Child Health and Genetic Medicine, University of Sydney, Sydney, NSW, Australia

¹⁰Department of Genetics and Molecular Pathology, 72 King William Rd, North Adelaide 5006, Australia

¹¹Discipline of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, Australia

¹²Victorian Clinical Genetics Services, Royal Children's Hospital, Melbourne, Victoria, Australia

¹³Genetic Services of Western Australia, King Edward Memorial Hospital, University of Western Australia, Perth, Australia

Abstract

Address Correspondence to: Vamsi K. Mootha, M.D. (vamsi@hms.harvard.edu), Ph: 617 643 3059, Fax: 617 643 2335
Massachusetts General Hospital/185 Cambridge Street Sixth Floor Boston, MA 02114/USA. David R. Thorburn, Ph.D.
(david.thorburn@mcri.edu.au), Ph: 613 8341 6235, Fax: 613 8341 6212 Murdoch Childrens Research Institute/Royal Children's
Hospital/Flemington Road Parkville Victoria 3052/Australia.
Authors Calvo and Compton contributed equally to this article.

Author contributions: This study was conceived and designed by SEC, DRT, SD, and VKM. Characterization of patient phenotypes was performed by JC, JMF, and JG. Next-generation sequencing analysis was performed by SGH, DSL, and SEC with mtDNA assembly performed by DBJ. Variant validation and follow-up on patient samples were performed by AGC, SCL, SL, CG, EJT, DLB, and AL. Molecular haplotyping was performed by SGH. The manuscript was written by SEC, AGC, SGH, SCL, SD, DRT, and VKM. All aspects of the study were supervised by DRT and VKM.

Competing interests: none.

Advances in next-generation sequencing (NGS) promise to facilitate diagnosis of inherited disorders. While in research settings NGS has pinpointed causal alleles using segregation in large families, the key challenge for clinical diagnosis is application to single individuals. To explore its diagnostic utility, we performed targeted NGS in 42 unrelated infants with clinical and biochemical evidence of mitochondrial oxidative phosphorylation disease, who were refractory to traditional molecular diagnosis. These devastating mitochondrial disorders are characterized by phenotypic and genetic heterogeneity, with over 100 causal genes identified to date. We performed “MitoExome” sequencing of the mitochondrial DNA (mtDNA) and exons of ~1000 nuclear genes encoding mitochondrial proteins and prioritized rare mutations predicted to disrupt function. Since patients and controls harbored a comparable number of such heterozygous alleles, we could not prioritize dominant acting genes. However, patients showed a five-fold enrichment of genes with two such mutations that could underlie recessive disease. In total, 23/42 (55%) patients harbored such recessive genes or pathogenic mtDNA variants. Firm diagnoses were enabled in 10 patients (24%) who had mutations in genes previously linked to disease. 13 patients (31%) had mutations in nuclear genes never linked to disease. The pathogenicity of two such genes, *NDUFB3* and *AGK*, was supported by cDNA complementation and evidence from multiple patients, respectively. The results underscore the immediate potential and challenges of deploying NGS in clinical settings.

INTRODUCTION

Advances in next-generation sequencing (NGS) technology are facilitating the discovery of novel disease genes and promise to transform the routine clinical diagnosis of inherited disease. Since the first reports in 2009 (1–2), NGS has aided in the discovery of over 50 novel disease genes in research settings. Although sequence-based clinical diagnosis has long been available for individual genes, NGS could have a particularly profound impact on conditions characterized by genetic heterogeneity, where individual genes cannot be readily prioritized for traditional sequencing. The success of NGS for clinical utility, however, hinges on distinguishing the small number of pathogenic alleles from the thousands of DNA variants present in each person. When DNA is available from large affected families or from individuals with a closely related phenotype, variants have been successfully prioritized on the basis of segregation with disease. However it is not clear how useful NGS-based diagnostics will be in the more typical clinical scenario where an individual has no clear family history of disease and where the mode of inheritance is ambiguous.

Human oxidative phosphorylation (OXPHOS) disease represents a challenging class of disorders that could benefit tremendously if sequence-based tests can provide accurate diagnosis. Inherited defects in OXPHOS affect at least 1 in 5000 live births (3) and are characterized by a biochemical defect in the respiratory chain. OXPHOS disease is clinically heterogeneous as it can present early in infancy or in adulthood, can be severe or mild in its presentation, and typically impacts multiple organ systems. Clinical manifestations can include, but are not limited to, myopathy, lactic acidosis, seizures, ataxia, peripheral neuropathy, blindness, deafness, GI dysmotility, liver failure, and bone marrow dysfunction (4). OXPHOS disease can show maternal, recessive, dominant, or X-linked inheritance, although most cases are sporadic and have no obvious family history (5–6). Biochemical diagnosis typically requires invasive biopsies and even then can be inconclusive.

Genetically, OXPHOS disease can be caused by mutations in either the mitochondrial DNA (mtDNA) or in the nuclear genome. It is possible to routinely re-sequence the mtDNA, but lesions in this tiny genome likely account for no more than 15–30% of all childhood cases (5, 7). A recent review listed 77 nuclear disease genes, identified largely through pedigree analysis, with up to 10 novel OXPHOS disease genes described each year (8). The majority involve recessive, loss-of-function alleles, although 9 genes can harbor dominant acting

mutations (*POLG*, *POLG2*, *C10orf2*, *RRM2B*, *SLC25A4*, *OPA1*, *CYCS*, *MFN2*, *HSPD1*) (8). Individual sequence-based diagnostic tests are readily available for each of 50 genes (9) and more recently NGS gene panels (10–11); however, pleiotropy makes it difficult for clinicians to predict which specific genes may be mutated in a given case (12). Moreover, the known disease genes likely account for only a fraction of the total genetic burden of these disorders (13). Given that 94% of the known nuclear disease genes encode mitochondrial-localized proteins (Supplemental Materials), the full set of ~1000 genes encoding mitochondrial proteins represent strong candidates for OXPHOS disease (14).

The extensive locus heterogeneity, allelic heterogeneity, and pleiotropy of OXPHOS disease, combined with the availability of a focused set of ~1000 high-confidence candidate genes encoding the known mitochondrial proteome, make OXPHOS disorders an ideal application area for sequence-based diagnostics. Three recent studies reported the technical feasibility of sequencing several hundred of these genes (13, 15–16) and NGS diagnostic tests are newly available for panels of disease-related genes (10–11). We performed targeted NGS to capture and sequence the entire mtDNA and the exons of the ~1000 genes encoding the mitochondrial proteome (the “MitoExome”). After benchmarking the performance on several DNA samples, we applied MitoExome sequencing to 42 unrelated patients with clinical and biochemical evidence of infantile OXPHOS disease in whom a molecular diagnosis was not previously available (Table 1).

RESULTS

MitoExome sequencing and validation

We targeted for sequencing the entire mitochondrial genome and all coding exons of 1034 nuclear genes encoding mitochondrial proteins based on the MitoCarta inventory (14) (Table S1). These regions were enriched using hybrid selection and then sequenced on an Illumina GAII. For each individual, we generated ~2 billion bases of sequence which yielded an average coverage of ~143X at each targeted nuclear base and ~25,457X at each mtDNA base (Table 2). On average, 96% of targeted bases were covered and 87% of targeted bases exceeded the 15X coverage threshold required for confident analysis, defined as 99% power to detect a variant (Table 2).

The accuracy and reproducibility of MitoExome variant detection was assessed using DNA obtained from the parents and daughter of a family with independent sequence data available through HapMap (17). Detection of nuclear single-nucleotide variants (SNVs) was 96% sensitive and 100% specific, based on 444 and 823 sites genotyped using independent technology. Specifically, there was 94% genotype concordance at heterozygous sites, 98% concordance at homozygous sites, and 99.3% concordance at sites with at least 15 reads. Similarly, nuclear SNV detection was 90% sensitive based on 856 variant sites detected by whole genome sequencing of these samples (98.4% concordance at sites with at least 15 reads). Two technical replicates of MitoExome sequencing showed 94% concordance at variant sites, consistent with the reported sensitivity. mtDNA SNV detection was 97% sensitive and 100% specific based on five patient samples for which independent data were available.

MitoExome sequencing was next applied to whole-genome amplified DNA obtained from 42 unrelated patients that fulfilled the clinical and biochemical criteria for OXPHOS disease prior to two years of age (Table 1). These patients included 11 from consanguineous families and one additional patient with an affected family member, although none of these families were large enough to identify a single locus by a traditional mapping approach. Each sequenced individual harbored approximately 758 nuclear SNVs, 15 small nuclear

insertions or deletions (indels), and 36 mtDNA variants compared to the reference human genome (Table 2 and Table S2).

Variant prioritization

The key challenge for interpreting sequence data for molecular diagnosis is to distinguish causal alleles from the thousands of variants present in each individual. Since we focused on rare, devastating clinical phenotypes, we hypothesized that the underlying causal variants would be uncommon in the general population, disruptive to protein function, and would either be inherited in an autosomal recessive fashion from unaffected carrier parents or be *de novo*, dominant acting mutations. We therefore highlighted all nuclear variants that were rare and predicted to be protein-modifying (Methods) as well as all variants previously associated with disease in the Human Gene Mutation Database (HGMD) (Table 2).

If rare, protein-modifying variants are enriched for causal alleles, we would expect increased prevalence in patients compared to healthy individuals. Therefore, we analyzed the MitoExome regions in 371 healthy individuals of European ancestry with available whole-exome sequence data. To avoid overestimating patient variants, this comparison was limited to the 31 patients from non-consanguineous families and to MitoExome regions well covered in all individuals (76% of all autosomal coding exons, see Methods). Genes containing only a single rare, protein-modifying variant showed only a 1.1 fold enrichment in patients compared to controls ($P=0.03$), and this subtle enrichment is likely caused by greater ethnic diversity in patients (Fig. 1A, Fig. S1). However, we observed a striking 5.5 fold enrichment of genes containing two such prioritized alleles in patients compared to controls ($P=3e-11$) (Fig. 1A and methods). There was no enrichment observed in the 347 auxiliary genes sequenced (that had only weak evidence of mitochondrial localization), suggesting that the enrichment is not due to population stratification (Fig. 1A). 45% of all non-consanguineous cases contained mitochondrial genes with two prioritized variants, compared to only 9% of controls (4.8 fold enrichment, $P=7e-5$) (Fig. 1B). Thus, the background rate of rare, predicted deleterious heterozygous variants makes it difficult to spotlight dominant acting, heterozygous causal alleles from individual DNA samples. However, the excess of genes harboring two prioritized variants provides a signal to distinguish recessive variants. Therefore, we prioritized only variants consistent with a recessive mode of pathogenesis, and denoted “prioritized genes” as those containing two such alleles.

We additionally prioritized two classes of mtDNA variants: variants annotated as pathogenic in MITOMAP (18) at greater than 10% heteroplasmy (as estimated by sequence reads) (19), and structural deletions and rearrangements as detected by *de novo* mtDNA assembly. Detection of heteroplasmic variants via deep sequencing was previously validated based on mixing experiments (19).

For the 9 genes known to harbor dominant mutations in OXPHOS disease, we carefully evaluated all rare, protein-modifying variants and all variants that were reported as pathogenic in HGMD. As described in the Supplemental Materials, none of the six such variants appeared to be dominant mutations associated with the observed patient phenotypes.

Enrichment of prioritized genes in patients

We note that the five-fold enrichment of prioritized genes in cases compared to controls is likely to be a conservative estimate, especially as databases of allele frequencies improve. Indeed, many prioritized variants showed greater than 0.005 allele frequency in the 371 control samples. When we excluded variants exceeding 0.005 allele frequency from cases

and 371 controls (similar to exclusion criteria based on 1000 genomes and dbSNP thresholds), we observed a 6.2 fold enrichment in non-consanguineous cases versus controls ($P=5e-11$) (Fig. 1C). This corresponded to a striking 23.9 fold enrichment ($P=2e-8$) specifically within the 77 established disease-related genes (8) and 5.2 fold enrichment ($P=2e-7$) within the remaining 957 mitochondrial genes (Fig. 1C). Thus, the 42 infantile patients are enriched for prioritized variants in both known pathogenic genes as well as in candidate genes not previously linked to disease.

Prioritized variants in 42 patients

Of the 42 patients with severe OXPHOS disorders, 23 (55%) harbored at least one “prioritized gene” or mtDNA variant: 10 (24%) of these were in known, nuclear OXPHOS disease-related genes, 12 (29%) were in nuclear genes not previously linked to disease, and 1 (2%) patient harbored a large mtDNA deletion (Fig. 2A). The remaining 19 (45%) patients lacked prioritized genes or mtDNA variants. Most prioritized genes harbored missense changes at amino acid residues highly conserved across evolution (Fig. 2B). Since there was a ~5- fold enrichment in patients with prioritized genes compared to controls, we can estimate that approximately 18 of the 42 patients (43%) will contain causal variants in prioritized genes. Determining which of the prioritized DNA variants are actually causal requires additional experimental evidence.

Molecular diagnoses in the mtDNA

Through de novo assembly of the mtDNA (Methods), we discovered a 7.2kb deletion in the mitochondrial genome in patient P33 (Fig. 3A). This patient presented with progressive developmental delay and regression, failure to thrive, microcephaly, lactic acidosis and death following a sudden metabolic and neurological decompensation. Muscle biopsy showed ragged red fibers and reduced COX staining in affected fibers. The deletion (m. 8407_15658 del7252 with imperfect tandem repeats at both ends) has been previously reported (20) and was present in 75% of sequence reads at this locus. The deletion was confirmed by quantitative PCR to be at 94% heteroplasmy in genomic DNA from affected muscle tissue. Due to mtDNA proliferation (288% compared to control) the actual amount of non-deleted mtDNA present in muscle of this patient was 18% of mean control level. The deletion was confirmed by long-range PCR (Fig. 3C) and the breakpoints were mapped by Sanger sequencing (Fig. 3B). Thus MitoExome sequencing enables simultaneous detection of mtDNA point mutations and deletions.

Molecular diagnoses in previously established nuclear disease genes

Ten patients harbored recessive mutations in eight genes previously shown to cause OXPHOS disease: *ACAD9* (21–22), *POLG* (23–24), *BCS1L* (25), *COX6B1* (26), *GFMI* (27), *TSM* (28), *AARS2* (29) and *TYMP* (30) (Fig. 2A). All such variants were confirmed through Sanger sequencing. We established compound heterozygosity via sequencing familial DNA, cDNA, cloning or molecular haplotyping (31) since short sequence reads do not typically provide the ability to phase variants.

With additional experiments, we established a firm genetic diagnosis for nine of these patients (Table 3). These methods included the demonstration that the detected alleles segregated with disease in a family, that they were rare in controls, and/or that they caused reduced cellular abundance of full-length mRNA transcripts, protein products, or OXPHOS subunits and assembly factors (Table 3 and Methods). The enzyme *ACAD9* (21–22), recently linked to complex I deficiency (MIM:252010), was mutated in a male infant with isolated complex I deficiency and lethal infantile mitochondrial disease (Fig. S2). A female infant with encephalopathy and clear complex I deficiency with less marked deficiency of complexes III and IV, harbored previously reported causal mutations in *POLG* (23–24),

which encodes the mtDNA polymerase and is the most commonly mutated nuclear gene underlying OXPHOS disease (7) (Fig. S3). The complex III assembly factor gene *BCSIL* (25), which is the most commonly mutated gene in complex III deficiency (MIM:124000) (7), was mutated in an infant girl with complex III deficiency (Fig. S4). Complex IV subunit COX6B1 (26) was mutated in a boy with neonatal onset of mitochondrial encephalopathy, metabolic acidosis, and complex IV deficiency (MIM:220110) (Fig. S5). The protein GFM1 (MIM:609060) (27), which is required for translational elongation of mtDNA-encoded proteins, was mutated in two unrelated patients with mitochondrial encephalopathy presenting within the first year of life (Fig. S6). A second elongation factor, TSFM (28), was mutated in one patient with combined OXPHOS deficiencies (MIM:610505) and in an unrelated patient with cardioencephalomyopathy (Fig. S7). A stillborn fetus with mitochondrial myopathy harbored one novel and one previously reported causal variant in the mitochondrial translational protein AARS2, which was recently shown to underlie fatal infantile cardiomyopathy (MIM:614096) (29) (Fig. S8).

Although in all nine cases the MitoExome genetic diagnosis was consistent with the patient phenotype, the clinical presentation alone was never specific enough to suggest the underlying gene as the most likely candidate for an available gene-based sequence diagnostic test, except perhaps for *BCSIL*. In the remaining tenth case, the compound heterozygous *TYMP* mutations in patient P12 do not fit with the clinical presentation, complex III defect, or familial consanguinity and it is possible that homozygous mutations in candidate genes *MTCH1* or *C6orf125* may prove to be causal.

Mutations in candidate genes

Approximately one-third of patients harbored prioritized DNA sequence variants in genes never before linked to OXPHOS disease. In total, 15 such candidate disease genes were identified in 13 patients (Fig. 2A and Table S2) after excluding mutations that were not compound heterozygous (*MRPL37*) and mutations that did not segregate with disease in available familial DNA (*COX10*, *FAM82A1*). Our collection of candidate genes is likely greatly enriched for novel, *bona fide* mitochondrial disease genes, but additional proof of pathogenicity is required. One method is to identify independent mutations of the same gene in unrelated individuals presenting with exquisitely similar phenotypes, as was recently shown for Miller (32), Kabuki (33), and Bartter (1) syndromes. A second definitive method is to complement a cellular defect present within patient cells by introducing a wild-type copy of the locus, either through cybrid fusions for mtDNA defects (34) or through overexpression of a nuclear gene (13). We obtained support of pathogenicity for two of the novel disease genes, as described below.

AGK mutations in two patients with myopathic mtDNA depletion

One novel candidate gene – acylglycerol kinase (*AGK*) – had recessive mutations in two unrelated patients. This protein (also known as multisubstrate lipid kinase, MuLK) phosphorylates monoacylglycerols and diacylglycerols to lysophosphatidic acid and phosphatidic acid respectively (35). The two unrelated patients each presented with severe myopathy, combined complex I, III, IV deficiency, bilateral cataracts, and severe mtDNA depletion in skeletal muscle (4% and 13% of normal). Detailed clinical histories are available in the Supplementary Materials. Currently, only two genes are known to underlie the myopathic form of mtDNA depletion (36) (*TK2*/MIM:612075 (37) and *RRM2B*/MIM:609560 (36)), and coding mutations in both these genes were excluded. Our two unrelated patients harbored 3 severe mutations in *AGK* (Fig. 4 and Fig. S9): patient P42 harbored a homozygous splice variant that caused a shortened transcript with a premature termination codon (c.1131+1G>T, p.S350EfsX19), and patient P41 harbored a compound heterozygous nonsense variant (p.Y390X) and splice variant that caused a shortened transcript with a

premature termination codon (c.297+2T>C, p.K75QfsX12), inherited respectively from her mother and father. All mRNAs were stable and not affected by nonsense mediated decay. All three mutations would cause deletion of a C-terminal region that is highly conserved across evolution and thus likely to be important for proper protein function, especially as it contains the conserved C5 domain from related sphingosine type kinases (35) (Fig. 4). We screened for coding *AGK* mutations in eight additional individuals with myopathic mtDNA depletion, but detected no further mutations.

Our identification of independent, likely deleterious mutations in *AGK* in patients with myopathic mtDNA depletion suggests that *AGK* is a novel locus underlying this syndrome and may offer a link between lipid metabolism and the control of mtDNA copy number.

Establishing the pathogenicity of *NDUFB3*

The complex I subunit *NDUFB3* was mutated in patient P3, a girl from a non-consanguineous family with complex I deficiency and lethal infantile mitochondrial disease. She harbored an apparently homozygous c.64T>C (p.W22R) mutation. The p.W22 residue is highly conserved (conserved in 34/34 vertebrates) and lies adjacent to an important lysine residue that undergoes N6-acetylation within the N-terminus (Fig. S10). Sanger sequencing identified the proband's mother as a heterozygous carrier for this mutation. Paternal DNA was unavailable for testing. To assess possible DNA deletions or uniparental isodisomy, we analyzed proband and maternal DNA using SNP arrays. No large deletions were evident and isodisomy of chromosome 2 was excluded based on the SNPduo identity by state algorithm (38). However, the cnvpartition v3.1.6 algorithm (Illumina) detected a 1.3Mb long contiguous stretch of homozygosity (LCSH) encompassing *NDUFB3* in the proband's DNA (Fig. S10). The maternal sample did not show statistically significant evidence of LCSH. Analysis of the patient's fibroblast cDNA (with and without cycloheximide) showed a single stable transcript of expected size indicating that the mutation did not alter mRNA splicing nor the presence of any small indels below the detection range of the cytogenetic arrays.

We next performed a complementation experiment to assess whether the introduction of wild-type cDNA into subject fibroblasts rescued the defect in complex I activity. Fibroblasts from this individual showed a strong complex I defect, with ~15% residual complex I activity when assayed by spectrophotometric enzyme assay and <2% residual complex I activity when assayed by dipstick enzyme assay. Using a lentiviral expression system, we transduced subject fibroblasts with wild-type cDNA. Expression of wild-type *NDUFB3* rescued complex I activity and protein levels in fibroblasts from subject P3 but not from a previously diagnosed complex I patient with confirmed mutations in *C8orf38*(14) and vice versa (Fig. 5), establishing *NDUFB3* as the causal gene in this case. This is the fifteenth nuclear encoded complex I subunit gene in which mutations have been shown to cause complex I deficiency in humans.

DISCUSSION

We performed MitoExome sequencing of the entire mtDNA and exons of 1034 nuclear genes encoding mitochondrial proteins, including all 77 nuclear OXPHOS disease-related genes that were reviewed recently (8). We applied MitoExome sequencing to 42 unrelated patients with a spectrum of early-onset OXPHOS disorders who lacked molecular diagnoses. Of note, we did not know what percent of unsolved OXPHOS cases would be expected to be due to mutations in established loci, since to our knowledge no study had sequenced even the 77 known disease-related genes in a collection of cases. We found that only 24% of unsolved cases were due to mutations in the known disease loci (including 1 mtDNA deletion and 9 nuclear gene defects shown in Fig. 2), highlighting the locus heterogeneity of OXPHOS disease. A further 31% of patients harbored rare, protein-

modifying, recessive variants in candidate genes not previously linked to OXPHOS disease. Since such variants exhibit five-fold enrichment in cases over background, they are likely enriched for truly causal alleles. We performed complementation experiments to firmly establish pathogenicity for *NDUFB3* in complex I deficiency, and based on independent mutations in P41 and P42, we suggest a novel link between the gene *AGK* and myopathic mtDNA depletion through an unknown mechanism.

Recessive mutations in 13 additional genes, never previously linked to mitochondrial OXPHOS disease, were identified. Formal proof of pathogenicity will be established by cDNA complementation experiments in cellular models or patient fibroblasts (13), when such cells exhibit an OXPHOS defect, or by detecting independent mutations in individuals with similar phenotype. We estimate that six of these 13 genes will prove to be *bona fide* disease genes. These candidates are particularly exciting since until recently such discoveries generally required familial forms of disease to narrow down the search region. Some of the candidate genes have established roles in OXPHOS biology consistent with the enzyme defect found in the relevant patient (e.g. *UQCRI0*, *LYRM4*, *EARS2*), while the majority of candidates encode poorly characterized proteins that have never before been linked to OXPHOS and will reveal fundamentally novel biochemical insights (e.g. *C1orf31*, *C6orf125*, *AKR1B15*).

Importantly, this pilot study showed that approximately half of the 42 sequenced patients lacked “smoking gun” prioritized variants. We can envision at least five possible explanations. First, we may have missed the causal variants due to a technical lack of sensitivity. Although we detected over 90% of SNVs present in the MitoExome, there is an unknown sensitivity for indels and exon deletions due to lack of training data. Second, the causal variant may have been located in a gene that we did not target with MitoExome sequencing. While possible, this explanation seems less likely since unbiased linkage and homozygosity mapping strategies to date have found that 94% of causal genes encode mitochondrial proteins. Third, the causal variant may be located in a non-targeted intronic or regulatory region. While these are very likely to exist, no robust methods are available yet for interpreting such variants. Fourth, and perhaps most likely, our stringent filters may have rejected the truly causal variant. For example, *de novo* dominant alleles (39), acting through gain-of-function or haploinsufficiency, were not prioritized since without parental DNA these alleles are difficult to distinguish from the high heterozygote burden of apparently deleterious alleles in healthy individuals. Similarly, it is difficult to distinguish benign from pathogenic mtDNA variants (40). By applying MitoExome sequencing to parental DNA, such *de novo* alleles may be identified in the future (41). Finally, and potentially most interesting, our assumptions on the genetic architecture of OXPHOS disease may be inaccurate. Nearly all of the nuclear genes discovered to date correspond to Mendelian syndromes, characterized by strong, highly penetrant alleles. It is possible that many of the sporadic cases of OXPHOS disease in our cohort are due to the combined action of multiple “weak” alleles, each with incomplete penetrance.

Our study underscores the need for clinical standards for interpretation of genetic variants to evolve as NGS is applied more widely. First, current guidelines for interpreting clinical genetic tests, such as the American College of Medical Genetics (ACMG) guidelines (42), are deliberately restricted to gene loci with established roles in disease. Second, it is notable that many variants purported to be causal for disease may, in some cases, be benign polymorphisms (43). For example, we detected 44 alleles previously reported as pathogenic, but only six actually appear to explain the phenotype while the remainder were heterozygous or present in patients with unrelated phenotypes (Supplementary Methods).

We anticipate that the success rate for establishing molecular diagnosis in unselected cases of infantile OXPHOS disease using NGS will be higher than that observed in this study. Indeed, our 42 patients were refractory to molecular diagnosis using traditional methods as most patients had been screened for common mutations in mtDNA or in relevant genes suggested by phenotype (e.g. *POLG* and *SURF1*) and were not from informative pedigrees. Analysis of a representative cohort of 291 unrelated infantile patients with “definite” OXPHOS disease from the Murdoch Childrens Research Institute, of which 124 cases had previous molecular diagnosis, suggests that MitoExome sequencing could enable diagnosis in roughly 47% of all infantile patients and prioritize candidates in a further 20% (see Supplementary Materials).

In the coming years, we anticipate that three advances will greatly aid interpretation of DNA variation for routine clinical diagnosis. First, NGS studies of Mendelian families will rapidly expand the set of validated OXPHOS disease-related genes from 77 to perhaps 200 nuclear loci. Second, NGS studies of ethnically diverse, healthy individuals will generate databases of allele frequencies that are necessary for filtering out common variants unlikely to cause severe disease, as was recently shown for interpretation of carrier screening (43) and has long been appreciated in the interpretation of mtDNA variation (45). Third, while costs currently support MitoExome sequencing (roughly one-third the cost of exome sequencing), future cost reductions will enable sequencing of the entire exome or genome, as well as simultaneous analysis of parental DNA in order to phase compound heterozygous variants and detect *de novo* mutations.

The subset of patients for whom a clear molecular etiology was possible spotlights the immediate promise of NGS in clinical diagnosis. For these patients, MitoExome sequencing, requiring a single blood sample, can accelerate clinical diagnosis and enable genetic counseling where appropriate. Genetic diagnosis will also enable rational subclassing of disease, which may help predict clinical course and severity, and lead to patient grouping for targeted therapeutics. However, we anticipate that even with improvements afforded by broader catalogs of genomic variation, interpretation of most sequence variants will be most useful when integrated with the broader clinical and biochemical presentation.

MATERIALS AND METHODS

Patients

We selected 42 patients with “definite” OXPHOS disease based on clinical and biochemical criteria from established diagnostic algorithms (46) who lacked a molecular diagnosis (Table 1). Of these, 38 patients were selected from the Murdoch Childrens Research Institute laboratory and 4 from the Columbia University H. Houston Merritt Clinical Research Center. All patients had severe biochemical OXPHOS defects and were selected to represent five main categories of OXPHOS defects: complex I deficiency (10 patients), complex III deficiency (6 patients), complex IV deficiency (9 patients), combined OXPHOS deficiencies (14 patients), and mtDNA depletion (3 patients) (Table 1). The severity of enzyme defects in Table I were classified as marked (<25%), moderate (26–40%), or equivocal (41–60%), corresponding to residual activities of normal control mean relative to citrate synthase or complex II. Inclusion criteria were infantile age of onset, defined for this study as under 2 years of age. Exclusion criteria were maternal family history or presence of a known nuclear or mtDNA mutation. In 16 cases, some candidate genes had previously been sequenced and excluded. The study protocols were approved by the ethics committee at the Royal Children’s Hospital, Melbourne, Columbia University, and the Massachusetts Institute of Technology. All samples were obtained as part of diagnostic investigations and families provided informed consent.

MitoExome target selection

The 4.1Mb of targeted DNA included the 16.6Kb mtDNA and all coding and untranslated exons of 1381 nuclear genes, including 1013 mitochondrial genes from the MitoCarta database (14), 21 genes with recent strong evidence of mitochondrial association, and 347 additional genes with weak evidence of mitochondrial association, all listed in Table S1. Gene transcripts from RefSeq (47) and UCSC known gene (48) collections were downloaded from the UCSC genome browser (49) assembly hg18 (February 2009). All analyses presented here, excepting one described in Fig. 1A, were restricted to the mtDNA and coding exons and splice sites of 1034 genes with confident evidence of mitochondrial association (1.4Mb), since no significant results were found in the untranslated regions (2.3 Mb) or in the coding regions of 347 auxiliary genes sequenced (0.4Mb).

Hybrid selection and Illumina sequencing

An in-solution hybridization capture method (50) was used to isolate targeted DNA, which was sequenced on the Illumina GAI platform (51) using paired 76bp reads. Single 120bp baits were synthesized (Agilent) for each target region, or tiled across targets that exceeded 120bp. A total of 42,923 baits were synthesized to cover 13,803 target regions. Each subject sample was whole-genome amplified using a QIAGEN REPLI-g Kit with 100 ng input DNA. HapMap and the National Institutes of Mental Health (NIMH) control samples were not whole-genome amplified. All samples were sequenced at the Broad Institute Sequencing Platform. Genomic DNA was sheared, ligated to Illumina sequencing adapters, and selected for lengths between 200–350bp. This “pond” of DNA was hybridized with an excess of Agilent baits in solution. The “catch” was pulled down by magnetic beads coated with streptavidin, then eluted (50, 52). Each patient or HapMap sample was sequenced in one lane of an Illumina GAI instrument. NIMH control samples were subjected to whole-exome hybrid selection using the same protocol and were sequenced in 2 or 3 lanes of an Illumina GAI instrument, using paired 76bp reads.

Sequence alignment, variant detection, and annotation

Illumina reads were aligned to the GRCh37 reference human genome assembly using the BWA (53) algorithm in SAMtools (2) within the Picard analysis pipeline (<http://picard.sourceforge.net/>). Illumina quality scores were recalibrated, realigned to GRCh37, duplicates were removed, and the alignment was modified to account for indels using the GATK software package (54) version v1.0.5159. Single nucleotide variants (SNVs) were detected and genotyped using the GATK UnifiedGenotyper in single-sample mode (with parameters `-im ALL -mbq 20 -mmq 20 -mm42 3 -deletions 0.05`). Variants were filtered using GATK VariantFiltration module (with filters “QUAL<50.0 & QD<5.0 & HRun>10 & DP<4” and parameters `-cluster 3 -window 10`). Indels were detected using GATK IndelGenotyperV2 (with parameters `-im ALL`) and filtered with a custom python module that removed sites with a `max_cons_av_ ≥ 1.9` (maximum average # of mismatches across reads supporting the indel) or `max_cons_nqs_av_mm ≥ 0.2` (maximum average mismatch rate in the 5bp NQS window around the indel, across indel-supporting reads). Indels were genotyped as homozygous (allele balance >0.85) or heterozygous (allele balance <0.85). Variants were annotated using the GATK GenomicAnnotator with data obtained from RefSeq transcripts(47), UCSC known gene transcripts(48), known disease-associated variants obtained from the HGMD(55) professional version 2010.3, allele frequency from dbSNP(56) version 131, the 1000 genomes project (57), and amino acid conservation scores calculated from 44-vertebrate species alignments downloaded from the UCSC genome browser (48).

847 variants likely to be alignment artifacts were excluded by filtering out variants supported by greater than two Illumina read-pairs that aligned to the genome in an aberrant

orientation. We observed stacks of aberrantly oriented reads at the boundaries of genomic regions with extremely high coverage, and sequence mismatches at the boundaries to these regions suggested that the underlying DNA had been circularized prior to amplification and fragmentation, perhaps during whole-genome amplification.

mtDNA sequence analysis

Analysis of mtDNA variants is challenging given the large copy number of mitochondrial genomes per cell, as well as the nuclear genome sequences of mitochondrial origin (NUMTs) (58). Subject DNA samples typically contain >100X copies of mtDNA molecules compared to nuclear DNA molecules. To avoid alignment artifacts caused by mtDNA reads being improperly assigned to NUMTs, all Illumina reads were separately aligned to the mtDNA revised Cambridge Reference Sequence (NC_012920) using BWA (53). Unlike the nuclear genome, read-pairs with identical alignment positions were retained (resulting in ~25,000X coverage) rather than excluded (resulting in ~1000X coverage) since the observed number of such duplicate read-pairs matched the expected number based on simulation of high mtDNA copy number. Variants were detected in each BAM file using the GATK UnifiedGenotyper version v1.0.2986 in pooled-sample mode in order to capture heteroplasmic variants (with parameters -mbq 2 -mmq 50 -mm42 3 -deletions 0.05 -exp -gm POOLED --poolSize 50 -confidence 0 -pl SOLEXA). Variants supported by > 10% of aligned reads were identified, since it is difficult to resolve low-heteroplasmy variants from variants present in NUMTs (19). mtDNA variants were annotated using data from MITOMAP (18) and allele frequency was obtained from mtDB (45).

mtDNA deletions or rearrangements were detected through *de novo* mtDNA assembly using a modified ALLPATHS assembly approach (59). Reads aligned to the mtDNA were randomly downsampled to 1000X coverage for computational tractability. Errors in reads were corrected using the ALLPATHS-LG (60) error correction algorithm. Next a k-mer graph (61) was constructed (k=40), and filtered using four steps (Supplementary Methods). Each resulting mtDNA assembly (in a circular, directed graph representation) was manually inspected to identify alterations supported by >10% of aligned reads.

Sensitivity, specificity and reproducibility

Sensitivity and specificity of detecting nuclear SNVs in HapMap individuals (NA12878, NA12891, NA12892) was estimated using independent genotype training data obtained from HapMap Phase 3 release 2 (17) as well as in Illumina whole genome sequence data from the 1000 genomes project (57) pilot 1. Genotype concordance was assessed at 444 targeted sites that differed between HapMap Phase 3 and the reference genome, as well as 856 variant sites based on the 1000 genomes project (57) pilot 1. Reproducibility was measured by technical duplicates of sample NA12878, from which one aliquot of lymphoblastoid cell line DNA was used to create two sequence libraries which were separately hybridized and sequenced, with genotype concordance observed in 776 of 826 sites detected as variant in either replicate. MtDNA sensitivity and specificity were calculated at 93 non-reference and 19,368 reference sites for which independent sequence data were available in five patient samples (Newcastle University).

Variant prioritization

Nuclear SNVs and indels that passed quality control metrics were prioritized according to three criteria: i) variants that were rare in healthy individuals, with SNV allele frequency below 0.005 within public databases (dbSNP (56) version 131 and the 1000 genomes project (57) release 20100804 including low-pass whole genome sequence data from 628 individuals, and OMNI chip genotype data from 764 individuals), consistent with the frequencies observed in 99% of all 895 reported pathogenic OXPHOS alleles within HGMD

version 2010.3 (62); Since estimates of indel allele frequency were less robust, only indels absent in the 1000 genomes data were considered rare. Variants too common among cases (>10% allele frequency) were also excluded. Variants associated with any HGMD disease phenotype were not excluded, unless associated with allele frequency exceeding 1%. ii) variants predicted to modify protein function, including nonsense, splice-site, coding indel, or missense variants at sites conserved across evolution, as previously described (13). Briefly, eight splice site locations were included (acceptor sites -1,-2,-3, and donor sites -1,1,2,3,5), and missense alleles were required to be at protein residues that were identical in at least 10 of 44 aligned vertebrate species. iii) variants consistent with a recessive model of pathogenesis, including homozygous variants or two heterozygous variants present in the same gene. All rare, protein-modifying, and recessive-type variants in both patient and control samples were manually reviewed to remove sequence artifacts and to exclude variants present that were not compound heterozygous based on read or read-pair evidence.

We prioritized mtDNA variants that were annotated as pathogenic in MITOMAP (18) and mtDNA deletions or rearrangements that were detected through *de novo* mtDNA assembly using the ALLPATHS algorithm (59).

Prevalence in cases and controls

Prevalence of prioritized variants was assessed in patients and in 371 healthy individuals of European ancestry obtained with permission from the NIMH Control Sample collection. Existing whole-exome data for the 371 controls, generated by the same Broad Institute Sequencing Platform protocol, were analyzed using the MitoExome analysis pipeline. Prevalence comparisons within the 1034 mitochondrial genes were restricted to the 6,872 autosomal coding exons that were well measured in both study designs, totaling 1Mb (76% of all autosomal coding exons), where each included exon had >80% of bases covered by >10 sequence reads (base quality >5, mapping quality > 0), in at least 90% of patient samples and at least 90% of control samples. Using the same criteria, comparisons within the 347 auxiliary genes were restricted to the 2,619 autosomal coding exons, totaling 381Kb (77% of all such coding exons), that were well measured in both study designs. Significance was assessed by one-sided, unpaired t-test with unequal variance (Fig. 1A) and by one-sided Wilcoxon rank sum (Fig. 1B,C). Details of variant prevalence in cases and controls are shown in Fig. S1. We note that for patients, we conservatively excluded heterozygous variants shown to be *in cis* based on haplotype phasing (1/11 cases) as well as variants not validated by Sanger sequencing (1/49 variants). These exclusions may underestimate the degree of enrichment in cases over controls.

Variant validation and phasing

All prioritized variants detected in patients were independently validated by Sanger sequencing (48/49 variants validated), and compound heterozygous variants were phased through sequencing cDNA (*GFMI*), cloned DNA (*BCS1L*, *C1orf31*, *TYMP*, *MTHFD1L*), familial DNA (*GFMI*, *AGK*, *EARS2*), or by a molecular haplotyping approach (31) (*ACAD9*, *AARS2*, *POLG*) described in Supplementary Methods. 10/11 putative compound heterozygous variants were confirmed and two instances are not yet phased (*ACOX3* and *ALDH1B1*). Two heterozygous variants in *MRPL37* were determined to be on the same maternal haplotype and therefore *MRPL37* was not listed as a prioritized candidate. One *C14orf159* variant was determined to have been introduced by whole genome amplification and was excluded. The mtDNA deletion was validated by long-range PCR using primers 2F (m.1650–1671) and D1R (m.19-1) and Sanger sequencing using internal primers near the breakpoint (available on request). The heteroplasmy level in skeletal muscle was determined by quantitative real-time PCR as previously described (63) and the relative abundance of

mtDNA versus nuclear DNA was measured in patient and control tissues as previously described (64).

Evidence of pathogenicity

Multiple assays were performed to support the pathogenicity of novel alleles within genes known to cause OXPHOS disease (Fig. S2–S8). Presence and segregation of alleles in family members (when available) was assayed by Sanger sequencing. The effect of mutations on mature RNA transcript splicing and abundance was measured in patient and control fibroblasts by RT-PCR, in the presence and absence of 100ng/μl cycloheximide for 24 hours to inhibit nonsense mediated decay (65). Relative abundance of full-length protein products and OXPHOS subunits or assembly factors from all 5 complexes was measured in patient and control fibroblast lines and skeletal muscle via SDS-PAGE western blot (13, 66). Prioritized genes were excluded if they did not segregate with disease in the family (*FAM82A1* in P22 and *COX10* in P26) or if variants showed >0.005 allele frequency in healthy individuals (*MUL* in P12 and *MRPL9* in P24). See Supplementary Methods for details. Prioritized variants were within individuals of western European ancestry, except as noted below, and therefore allele frequencies were obtained from the 1000 genomes project (Table S2). Prioritized variants within patients of Lebanese descent (P12, P13, P19, P21, P22, P26, P27, P28, P30) were screened using Sequenom in 168 ethnically matched chromosomes. Only the *ACADSB* variant (homozygous in P13) was present in any other sample, where it was observed in heterozygous state in 2/168 Lebanese chromosomes (0.012 allele frequency). We note that ethnically matched allele frequencies were not obtained for prioritized variants in P20 (Vietnamese ancestry), P42 (Pakistani ancestry), P14 or P15 (uncertain ancestry).

Molecular karyotyping

Molecular karyotyping of proband P3 and maternal DNA samples was performed using the Illumina HumanCytoSNP-12 array (version 2.1) as previously described (67). Automated LCSH detection was performed using the cnvPartition v3.1.6 algorithm in KaryoStudio software. SNP genotypes were generated in GenomeStudio software (Illumina) using data from a set of 102 intra-run samples. Both the proband and maternal DNA samples yielded a SNP call rate of 99.5%. Identity by state (IBS) analysis, assessing the sharing of alleles between individuals, was performed using the online SNPduo program (38) (pevsnerlab.kennedykrieger.org/SNPduo) with default parameters.

NDUFB3 complementation

The *NDUFB3* open reading frame (ORF) was amplified from control fibroblast cDNA and cloned into the 4-hydroxytamoxifen-inducible lentiviral vector pF-5xUAS-MCS-SV40-puroGal4ER^{T2}VP16 (GEV16)-W (68) using 5' *BclI* and 3' *XbaI*. The pF-5xUAS-*C8orf38*-SV40 puroGEV16-W vector was described previously (PMID 22019594). To generate lentiviral particles, HEK293T cells were grown on 10-cm plates to 60% confluence and cotransfected using Effectene reagents (Qiagen) with packaging plasmid (pCMVδR8.2), a pseudotyping plasmid (pCMV-VSVg) and either pF-5xUAS-*NDUFB3*-SV40 puroGEV16-W or pF-5xUAS-*C8orf38*-SV40 puroGEV16-W. Fresh medium was applied to the cells 16 h after transfection and, after another 24 h incubation, supernatants containing packaged virus were harvested and filtered through a 0.45-μM membrane filter. Primary human fibroblasts were infected with viral supernatant along with 5 μg/ml polybrene (Sigma) for 48 h. Cells were grown in antibiotic-free medium for 30 h before applying selection medium containing 1 μg/ml puromycin. We found that our lenti-vectors were “leaky” in primary human fibroblast lines and that sufficient expression of *C8orf38* and *NDUFB3* was achieved without tamoxifen induction. Following 21 days of selection, cells were harvested for enzyme activity assays and SDS-PAGE western blotting.

Complex I and IV enzyme activity assays

Complex I and complex IV dipstick activity assays (Mitosciences) were performed on 15 µg and 20 µg, respectively, of cleared cell lysates according to the manufacturer's protocol. A Hamamatsu ICA-1000 immunochromatographic dipstick reader was used for densitometry. Two-way repeated measures analysis of variance (ANOVA) was used for comparisons of groups followed by post hoc analysis using the Bonferroni method to determine statistically significant differences.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank C. Guiducci, C. Sougnez, L. Ambrogia, J. Wilkinson, for assistance with sample preparation and sequencing, R.W. Taylor for provision of data on previously sequenced mtDNA samples and long-range PCR primers, M. Ryan and M. McKenzie for gifts of CIA30 and ECSIT antibodies, S. Tregoning and W. Salter for enzyme analyses, J. Marum for designing and performing the Sequenom assay, P. Dear for advice on molecular haplotyping, T. Fennel, M. DePristo, E. Banks, and K. Garimella for assistance with bioinformatic data analysis, S. Flynn for assistance with IRBs, and the subjects and referring physicians who participated in the study.

Funding: Supported by grants from the National Institutes of Health (R01GM077465, 1R01GM097136, to VKM; HD32062 to SD), a grant and Principal Research Fellowship from the Australian National Health & Medical Research Council (436901, 436906 to DRT), a National Defense Science and Engineering Graduate (NDSEG) fellowship (SGH), a Melbourne Research Scholarship (SCL), an Australian Postgraduate Award (EJT), the Marriott Mitochondrial Disorders Clinical Research Fund (SD) and by the Victorian Government's Operational Infrastructure Support Program (DRT).

References

- Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Özen S, Sanjad S, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences*. 2009; 106:19096–19101.
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009; 461:272–276. [PubMed: 19684571]
- Skladal D, Halliday J, Thorburn DR. Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain*. 2003; 126:1905–12. [PubMed: 12805096]
- DiMauro, S.; Hirano, M.; Schon, EA. *Mitochondrial Medicine*. Informa Healthcare; New York: 2006. p. 368
- Dimauro S, Davidzon G. Mitochondrial DNA and disease. *Ann Med*. 2005; 37:222–32. [PubMed: 16019721]
- Chinnery PF, Turnbull DM. Epidemiology and treatment of mitochondrial disorders. *Am J Med Genet*. 2001; 106:94–101. [PubMed: 11579428]
- Kirby DM, Thorburn DR. Approaches to finding the molecular basis of mitochondrial oxidative phosphorylation disorders. *Twin Res Hum Genet*. 2008; 11:395–411. [PubMed: 18637740]
- Tucker EJ, Compton AG, Thorburn DR. Recent advances in the genetics of mitochondrial encephalopathies. *Curr Neurol Neurosci Rep*. 2010; 10:277–85. [PubMed: 20446063]
- M. G. L. Baylor College of Medicine. 2011
- MEDomics. Sep 30. 2011 www.medomics.com/MitoNucleomeDx
- GeneDX. Next-Generation Sequence Analysis of 24 Nuclear Genes Important for Normal Mitochondrial Function. Sep 30. 2011 www.genedx.com/pdf_files/Nuclear_panel_24_gene_descriptions.pdf
- Scharfe C, Lu HHS, Neuenburg JK, Allen EA, Li GC, Klopstock T, Cowan TM, Enns GM, Davis RW. Mapping Gene Associations in Human Mitochondria using Clinical Disease Phenotypes. *PLoS Comput Biol*. 2009; 5:e1000374. [PubMed: 19390613]

13. Calvo SE, Tucker EJ, Compton AG, Kirby DM, Crawford G, Burt NP, Rivas M, Guiducci C, Bruno DL, Goldberger OA, et al. High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat Genet.* 2010; 42:851–8. [PubMed: 20818383]
14. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, Walford GA, Sugiana C, Boneh A, Chen WK, et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell.* 2008; 134:112–23. [PubMed: 18614015]
15. Shen P, Wang W, Krishnakumar S, Palm C, Chi AK, Enns GM, Davis RW, Speed TP, Mindrinos MN, Scharfe C. High-quality DNA sequence capture of 524 disease candidate genes. *Proceedings of the National Academy of Sciences.* 2011; 108:6549–6554.
16. Vasta V, Ng S, Turner E, Shendure J, Hahn SH. Next generation sequence analysis for mitochondrial disorders. *Genome Med.* 2009; 1:100. [PubMed: 19852779]
17. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010; 467:52–8. [PubMed: 20811451]
18. Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, Yi C, Kreuziger J, Baldi P, Wallace DC. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res.* 2007; 35:D823–8. [PubMed: 17178747]
19. Li M, Schonberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet.* 2010; 87:237–49. [PubMed: 20696290]
20. Hirt L, Magistretti PJ, Bogousslavsky J, Boulat O, Borruat FX. Large deletion (7.2 kb) of mitochondrial DNA with novel boundaries in a case of progressive external ophthalmoplegia. *Journal of Neurology, Neurosurgery & Psychiatry.* 1996; 61:422–423.
21. Nouws J, Nijtmans L, Houten SM, van den Brand M, Huynen M, Venselaar H, Hoefs S, Gloerich J, Kronick J, Hutchin T, et al. Acyl-CoA dehydrogenase 9 is required for the biogenesis of oxidative phosphorylation complex I. *Cell Metab.* 2010; 12:283–94. [PubMed: 20816094]
22. Haack TB, Danhauser K, Haberberger B, Hoser J, Strecker V, Boehm D, Uziel G, Lamantea E, Invernizzi F, Poulton J, et al. Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat Genet.* 2010; 42:1131–4. [PubMed: 21057504]
23. Stumpf JD, Copeland WC. Mitochondrial DNA replication and disease: insights from DNA polymerase gamma mutations. *Cell Mol Life Sci.* 2010; 68(2):219–33. [PubMed: 20927567]
24. Van Goethem G, Dermaut B, Lofgren A, Martin JJ, Van Broeckhoven C. Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. *Nat Genet.* 2001; 28:211–2. [PubMed: 11431686]
25. de Lonlay P, Valnot I, Barrientos A, Gorbatyuk M, Tzagoloff A, Taanman JW, Benayoun E, Chretien D, Kadhom N, Lombes A, et al. A mutant mitochondrial respiratory chain assembly protein causes complex III deficiency in patients with tubulopathy, encephalopathy and liver failure. *Nat Genet.* 2001; 29:57–60. [PubMed: 11528392]
26. Massa V, Fernandez-Vizarra E, Alshahwan S, Bakhsh E, Goffrini P, Ferrero I, Mereghetti P, D'Adamo P, Gasparini P, Zeviani M. Severe infantile encephalomyopathy caused by a mutation in COX6B1, a nucleus-encoded subunit of cytochrome c oxidase. *Am J Hum Genet.* 2008; 82:1281–9. [PubMed: 18499082]
27. Coenen MJ, Antonicka H, Ugalde C, Sasarman F, Rossi R, Heister JG, Newbold RF, Trijbels FJ, van den Heuvel LP, Shoubridge EA, et al. Mutant mitochondrial elongation factor G1 and combined oxidative phosphorylation deficiency. *N Engl J Med.* 2004; 351:2080–6. [PubMed: 15537906]
28. Smeitink JA, Elpeleg O, Antonicka H, Diepstra H, Saada A, Smits P, Sasarman F, Vriend G, Jacob-Hirsch J, Shaag A, et al. Distinct clinical phenotypes associated with a mutation in the mitochondrial translation elongation factor EFTs. *Am J Hum Genet.* 2006; 79:869–77. [PubMed: 17033963]
29. Gotz A, Tyynismaa H, Euro L, Ellonen P, Hyotylainen T, Ojala T, Hamalainen RH, Tommiska J, Raivio T, Oresic M, et al. Exome Sequencing Identifies Mitochondrial Alanine-tRNA Synthetase Mutations in Infantile Mitochondrial Cardiomyopathy. *Am J Hum Genet.* 2011; 88:635–42. [PubMed: 21549344]

30. Nishino I, Spinazzola A, Hirano M. Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science*. 1999; 283:689–92. [PubMed: 9924029]
31. Konfortov BA, Bankier AT, Dear PH. An efficient method for multi-locus molecular haplotyping. *Nucleic Acids Res*. 2007; 35:e6. [PubMed: 17158153]
32. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010; 42:30–5. [PubMed: 19915526]
33. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*. 2010; 42:790–3. [PubMed: 20711175]
34. King MP, Koga Y, Davidson M, Schon EA. Defects in mitochondrial protein synthesis and respiratory chain activity segregate with the tRNA(Leu(UUR)) mutation associated with mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes. *Mol Cell Biol*. 1992; 12:480–90. [PubMed: 1732728]
35. Bektas M, Payne SG, Liu H, Goparaju S, Miltstien S, Spiegel S. A novel acylglycerol kinase that produces lysophosphatidic acid modulates cross talk with EGFR in prostate cancer cells. *J Cell Biol*. 2005; 169:801–11. [PubMed: 15939762]
36. Spinazzola A, Invernizzi F, Carrara F, Lamantea E, Donati A, Dirocco M, Giordano I, Meznaric-Petrusa M, Baruffini E, Ferrero I, et al. Clinical and molecular features of mitochondrial DNA depletion syndromes. *J Inherit Metab Dis*. 2009; 32:143–58. [PubMed: 19125351]
37. Saada A, Shaag A, Mandel H, Nevo Y, Eriksson S, Elpeleg O. Mutant mitochondrial thymidine kinase in mitochondrial DNA depletion myopathy. *Nat Genet*. 2001; 29:342–4. [PubMed: 11687801]
38. Roberson ED, Pevsner J. Visualization of shared genomic regions and meiotic recombination in high-density SNP data. *PLoS One*. 2009; 4:e6711. [PubMed: 19696932]
39. Zaghoul NA, Liu Y, Gerdes JM, Gascue C, Oh EC, Leitch CC, Bromberg Y, Binkley J, Leibel RL, Sidow A, et al. Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. *Proc Natl Acad Sci U S A*. 2010; 107:10602–7. [PubMed: 20498079]
40. Tuppen HA, Fattori F, Carrozzo R, Zeviani M, DiMauro S, Seneca S, Martindale JE, Olpin SE, Treacy EP, McFarland R, et al. Further pitfalls in the diagnosis of mtDNA mutations: homoplasmic mt-tRNA mutations. *J Med Genet*. 2008; 45:55–61. [PubMed: 18178636]
41. O’Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, MacKenzie AP, Ng SB, Baker C, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet*. 2011; 43:585–589. [PubMed: 21572417]
42. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*. 2008; 10:294–300. [PubMed: 18414213]
43. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med*. 2011; 3:65ra4.
44. Thorburn DR, Smeitink J. Diagnosis of mitochondrial disorders: clinical and biochemical approach. *J Inherit Metab Dis*. 2001; 24:312–6. [PubMed: 11405352]
45. Ingman M, Gyllensten U. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res*. 2006; 34:D749–51. [PubMed: 16381973]
46. 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. [PubMed: 12427892]
47. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2007; 35:D61–5. [PubMed: 17130148]

48. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, et al. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* 39:D876–82. [PubMed: 20959295]
49. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res.* 2002; 12:996–1006. [PubMed: 12045153]
50. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology.* 2009; 27:182–9.
51. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 2008; 456:53–9. [PubMed: 18987734]
52. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A, Ambrogio L, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* 2011; 12:R1. [PubMed: 21205303]
53. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25:1754–60. [PubMed: 19451168]
54. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–303. [PubMed: 20644199]
55. Stenson PD, Ball EV, Howells K, Phillips AD, Mort M, Cooper DN. The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics.* 2009; 4:69–72. [PubMed: 20038494]
56. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001; 29:308–11. [PubMed: 11125122]
57. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature.* 467:1061–73. [PubMed: 20981092]
58. Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* 6:e1000834. [PubMed: 20168995]
59. Maccallum I, Przybylski D, Gnerre S, Burton J, Shlyakhter I, Gnirke A, Malek J, McKernan K, Ranade S, Shea TP, et al. ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. *Genome Biol.* 2009; 10:R103. [PubMed: 19796385]
60. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A.* 2011; 108:1513–8. [PubMed: 21187386]
61. Pevzner PA, Tang H, Waterman MS. An Eulerian path approach to DNA fragment assembly. *Proc Natl Acad Sci U S A.* 2001; 98:9748–53. [PubMed: 11504945]
62. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN. The Human Gene Mutation Database: 2008 update. *Genome Med.* 2009; 1:13. [PubMed: 19348700]
63. He L, Chinnery PF, Durham SE, Blakely EL, Wardell TM, Borthwick GM, Taylor RW, Turnbull DM. Detection and quantification of mitochondrial DNA deletions in individual cells by real-time PCR. *Nucleic Acids Research.* 2002; 30:e68. [PubMed: 12136116]
64. Pagnamenta AT, Taanman J-W, Wilson CJ, Anderson NE, Marotta R, Duncan AJ, Glindzicz MB, Taylor RW, Laskowski A, Thorburn DR, et al. Dominant inheritance of premature ovarian failure associated with mutant mitochondrial DNA polymerase gamma. *Human Reproduction.* 2006; 21:2467–2473. [PubMed: 16595552]
65. Lamande SR, Bateman JF, Hutchison W, McKinlay Gardner RJ, Bower SP, Byrne E, Dahl HH. Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency. *Hum Mol Genet.* 1998; 7:981–9. [PubMed: 9580662]
66. Cooper ST, Lo HP, North KN. Single section Western blot: improving the molecular diagnosis of the muscular dystrophies. *Neurology.* 2003; 61:93–7. [PubMed: 12847163]

67. Bruno DL, Stark Z, Amor DJ, Burgess T, Butler K, Corrie S, Francis D, Ganesamoorthy D, Hills L, James PA, et al. Extending the scope of diagnostic chromosome analysis: Detection of single gene defects using high-resolution SNP microarrays. *Hum Mutat.* 2011; 32:12.
68. Yeap YY, Ng IH, Badrian B, Nguyen TV, Yip YY, Dhillon AS, Mutsaers SE, Silke J, Bogoyevitch MA, Ng DC. c-Jun N-terminal kinase/c-Jun inhibits fibroblast proliferation by negatively regulating the levels of stathmin/oncoprotein 18. *Biochem J.* 2010; 430:345–54. [PubMed: 20594188]
69. Lamantea E, Tiranti V, Bordoni A, Toscano A, Bono F, Servidei S, Papadimitriou A, Spelbrink H, Silvestri L, Casari G, et al. Mutations of mitochondrial DNA polymerase gammaA are a frequent cause of autosomal dominant or recessive progressive external ophthalmoplegia. *Ann Neurol.* 2002; 52:211–9. [PubMed: 12210792]
70. Agostino A, Valletta L, Chinnery PF, Ferrari G, Carrara F, Taylor RW, Schaefer AM, Turnbull DM, Tiranti V, Zeviani M. Mutations of ANT1, Twinkle, and POLG1 in sporadic progressive external ophthalmoplegia (PEO). *Neurology.* 2003; 60:1354–6. [PubMed: 12707443]
71. Visapaa I, Fellman V, Vesa J, Dasvarma A, Hutton JL, Kumar V, Payne GS, Makarow M, Van Coster R, Taylor RW, et al. GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. *Am J Hum Genet.* 2002; 71:863–76. [PubMed: 12215968]

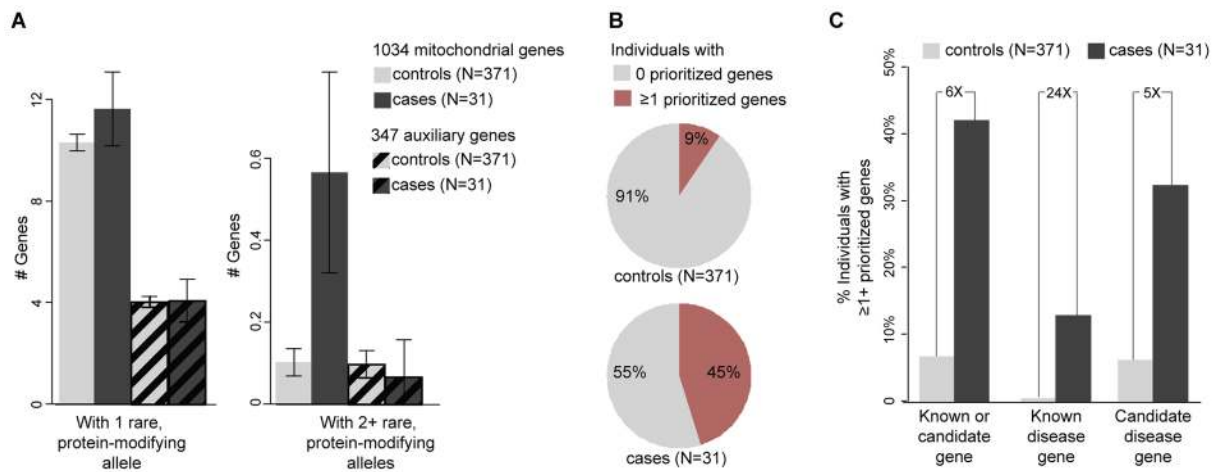


Figure 1. Enrichment of prioritized variants in cases compared to healthy individuals

(A) Mean number of genes containing rare, protein-modifying alleles in cases and healthy controls within the 1034 mitochondrial genes and the 347 auxiliary genes sequenced. Error bars indicate standard error. (B) Percent of cases and controls containing prioritized mitochondrial genes (red), defined as harboring 2 rare, protein-modifying alleles. (C) Percent of cases and controls containing prioritized mitochondrial genes, excluding variants with allele frequency >0.005 in either cases or controls, with enrichment in cases versus controls displayed above. Analysis was restricted to regions well-sequenced in all individuals, and to the 31 cases from non-consanguineous families.

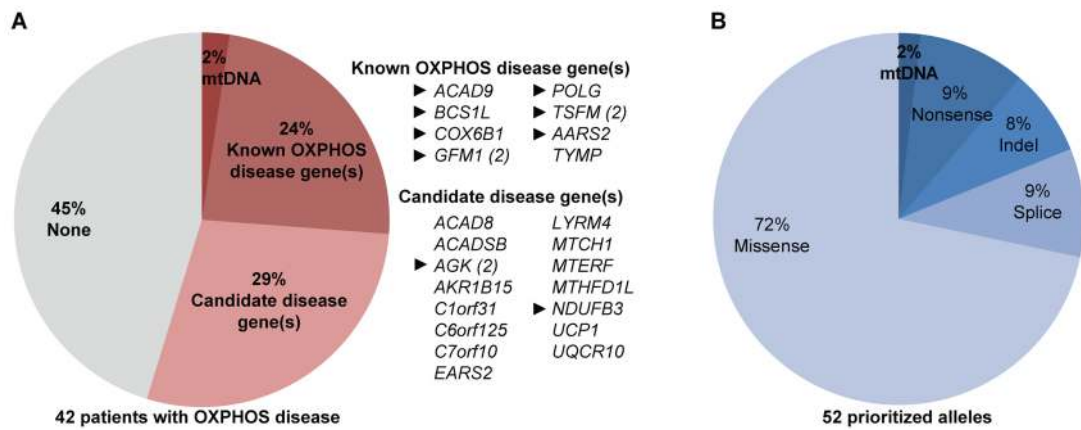


Figure 2. Prioritized variants in 42 patients with OXPHOS disease

(A) 42 patients categorized by presence of prioritized genes (red). The gene names are listed alphabetically at right, with parentheses indicating genes prioritized in two unrelated patients, and triangle indicating support of pathogenicity. (B) 52 prioritized alleles, categorized by type of protein modification.

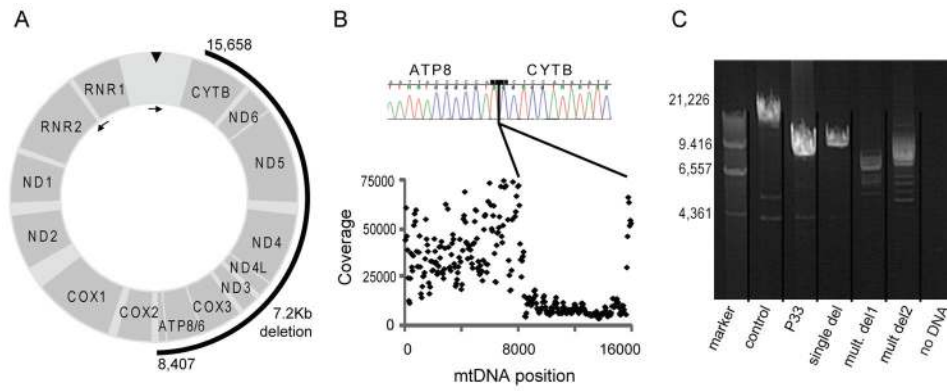


Figure 3. mtDNA deletion in patient P33

(A) Schematic diagram of mtDNA indicates deletion (black arc) relative to position 0/16,569 (triangle). Arrows indicate long-range PCR primers. (B) mtDNA sequence coverage in 50bp windows. Inset shows Sanger electropherogram of breakpoint. (C) Gel electrophoresis of long-range PCR amplicon displays 14,933 bp fragment in control DNA and 7,681 bp fragment in P33, along with marker II (Roche) and mtDNA from individuals with confirmed single and multiple deletions.

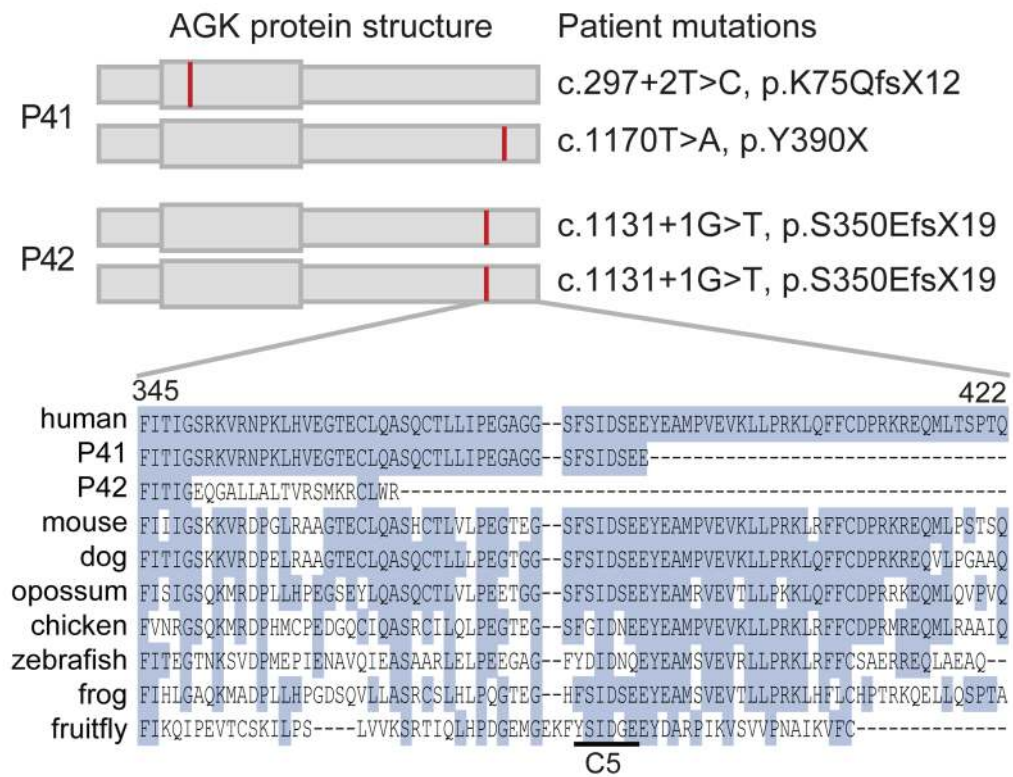


Figure 4. Patient mutations in acylglycerol kinase (AGK)

Schematic diagram shows protein structure with the location of truncating mutations in unrelated patients P41 and P42 shown in red. Conservation of C-terminal region is shown below with identical residues shaded and the C5 domain shared with yeast sphingosine kinases underlined.

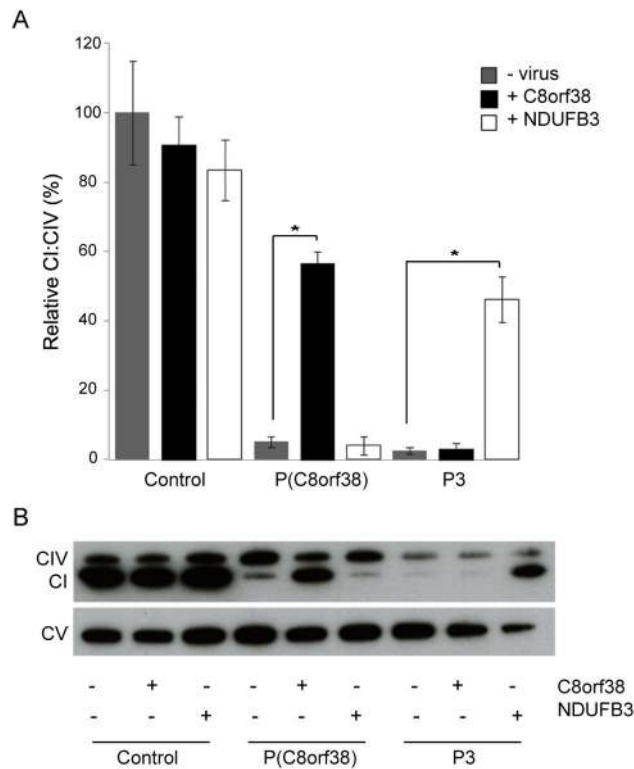


Figure 5. *NDUFB3* complementation of complex I defects in subject P3 fibroblasts

(A) Barplots show complex I (CI) activity, normalized by complex IV (CIV) activity, in fibroblasts from patient P3, a healthy control individual, and an unrelated complex I patient with a defined *C8orf38* mutation P(*C8orf38*). Enzyme activity was measured before and after transduction with wild-type *C8orf38* or *NDUFB3* mRNA. Data shown are mean of three biological replicates \pm s.e.m. * $P < 0.001$. (B) Western blot shows protein expression of complex I, complex IV, and loading control complex V (CV).

Table 1

Patient clinical, biochemical and molecular characteristics

ID	Sex	Age of onset (death)	Clinical Presentation	Family Hx	Biochemical Features					Gene Candidate(s)	
					Tissue(s)	I	II	III	IV		mtDNA quantity
Complex I	P1	<1mo	lethal infantile mitochondrial disease		fib	↓↓				nd	*ACAD9
	P2	<1wk (3wk)	hypotonia, lethargy, respiratory distress, metabolic acidosis, FTT		fib	↓↓		↓		nd	*POLG
	P3	<1wk (4mo)	severe IUGR, FTT, recurrent episodes of metabolic acidosis		skm, fib	↓↓				nd	NDUFB3
	P4	<i>in utero</i> (<i>in utero</i>)	multiple pterygium syndrome, severe fetal hydrops, IUGR, terminated at 21w	cons. (1 st)	liv fib	↓↓	~			nd	AKR1B15
	P5	<1wk (1yr)	lethargy, tachypnea, organomegaly, renal & liver failure		skm, liv, hea	↓↓				49% liv	-
	P6	<1wk	bilateral optic nerve hypoplasia, lethargy, FTT		skm, liv, fib	↓				nd	-
	P7	<6mo (28mo)	cystic leukodystrophy, FTT, severe metabolic acidosis		skm, liv, fib	↓↓				nd	-
Complex III	P8	<3m (8yr)	leukodystrophy, DD, abnormal LFTs, vasculitic rash		skm fib liv	↓↓	↓↓			nd	-
	P9	<i>in utero</i> (4d)	leukodystrophy, IUGR, metabolic acidosis, seizures		liv, fib skm	↓↓	↓↓			460% liv	-
	P10	<1yr (na)	FTT, DD, seizures, hypotonia, prominent extra-axial CSF space		skm	↓				nd	-
	P11	<1wk	severe global DD, tachypnea, metabolic acidosis, renal tubular acidosis, FTT		skm, liv fib			↓↓	nl	47% skm	*BCSL1
	P12	<1wk (na)	severe IUGR, lethargy, metabolic acidosis, renal tubular acidosis, dysmorphic facies	cons. (1 st).	skm fib	↓	~	↓↓	↓↓	nd	*TYMP, MTCH1, C6orf125
	P13	<6mo	hypotonia, FTT, DD, short stature, rickets, pruritis, dysmorphic facies	cons. (1 st)	skm liv			↓↓	~	nd	ACADSB
	P14	<1wk (na)	fetal hypokinesia, Pierre Robin sequence, intra-abdominal calcification		skm, fib			↓↓		nd	UCP1, MTHFD1L
Complex IV	P15	<i>in utero</i> (5yr)	ventriculomegaly, apnea, dev. regression, hypotonia, seizures		skm fib	~	↓	↓↓	↓↓	nd	UQCRI0
	P16	<1wk	cardiorespiratory arrest, metabolic acidosis, renal tubular acidosis		fib			↓↓		nd	-
	P17	<1wk	neonatal lactic acidosis, cystic leukodystrophy		skm fib				↓↓	nl	*COX6B1
	P18	<1wk	hypsarhythmia, FTT, dystonia, squint		skm fib	~		~	↓↓	↓	*GFML, ACOX3
	P19	<1mo (12mo)	DD, seizure, hypotonia, doesn't fix & follow, cortical atrophy	cons. (1 st +1)	fib				↓↓	nd	ACAD8
	P20	<1yr	mild global DD, hypotonia, mild cerebral atrophy	sib.	skm fib				↓↓	nl	C7orf10, MTERF

ID	Sex	Age of onset (death)	Clinical Presentation	Family Hx	Biochemical Features					Gene Candidate(s)	
					Tissue(s)	I	II	III	IV		mtDNA quantity
P21	M	1 mo (7mo)	Leigh syndrome, global DD, FTT, hypotonia, seizures	cons. (2 nd)	fib				↓↓	nd	-
P22	F	<i>in utero</i>	IUGR, oligohydramnios, severe global DD, seizures, microcephaly, hyperthermia	cons. (1 st), sib.	skm, liv fib				↓↓ nl	nd	-
P23	F	<1mo	profound hypotonia, paucity of movement		skm				↓↓	37% skm	-
P24	M	<1yr	DD, hypotonia, apnea, spastic diplegia, dysmorphic		skm fib				↓↓ ↓	84% skm	-
P25	M	<6mo	Leigh-like syndrome, respiratory failure, lactic acidosis, RRF		skm				↓↓	nd	-
P26	M	3wk	FTT, metabolic acidosis, hepatomegaly, apnea	cons. (1 st)	skm liv fib	↓↓ ~ nl	↓↓ ~ nl	↓↓ ~ nl		nd	<i>LYRM4</i>
P27	M	<1wk (2mo)	FTT, hypotonia, hypothermia, hypertrophic cardiomyopathy, hepatomegaly	cons. (1 st), sib.	liv fib	↓↓ ↓	↓↓ ↓	↓↓ ↓	↓↓ ↓	800% liv	<i>*TSFM</i>
P28	F	<1mo (1mo)	FTT, hypertonia, hypertrophic cardiomyopathy, respiratory arrest	cons. (1 st), sib.	fib	↓		↓	↓	nd	<i>*TSFM</i>
P29	F	<1yr	DD, seizures, hypotonia, episodic metabolic acidosis		liv	↓↓			↓↓	167% liv	<i>*GFMI</i>
P30	M	<1mo	FTT, DD, sensorineural deafness, renal failure, dysmorphic	cons. (1 st), sib.	skm fib	↓↓ ↓	↓↓ ↓	↓↓ ↓	↓↓ ↓	42% skm	
P31	M	<1yr	hypertrophic cardiomyopathy		hea fib	↓↓ nl			↓↓ nl	nd	<i>C1orf31</i>
P32	M	<6mo	gastroesophageal reflux, FTT, hypotonia, respiratory distress, lactic acidosis		skm	↓↓		↓↓	↓↓	nd	<i>EARS2</i>
P33	M	<1yr (20mo)	FTT, dev. regression, microcephaly		skm liv	↓↓ ~	↓↓ ~	↓↓ ~	↓↓ ~	290% skm	<i>ALDH1B1, mtDNA del.</i>
P34	F	<1wk (4mo)	FTT, cardiac arrest, hypertrophic cardiomyopathy		skm fib liv	↓↓ ↓ ~	↓↓ ↓ ~	↓↓ ↓ ~	↓↓ ~ nl	88% skm	-
P35	F	<1wk (10d)	metabolic acidosis, cardiac failure, hemorrhagic brain infarct		liv fib	↓↓ nl			↓↓ nl	190% liv	-
P36	M	<1yr	sudden-onset encephalopathy with seizures after URTI		skm liv	↓↓ ~		~	↓↓ ↓	92% skm	-
P37	F	<1yr	DD, ptosis, microcephaly, GI dysmotility		skm liv	↓↓ nl			↓↓ nl	53% skm	-
P38	F	<1yr	leukodystrophy, speech delay, dysaesthesia in hands, dev. regression		skm fib	↓↓ nl		~	↓↓ nl	112% skm	-
P39	M	<2yr	Leigh syndrome, global DD, deterioration after URTI		skm fib	↓ ↓		↓ ↓	↓ ~	nd	-
P40	F	<i>in utero</i> (stillborn)	stillborn with hypotonia and multiple fractures		skm fib	↓↓ nl	↓↓ nl	↓↓ nl	↓↓ nl	8% skm	<i>*AARS2</i>

Combined OXPHOS deficiencies

mtDNA dep.

\$watermark-text

\$watermark-text

\$watermark-text

ID	Sex	Age of onset (death)	Clinical Presentation	Family Hx	Biochemical Features					Gene Candidate(s)	
					Tissue(s)	I	II	III	IV		mtDNA quantity
P41	F	<1yr (18y)	cardioskeletal myopathy, cataracts, FTT, fatigue		skm	↓↓		↓↓	↓↓	4% skm	<i>AGK</i>
P42	F	<1wk (4d)	marked respiratory distress, pulmonary hypertension, cataracts	cons. (1 st)	skm liv	↓↓ nl		↓↓ nl	↓↓ nl	13% skm 88% liv	<i>AGK</i>

Patients are ordered by primary biochemical defect. Where appropriate, age at death is shown in parentheses. Family history (Hx) includes consanguinity (cons.), with level of consanguinity noted (1st cousin, 2nd cousin, or 1st cousin once removed indicated by 1st, 2nd, and 1st+1) and/or sibling with features of OXPPOS disease (sib.). Biochemical features were measured in up to four tissues (skm, skeletal muscle; liv, liver; fib, fibroblasts; hea, heart) and expressed relative to the marker enzymes citrate synthase and complex II. Enzyme activity per tissue per complex (I, II, III, IV) were categorized as ↓↓ markedly deficient; ↓ moderately deficient; ~, equivocal; or nl, normal. mtDNA quantity denotes % in given tissue relative to the mean value for normal controls, with values <20% regarded as mtDNA depletion. Gene names shown in bold represent genes with mutations that are believed to cause the clinical and biochemical phenotype. Previously reported OXPPOS disease genes are marked with an asterisk. Other abbreviations: na/nd, not available/not determined; d/wk/mo/yr, day/week/month/year; M/F, Male/Female; DD, Developmental Delay; Dev., Developmental; FTT, Failure To Thrive; IUGR, Intra-Uterine Growth Restriction; LFT, Liver Function Test; CSF, CerebroSpinal Fluid; URTI, Upper Respiratory Tract Infection; GI, GastroIntestinal; del, Deletion; dep., Depletion.

Table 2

MitoExome sequencing statistics

Targeted DNA	Nuclear DNA *	mtDNA *
# gene loci	1034	37
target size (bp)	1,392,193	16,569
coverage †		
mean coverage	142 [72–257]	25,457 [1,477–87,135]
% target bp covered $\geq 1X$	96 [91–98]	100 [100]
% target bp covered $\geq 10X$	89 [76–94]	100 [100]
% target bp covered $\geq 15X$	87 [71–92]	100 [100]
# variants compared to reference DNA	773 [571–912]	36 [11–50]
# rare	29 [17–56]	3 [0–23]
# rare, protein-modifying	17 [9–33]	0 [0–4]
# genes with 2 rare, protein-modifying alleles	1 [0–3]	NA

* Median values per patient, with range across patients indicated in brackets

† excluding reads with ambiguous alignment

Table 3

Support of molecular diagnosis in 13 patients

ID	Disease subtype	Gene	Type	Mutations	ACMG category	Support of pathogenicity
P1	complex I deficiency (MIM:252010)	* <i>ACAD9</i>	cmpd het	c.976G>C, p.A326P(22) c.260T>A, p.I87N	1 3	conservation, protein degradation
P2	complex I deficiency (MIM:252010)	* <i>POLG</i>	cmpd het	c.2542G>A, p.G848S(69) c.679C>T, p.R227W(70)	1 1	Mother carrier, no deletion or isoparental disomy detected by Affy2.7M & Illumina 300K arrays, protein degradation, conservation, correction studies
P3	complex I deficiency (MIM:252010)	<i>NDUFB3</i>	hom.	c.64T>C, p.W22R	3	conservation, mRNA and protein degradation
P11	complex III deficiency (MIM:124000)	* <i>BCSIL</i>	cmpd het	c.166C>T, p.R56X(71) c.268C>T, p.R90C	1 3	protein degradation family segregation
P17	complex IV deficiency (MIM:220110)	* <i>COX6BI</i>	hom.	c.241A>C, p.T81P	3	mRNA degradation protein degradation family segregation
P18	complex IV deficiency	* <i>GFM1</i>	cmpd het	c.720delT, p.E241NfsX1 c.2011C>T, p.R671C	2 3	mRNA degradation protein degradation maternal carrier c.720delT
P29	combined deficiency (MIM:609060)	* <i>GFM1</i>	cmpd het	c.720delT, p.E241NfsX1 c.910A>G, p.K304E	2 3	family segregation
P27	combined deficiency (MIM:610505)	* <i>TSFM</i>	hom.	c.997C>T, p.R333W(28)	1	family segregation
P28	combined deficiency (MIM:610505)	* <i>TSFM</i>	hom.	c.997C>T, p.R333W(28)	1	family segregation
P33	combined deficiency	mtDNA deletion	%94 mut	m.8407_15658del7252(19)	1	qPCR, LR-PCR and sequencing
P40	mtDNA depletion	* <i>AARS2</i>	cmpd het	c.986G>A, p.R329H c.1774C>T, p.R592W(29)	3 1	conservation, protein degradation
P41	myopathic mtDNA depletion (MIM:609560)	<i>AGK</i>	cmpd het	c.297+2T>C, p.K750fsX12 c.1170T>A, p.Y390X	2	independent mutations in 2 patients
P42	myopathic mtDNA depletion (MIM:609560)	<i>AGK</i>	hom.	c.1131+1G>T, p.S350EfsX19	2	independent mutations in 2 patients family segregation

* Indicates genes previously associated with OXPHOS disease. Abbreviations: cmpd het, compound heterozygous; hom., homozygous; het, heterozygous; mut, heteroplasmic mutant load.

classification of pathogenicity for each mutation using the American College of Medical Genetics (ACMG) categories(42).