Editor's Summary

Getting to the Genetic Root of Mitochondrial Disease

Next-generation DNA sequencing is being applied with great success in research settings to uncover new disease genes. Despite these successes, it is unclear how useful the technology will be for routine clinical diagnosis given the challenge of interpreting DNA variations in individual patients. In a new study, Calvo and colleagues apply next-generation sequencing to infants with mitochondrial disorders, a large collection of inherited diseases that are notoriously difficult to diagnose because of the multitude of candidate genes and the highly variable nature of the clinical presentation.

First, the authors selected 42 unrelated infants with mitochondrial diseases that were refractory to standard clinical genetic testing. Then, for each child, they sequenced the DNA of the mitochondrial genome, the 100 genes previously linked to mitochondrial disease, and the ~1000 additional genes that are known to participate in mitochondrial biology. Of all the DNA differences present in the patients, the researchers prioritized those that were rare in the general population, predicted to disrupt protein function, and inherited in a recessive fashion. Such variants showed fivefold enrichment in the patients compared to that in healthy control individuals. In 10 patients (24%), firm molecular diagnoses were made in genes previously linked to mitochondrial diseases; 13 patients (31%) had prioritized recessive mutations in genes not previously linked to disease. For two of these genes, the authors were able to show that the mutations caused the mitochondrial disorder. These results suggest that next-generation sequencing may be able to provide a molecular diagnosis for ~25% of currently unsolved cases of infantile mitochondrial disease. An additional 25% of cases could be solved in the coming few years as more genes are formally proven to be linked to mitochondrial disease. The remaining 50% of patients in whom diagnosis was not possible underscores the challenge of interpreting DNA sequence data for clinical diagnosis. Nevertheless, the study by Calvo and colleagues will help to calibrate clinicians' expectations regarding the diagnostic use of next-generation sequencing.

A complete electronic version of this article and other services, including high-resolution figures, can be found at:
http://stm.sciencemag.org/content/4/118/118ra10.full.html

Supplementary Material can be found in the online version of this article at:
http://stm.sciencemag.org/content/suppl/2012/01/23/4.118.118ra10.DC1.html

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl

HUMAN GENETICS

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

Sarah E. Calvo,[1,2,3]* Alison G. Compton,[4]* Steven G. Hershman,[1,2,3] Sze Chern Lim,[4,5] Daniel S. Lieber,[1,2,3] Elena J. Tucker,[4,5] Adrienne Laskowski,[4] Caterina Garone,[6,7] Shangtao Liu,[1] David B. Jaffe,[3] John Christodoulou,[8,9] Janice M. Fletcher,[10,11] Damien L. Bruno,[4,12] Jack Goldblatt,[13] Salvatore DiMauro,[6] David R. Thorburn,[4,5,12]† Vamsi K. Mootha[1,2,3]†

Advances in next-generation sequencing (NGS) promise to facilitate diagnosis of inherited disorders. Although 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 use, we performed targeted NGS in 42 unrelated infants with clinical and biochemical evidence of mitochondrial oxidative phosphorylation disease. These devastating mitochondrial disorders are characterized by phenotypic and genetic heterogeneity, with more than 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. Because patients and healthy control individuals harbored a comparable number of such heterozygous alleles, we could not prioritize dominant-acting genes. However, patients showed a fivefold enrichment of genes with two such mutations that could underlie recessive disease. In total, 23 of 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. Thirteen patients (31%) had mutations in nuclear genes not previously linked to disease. The pathogenicity of two such genes, NDUFB3 and AGK, was supported by complementation studies and evidence from multiple patients, respectively. The results underscore the potential and challenges of deploying NGS in clinical settings.

INTRODUCTION

Advances in next-generation sequencing (NGS) technology are facilitating the discovery of new 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 more than 50 new disease genes in research settings. Although sequence-based clinical diagnosis has long been available for individual genes, NGS could be particularly useful for conditions characterized by genetic heterogeneity, where individual genes cannot be readily prioritized for traditional sequencing. The success of NGS for clinical use, 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 could 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 because it can present early in infancy or in adulthood, can be severe or mild in its presentation, and typically affects multiple organ systems. Clinical manifestations can include, but are not limited to, myopathy, lactic acidosis, seizures, ataxia, peripheral neuropathy, blindness, deafness, gastrointestinal 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 either in the mitochondrial DNA (mtDNA) or in the nuclear genome. It is possible to routinely resequence the mtDNA, but lesions in this tiny genome likely account for no more than 15 to 30% of all childhood cases (5, 7). A recent review listed 77 nuclear disease genes, identified largely through pedigree analysis, with up to 10 new OXPHOS disease genes described each year (8). Most involve recessive, loss-of-function alleles, although nine genes can harbor dominant-acting mutations (POLG,

[1]Center for Human Genetic Research and Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street, Sixth Floor, Boston, MA 02114, USA. [2]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA. [3]Broad Institute of Harvard and Massachusetts Institute of Technology, 7 Cambridge Center, Cambridge, MA 02141, USA. [4]Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville, VIC 3052, Australia. [5]Department of Paediatrics, University of Melbourne, Melbourne, VIC 3052, Australia. [6]Department of Neurology, Columbia University College of Physicians and Surgeons, 710 West 168th Street, New York, NY 10032, USA. [7]Human Genetics Joint PhD Programme, University of Turin, 10125 Turin, Italy and University of Bologna, 40125 Bologna, Italy. [8]Genetic Metabolic Disorders Research Unit, Children's Hospital at Westmead, 212 Hawkesbury Road, Westmead, NSW 2145, Australia. [9]Disciplines of Paediatrics and Child Health and Genetic Medicine, University of Sydney, Sydney, NSW 2006, Australia. [10]Department of Genetics and Molecular Pathology, 72 King William Road, North Adelaide, SW 5006, Australia. [11]Discipline of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, SA 5005, Australia. [12]Victorian Clinical Genetics Services, Royal Children's Hospital, Melbourne, VIC 3052, Australia. [13]Genetic Services of Western Australia, King Edward Memorial Hospital, University of Western Australia, Perth, WA 6009, Australia.
*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: vamsi@hms.harvard.edu (V.K.M.); david.thorburn@mcri.edu.au (D.R.T.)

POLG2, C10orf2, RRM2B, SLC25A4, OPA1, CYCS, MFN2, and 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 (Supplementary Material), the full set of ~1000 genes encoding mitochondrial proteins represents 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 ~142× at each targeted nuclear base and ~25,457× at each mtDNA base (Table 2). On average, 96% of targeted bases were covered and 87% of targeted bases exceeded the 15× 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 who fulfilled the clinical and biochemical criteria for OXPHOS disease before 2 years of age (Table 1). These patients included 11 from consanguineous families and 1 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 about 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. As 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 (Materials and 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, we limited this comparison to the 31 patients from nonconsanguineous families and to MitoExome regions well covered in all individuals (76% of all autosomal coding exons, see Materials and 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 and fig. S1). However, we observed a marked 5.5-fold enrichment of genes containing two such prioritized alleles in patients compared to controls ($P = 3 \times 10^{-11}$) (Fig. 1A and Materials 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). Forty-five percent of all nonconsanguineous cases contained mitochondrial genes with two prioritized variants compared to only 9% of controls (4.8-fold enrichment, $P = 7 \times 10^{-5}$) (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 on the basis of mixing experiments (19).

For the nine 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 Supplementary Material, none of the six such variants appeared to be dominant mutations associated with the observed patient phenotypes.

Table 1. Patient clinical, biochemical, and molecular characteristics. Patients are ordered by primary biochemical defect. Where appropriate, age at death is shown in parentheses. Family history includes consanguinity (cons.), with level of consanguinity noted (first cousin, second cousin, or first cousin once removed indicated by 1st, 2nd, and 1st+1), and/or sibling with features of OXPHOS 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) was categorized as markedly deficient (↓↓), moderately deficient (↓), equivocal (~), or normal (nl). mtDNA quantity denotes percentage 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. na, not available; nd, not determined; M, male; F, female; DD, developmental delay; FTT, failure to thrive; IUGR, intrauterine growth restriction; LFT, liver function test; CSF, cerebrospinal fluid; URTI, upper respiratory tract infection; GI, gastrointestinal; del., deletion; dep., depletion.

| | ID | Sex | Age of onset (death) | Clinical presentation | Family history | Biochemical features | | | | | | Gene candidate(s) |
						Tissue(s)	I	II	III	IV	mtDNA quantity	
Complex I	P1	M	<1 month	Cardiomyopathy, primary lactic acidosis		fib	↓↓				nd	**ACAD9***
	P2	F	<1 week (3 weeks)	Hypotonia, lethargy, respiratory distress, metabolic acidosis, FTT		fib	↓↓		↓	↓	nd	**POLG***
	P3	F	<1 week (4 months)	Severe IUGR, FTT, recurrent episodes of metabolic acidosis		skm, fib	↓↓				nd	**NDUFB3**
	P4	F	In utero (in utero)	Multiple pterygium syndrome, severe fetal hydrops, IUGR, terminated at 21 weeks	cons. (1st)	liv fib	↓↓ ~				nd nd	**AKR1B15**
	P5	F	<1 week (1 year)	Lethargy, tachypnea, organomegaly, renal and liver failure		liv skm, hea	↓↓ ↓↓				49% nd	—
	P6	M	<1 week	Bilateral optic nerve hypoplasia, lethargy, FTT		skm, liv, fib	↓				nd	—
	P7	M	<6 months (28 months)	Cystic leukodystrophy, FTT, severe metabolic acidosis		skm, liv, fib	↓↓				nd	—
	P8	F	<3 months (8 years)	Leukodystrophy, DD, abnormal LFTs, vasculitic rash		skm fib liv	↓↓ ↓↓ ~				nd nd nd	—
	P9	F	In utero (4 days)	Leukodystrophy, IUGR, metabolic acidosis, seizures		liv fib, skm	↓↓ ↓↓				460% nd	—
	P10	F	<1 year (na)	FTT, DD, seizures, hypotonia, prominent extra-axial CSF space		skm	↓				nd	—
Complex III	P11	F	<1 week	Severe global DD, tachypnea, metabolic acidosis, renal tubular acidosis, FTT		skm, liv fib			↓↓ nl		47% nd	**BCS1L***
	P12	M	<1 week (na)	Severe IUGR, lethargy, metabolic acidosis, renal tubular acidosis, dysmorphic facies	cons. (1st).	skm fib	↓ ~		↓↓ ↓↓		nd nd	TYMP*, MTCH1, C6orf125
	P13	M	<6 months	Hypotonia, FTT, DD, short stature, rickets, pruritus, dysmorphic facies	cons. (1st)	skm liv			↓↓ ~		nd nd	ACADSB

continued on next page

	ID	Sex	Age of onset (death)	Clinical presentation	Family history	Biochemical features						Gene candidate(s)
						Tissue(s)	I	II	III	IV	mtDNA quantity	
	P14	M	<1 week (na)	Fetal hypokinesia, Pierre Robin sequence, intra-abdominal calcification		skm, fib			↓↓		nd	UCP1, MTHFD1L
	P15	F	In utero (5 years)	Ventriculomegaly, apnea, developmental regression, hypotonia, seizures		skm fib	~ ↓		↓↓ ↓↓		nd nd	UQCR10
	P16	M	<1 week	Cardiorespiratory arrest, metabolic acidosis, renal tubular acidosis		fib			↓↓		nd	—
Complex IV	P17	M	<1 week	Neonatal lactic acidosis, cystic leukodystrophy		skm fib				↓↓ nl	nd nd	**COX6B1***
	P18	M	<1 week	Hypsarrhythmia, FTT, dystonia, squint		skm fib	~		~	↓↓ ↓	nd nd	**GFM1***, ACOX3
	P19	F	<1 month (12 months)	DD, seizure, hypotonia, does not fix and follow, cortical atrophy	cons. (1st+1)	fib				↓↓	nd	ACAD8
	P20	F	<1 year	Mild global DD, hypotonia, mild cerebral atrophy	sib.	skm fib				↓↓ nl	nd nd	C7orf10, MTERF
	P21	M	1 month (7 months)	Leigh syndrome, global DD, FTT, hypotonia, seizures	cons. (2nd)	fib				↓↓	nd	—
	P22	F	In utero	IUGR, oligohydramnios, severe global DD, seizures, microcephaly, hyperthermia	cons. (1st), sib.	skm, liv fib				↓↓ nl	nd nd	—
	P23	F	<1 month	Profound hypotonia, paucity of movement		skm				↓↓	37%	—
	P24	M	<1 year	DD, hypotonia, apnea, spastic diplegia, dysmorphic		skm fib				↓↓ ↓	84% nd	—
	P25	M	<6 months	Leigh-like syndrome, respiratory failure, lactic acidosis, RRF		skm				↓↓	nd	—
Combined OXPHOS deficiencies	P26	M	3 week	FTT, metabolic acidosis, hepatomegaly, apnea	cons. (1st)	skm liv fib	↓↓ ~ nl	↓↓ ~ nl	↓↓ ~ nl		nd nd nd	LYRM4
	P27	M	<1 week (2 months)	FTT, hypotonia, hypothermia, hypertrophic cardiomyopathy, hepatomegaly	cons. (1st), sib.	liv fib	↓↓ ↓		↓↓ ↓	↓↓ ↓	800% nd	**TSFM***
	P28	F	<1 month (1 month)	FTT, hypertonia, hypertrophic cardiomyopathy, respiratory arrest	cons. (1st), sib.	fib	↓		↓	↓	nd	**TSFM***
	P29	F	<1 year	DD, seizures, hypotonia, episodic metabolic acidosis		liv	↓↓			↓↓	167%	**GFM1***
	P30	M	<1 month	FTT, DD, sensorineural deafness, renal failure, dysmorphic	cons. (1st), sib.	skm fib	↓↓ ↓		↓↓ ↓	↓↓ ↓	42% nd	

continued on next page

	ID	Sex	Age of onset (death)	Clinical presentation	Family history	Biochemical features						Gene candidate(s)
						Tissue(s)	I	II	III	IV	mtDNA quantity	
	P31	M	<1 year	Hypertrophic cardiomyopathy		hea fib	↓↓ nl			↓↓ nl	nd nd	C1orf31
	P32	M	<6 months	Gastroesophageal reflux, FTT, hypotonia, respiratory distress, lactic acidosis		skm	↓↓		↓↓	↓↓	nd	EARS2
	P33	M	<1 year (20 months)	FTT, developmental regression, microcephaly		skm liv	↓↓ ~		↓↓ ↓↓	↓↓ ↓↓	290% nd	ALDH1B1, **mtDNA del.**
	P34	F	<1 week (4 months)	FTT, cardiac arrest, hypertrophic cardiomyopathy		skm fib liv	↓↓ ↓ ~		↓↓ ↓ nl	↓↓ ~ nl	88% nd nd	—
	P35	F	<1 week (10 days)	Metabolic acidosis, cardiac failure, hemorrhagic brain infarct		liv fib	↓↓ nl			↓↓ nl	190% nd	—
	P36	M	<1 year	Sudden-onset encephalopathy with seizures after URTI		skm liv	↓↓ ~		~ ~	↓↓ ↓	92% nd	—
	P37	F	<1 year	DD, ptosis, microcephaly, GI dysmotility		skm liv	↓↓ nl			↓↓ nl	53% nd	—
	P38	F	<1 year	Leukodystrophy, speech delay, dysesthesia in hands, developmental regression		skm fib	↓↓ nl		~ nl	↓↓ nl	112% nd	—
	P39	M	<2 years	Leigh syndrome, global DD, deterioration after URTI		skm fib	↓ ↓		↓ ↓	↓ ~	nd nd	—
mtDNA dep.	P40	F	In utero (stillborn)	Stillborn with hypotonia and multiple fractures		skm fib	↓↓ nl		↓↓ nl	↓↓ nl	8% nd	**AARS2***
	P41	F	<1 year (18 years)	Cardioskeletal myopathy, cataracts, FTT, fatigue		skm	↓↓		↓↓	↓↓	4%	**AGK**
	P42	F	<1 week (4 days)	Marked respiratory distress, pulmonary hypertension, cataracts	cons. (1st)	skm liv	↓↓ nl		↓↓ nl	↓↓ nl	13% 88%	**AGK**

*Previously reported OXPHOS disease genes.

Enrichment of prioritized genes in patients

We note that the fivefold 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 nonconsanguineous cases versus controls ($P = 5 \times 10^{-11}$) (Fig. 1C). This corresponded to a marked 23.9-fold enrichment ($P = 2 \times 10^{-8}$) specifically within the 77 established disease-related genes (8) and 5.2-fold enrichment ($P = 2 \times 10^{-7}$) within the remaining 957 mitochondrial genes (Fig. 1C). Thus, the 42 infantile patients are enriched for prioritized variants in both known pathogenic genes and 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 patient (2%) 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). Because there was about fivefold enrichment in patients with prioritized genes compared to controls, we can estimate that about 18 of the 42 patients (43%) will contain causal variants in prioritized genes. Determining which of the prioritized DNA variants are actually causal will require additional experimental evidence.

Molecular diagnoses in the mtDNA

Through de novo assembly of the mtDNA (Materials and Methods), we discovered a 7.2-kb 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 after a sudden metabolic and neurological decompensation. Muscle biopsy showed ragged red fibers and reduced

cytochrome c oxidase (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 polymerase chain reaction (qPCR) to be at 94% heteroplasmy in genomic DNA from affected muscle tissue. Because of mtDNA proliferation (288% compared to control), the actual amount of nondeleted mtDNA present in muscle of this patient was 18% of control. The deletion was confirmed by long-range PCR (LR-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), *GFM1* (27), *TSFM* (28), *AARS2* (29), and *TYMP* (30) (Fig. 2A). All such variants were confirmed through Sanger sequencing. We established compound heterozygosity via sequencing familial DNA, complementary DNA (cDNA), cloning, or molecular haplotyping (31) because short sequence reads do not typically provide the ability to determine whether variants are in cis or trans.

With additional experiments, we established a firm genetic diagnosis for nine of these patients (Table 3). These methods included the demon-

Table 2. MitoExome sequencing statistics. NA, not available.

| Targeted DNA	Nuclear DNA*	mtDNA*
Number of gene loci	1,034	37
Target size (bp)	1,392,193	16,569
Coverage[†]		
Mean coverage	142 (72–257)	25,457 (1,477–87,135)
% Target bp covered ≥1×	96 (91–98)	100 (100)
% Target bp covered ≥10×	89 (76–94)	100 (100)
% Target bp covered ≥15×	87 (71–92)	100 (100)
Number of variants compared to reference DNA	773 (571–912)	36 (11–50)
Rare	29 (17–56)	3 (0–23)
Rare and protein-modifying	17 (9–33)	0 (0–4)
Number of genes with two rare, protein-modifying alleles	1 (0–3)	NA

*Median values per patient, with range across patients indicated in parentheses.   [†]Excluding reads with ambiguous alignment.
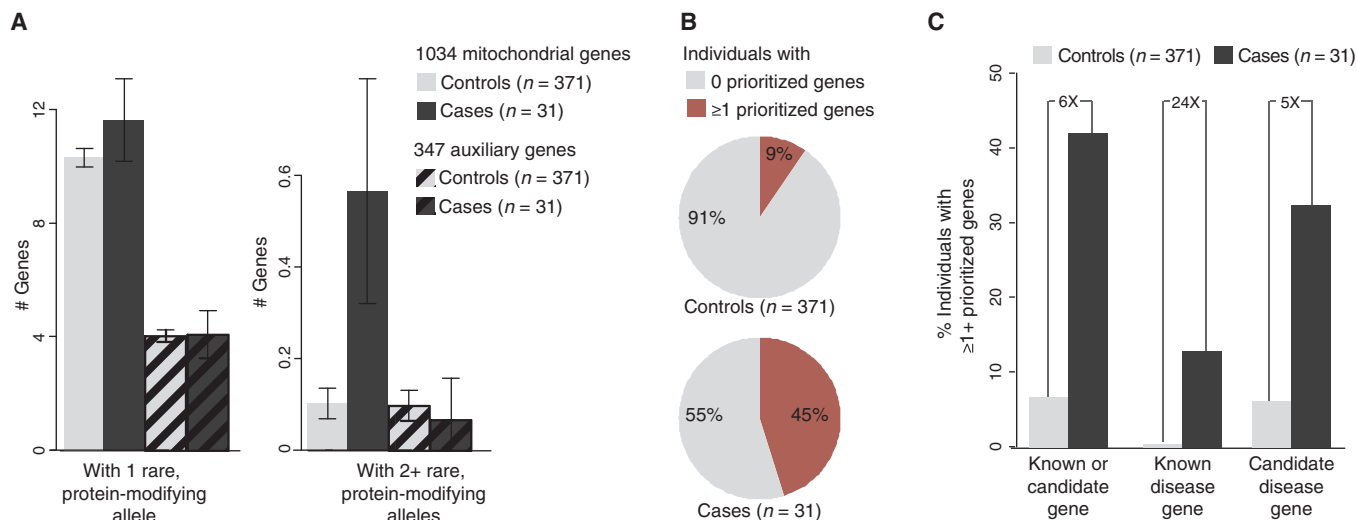
Fig. 1. Enrichment of prioritized variants in cases compared to healthy individuals. (**A**) Mean number of genes containing rare, protein-modifying alleles in cases and healthy controls within the 1034 mitochondrial genes and the 347 auxiliary genes sequenced. Error bars indicate SE. (**B**) Percent of cases and controls containing prioritized mitochondrial genes (red), defined as harboring two rare, protein-modifying alleles. (**C**) Percent of cases and controls containing prioritized mitochondrial genes, excluding variants with allele frequency of >0.005 in either cases or controls, with enrichment in cases versus controls displayed above. Analysis was restricted to regions with sequence coverage in all individuals and to the 31 cases from nonconsanguineous families.
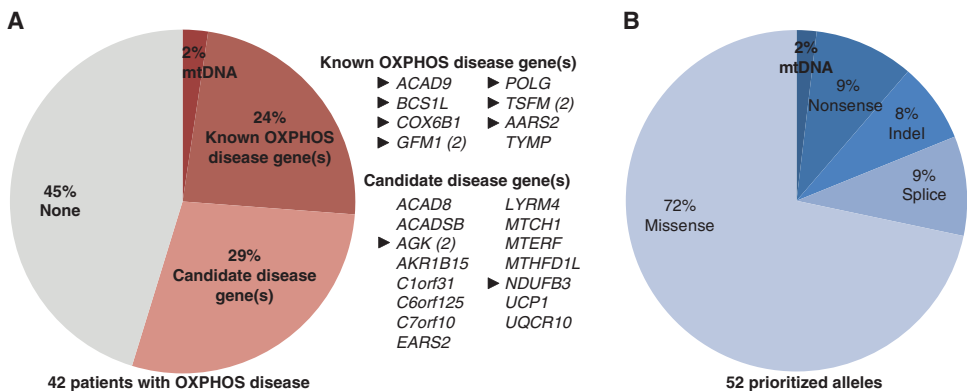
Fig. 2. Prioritized variants in 42 patients with OXPHOS disease. (**A**) Forty-two patients categorized by the presence of prioritized genes (red). The gene names are listed alphabetically at right, with parentheses indicating genes prioritized in two unrelated patients, and triangle indicating support of pathogenicity. (**B**) Fifty-two prioritized alleles, categorized by type of protein modification.

stration 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 Materials and Methods). The enzyme ACAD9 (*21*, *22*), recently linked to complex I deficiency [Mendelian Inheritance in Man (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 BCS1L (*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 new 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 BCS1L. In the remaining 10th 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

About 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 new, 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
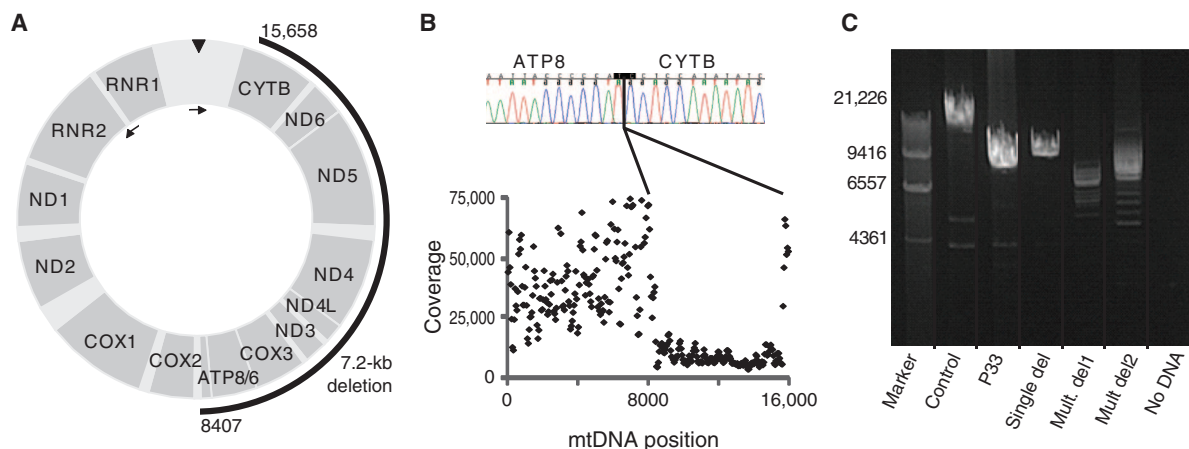
Fig. 3. mtDNA deletion in patient P33. (**A**) Schematic diagram of mtDNA indicates deletion (black arc) relative to position 0/16,569 (black triangle). Arrows indicate LR-PCR primers. (**B**) mtDNA sequence coverage in 50-bp windows. Inset shows Sanger electropherogram of breakpoint. (**C**) Gel electrophoresis of LR-PCR amplicon displays a 14,939-bp fragment in control DNA and 7681-bp fragment in P33, along with marker II (Roche) and mtDNA from individuals with confirmed single and multiple deletions.

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 new disease genes, as described below.

AGK mutations in two patients with myopathic mtDNA depletion

One new candidate gene, acylglycerol kinase (*AGK*), had recessive mutations in two unrelated patients. This protein [also known as multi-substrate 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, and IV deficiency, bilateral cataracts, and severe mtDNA depletion in skeletal muscle (4 and 13% of normal). Detailed clinical histories are available in the Supplementary Material. 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 three 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

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)	*ACAD9*[†]	Compound Heterozygous	c.976G>C, p.A326P (*22*) c.260T>A, p.I87N	1 3	Conservation, protein degradation
P2	Complex I deficiency (MIM 252010)	*POLG*[†]	Compound Heterozygous	c.2542G>A, p.G848S (*69*) c.679C>T, p.R227W (*70*)	1 1	
P3	Complex I deficiency (MIM 252010)	*NDUFB3*	Homozygous	c.64T>C, p.W22R	3	Mother carrier, no deletion or isoparental disomy detected by Affy2.7M and Illumina 300K arrays, protein degradation, conservation, correction studies
P11	Complex III deficiency (MIM 124000)	*BCS1L*[†]	Compound Heterozygous	c.166C>T, p.R56X (*71*) c.268C>T, p.R90C	1 3	Conservation, mRNA and protein degradation
P17	Complex IV deficiency (MIM 220110)	*COX6B1*[†]	Homozygous	c.241A>C, p.T81P	3	Protein degradation, family segregation
P18	Complex IV deficiency	*GFM1*[†]	Compound Heterozygous	c.720delT, p.E241NfsX1 c.2011C>T, p.R671C	2 3	mRNA degradation, protein degradation, family segregation
P29	Combined deficiency (MIM 609060)	*GFM1*[†]	Compound Heterozygous	c.720delT, p.E241NfsX1 c.910A>G, p.K304E	2 3	mRNA degradation, protein degradation, maternal carrier c.720delT
P27	Combined deficiency (MIM 610505)	*TSFM*[†]	Homozygous	c.997C>T, p.R333W (*28*)	1	Family segregation
P28	Combined deficiency (MIM 610505)	*TSFM*[†]	Homozygous	c.997C>T, p.R333W (*28*)	1	Family segregation
P33	Combined deficiency	**mtDNA deletion**	94% heteroplasmic mutant load	m.8407_15658del7252 (*20*)	1	qPCR, LR-PCR, and sequencing
P40	mtDNA depletion	*AARS2*[†]	Compound Heterozygous	c.986G>A, p.R329H c.1774C>T, p.R592W (*29*)	3 1	Conservation, protein degradation
P41	Myopathic mtDNA depletion (MIM 609560)	*AGK*	Compound Heterozygous	c.297+2T>C, p.K75QfsX12 c.1170T>A, p.Y390X	2 2	Independent mutations in two patients
P42	Myopathic mtDNA depletion (MIM 609560)	*AGK*	Homozygous	c.1131+1G>T, p.S350EfsX19	2	Independent mutations in two patients, family segregation

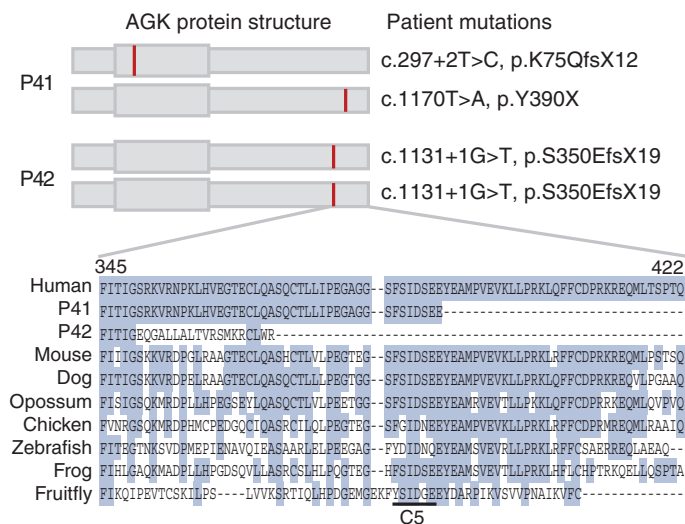*Classification of pathogenicity for each mutation using the ACMG categories (*42*).   †Genes previously associated with OXPHOS disease.

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 because 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 new 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 nonconsanguineous 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 of 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 single-nucleotide polymorphism (SNP) arrays. No large deletions were evident, and isodisomy of chromosome 2 was excluded on the basis of the SNPduo identity by state algorithm (38). However, the CNVPartition v3.1.6 algorithm (Illumina) detected a 1.3-Mb 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 or 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 patient with complex I deficiency, who had pathogenic mutations in *C8orf38* (14) and vice versa (Fig. 5), establishing *NDUFB3* as the causal gene in this case. This is the 15th 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. Notably, we did not know what percent of unsolved OXPHOS cases would be expected to be due to mutations in established loci, because, 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 one mtDNA deletion and nine 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. Given that such variants exhibit fivefold 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 new link between the gene *AGK* and myopathic mtDNA depletion through an unknown mechanism.

Recessive mutations in 13 additional genes, not 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 a similar phenotype. We estimate that 6 of these 13 genes will prove to be bona fide disease genes. These candidates are particularly exciting because 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 (for example, *UQCR10*, *LYRM4*, and *EARS2*), whereas most of the candidates encode poorly characterized proteins that have never before been linked to OXPHOS and will reveal fundamentally new biochemical insights (for example, *C1orf31*, *C6orf125*, and *AKR1B15*).

This pilot study showed that about half of the 42 sequenced patients lacked "smoking gun" prioritized variants. We can envision at least five

Fig. 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 and a box indicating the diacylglycerol kinase catalytic domain. Conservation of C-terminal region is shown below, with identical residues shaded and the C5 domain shared with yeast sphingosine kinases underlined.

possible explanations. First, we may have missed the causal variants because of a technical lack of sensitivity. Although we detected more than 90% of SNVs present in the MitoExome, there is an unknown sensitivity for indels and exon deletions because of a lack of training data. Second, the causal variant may have been located in a gene that we did not target with MitoExome sequencing. Although possible, this explanation seems less likely because 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 nontargeted intronic or regulatory region. Although these are 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 because, 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 6 actually appear to explain the phenotype, whereas 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 because most patients had been screened for common mutations in mtDNA or in relevant genes suggested by phenotype (for example, *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 fig. S11 and Supplementary Material).
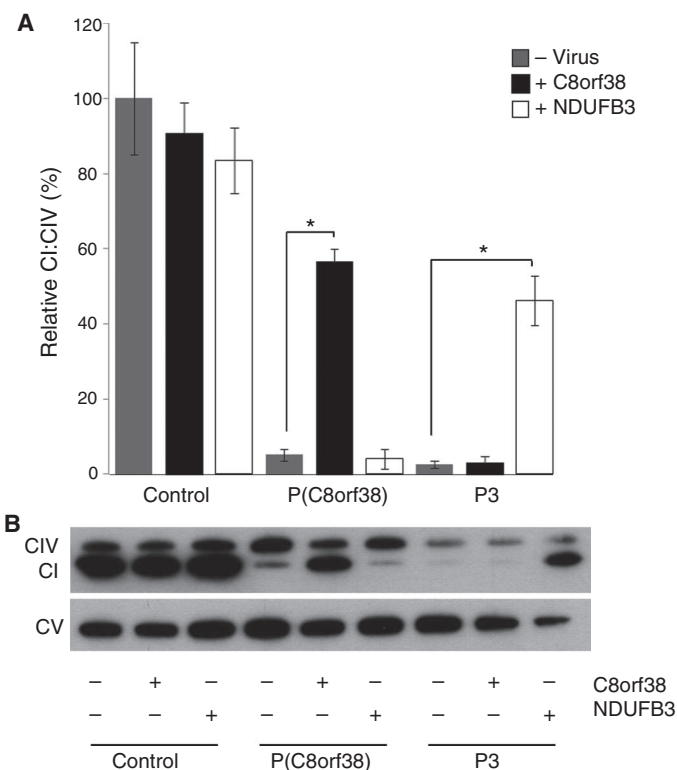
Three advances will greatly aid interpretation of DNA variation for routine clinical diagnosis. First, NGS studies of Mendelian families will rapidly expand the set of validated OXPHOS disease–related genes from 77 to perhaps 200 nuclear loci. Second, NGS studies of ethnically diverse, healthy individuals 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 (*44*, *45*). Third, whereas 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 to detect de novo mutations and to confirm the mode of transmission.

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 or tissue sample, can accelerate clinical diagnosis and enable genetic counseling where appropriate. Genetic diagnosis will also enable rational subclassification of disease, which may help to 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 of greatest use when integrated with the broader clinical and biochemical presentation.

Fig. 5. *NDUFB3* complementation of complex I defects in subject P3 fibroblasts. (**A**) Bar plots show complex I (CI) activity, normalized by complex IV (CIV) activity, in fibroblasts from patient P3, a healthy control individual, and an unrelated patient with complex I deficiency because of a *C8orf38* mutation P(*C8orf38*). Enzyme activity was measured before and after transduction with wild-type *C8orf38* or *NDUFB3* mRNA. Data are means of three biological replicates ± SEM. *$P < 0.001$, Student's *t* test. (**B**) Western blot shows protein expression of complex I, complex IV, and complex V (CV) as the loading control.

MATERIALS AND METHODS

Patients

On the basis of clinical and biochemical criteria from established diagnostic algorithms (*46*), we selected 42 patients with definite OXPHOS

disease 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 1 was classified as marked (<25%), moderate (26 to 40%), or equivocal (41 to 60%), corresponding to residual activities of normal control mean relative to citrate synthase or complex II. The inclusion criterion was infantile age of onset, defined for this study as under 2 years of age. The exclusion criterion was 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 4.1 Mb of targeted DNA included the 16.6-kb 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 University of California Santa Cruz (UCSC) known gene (48) collections were downloaded from the UCSC genome browser (49) assembly hg18 (February 2009). All analyses presented here, except 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 (1.4 Mb), because no significant results were found in the untranslated regions (2.3 Mb) or in the coding regions of 347 auxiliary genes sequenced (0.4 Mb).

Hybrid selection and Illumina sequencing

An in-solution hybridization capture method (50) was used to isolate targeted DNA, which was sequenced on the Illumina GAII platform (51) with paired 76–base pair (bp) reads. Single 120-bp baits were synthesized (Agilent) for each target region or tiled across targets that exceeded 120 bp. A total of 42,923 baits were synthesized to cover 13,803 target regions. Each subject sample was whole genome–amplified with a Qiagen REPLI-g Kit with 100 ng of 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 and 350 bp. 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 and then eluted (50, 52). Each patient or HapMap sample was sequenced in one lane of an Illumina GAII instrument. NIMH control samples were subjected to whole-exome hybrid selection with the same protocol and were sequenced in two or three lanes of an Illumina GAII instrument with paired 76-bp reads.

Sequence alignment, variant detection, and annotation

Illumina reads were aligned to the GRCh37 reference human genome assembly with the BWA (53) algorithm in SAMtools (2) within the Picard analysis pipeline (http://picard.sourceforge.net/). Illumina quality scores were recalibrated and realigned to GRCh37, duplicates were removed, and the alignment was modified to account for indels with the GATK software package (54) version v1.0.5159. SNVs were detected and genotyped with the GATK UnifiedGenotyper in single-sample mode (with parameters -im ALL -mbq 20 -mmq 20 -mm42 3 -deletions 0.05). Variants were filtered with GATK VariantFiltration module (with filters "QUAL<50.0 & QD<5.0 & HRun>10 & DP<4" and parameters -cluster 3 -window 10). Indels were detected with 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 number of mismatches across reads supporting the indel) or max_cons_nqs_av_mm ≥0.2 (maximum average mismatch rate in the 5-bp 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 with 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).

Eight hundred and forty-seven 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 before 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 >100× copies of mtDNA molecules compared to nuclear DNA molecules. To avoid alignment artifacts caused by mtDNA reads being improperly assigned to NUMTs, we separately aligned all Illumina reads 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 ~25,000× coverage) rather than excluded (resulting in ~1000× coverage) because 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 with the GATK UnifiedGenotyper version v1.0.2986 in pooled-sample mode 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, because it is difficult to resolve low-heteroplasmy variants from variants present in NUMTs (19). mtDNA variants were annotated with data from MITOMAP (18), and allele frequency was obtained from mtDB (45).

mtDNA deletions or rearrangements were detected through de novo mtDNA assembly with a modified ALLPATHS assembly approach (59). Reads aligned to the mtDNA were randomly down-sampled to 1000× coverage for computational tractability. Errors in reads were corrected with 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, and NA12892) were estimated with 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 that 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 nonreference 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). Because 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 with 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 with the MitoExome analysis pipeline. Prevalence comparisons within the 1034 mitochondrial genes were restricted to the 6872 autosomal coding exons that were well measured in both study designs, totaling 1 Mb (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, we restricted comparisons within the 347 auxiliary genes to the 2619 autosomal coding exons, totaling 381 kb (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. 1, B and C). Details of variant prevalence in cases and controls are shown in fig. S1. For patients, we conservatively excluded heterozygous variants shown to be in cis based on haplotype phasing (1 of 11 cases) as well as variants not validated by Sanger sequencing (1 of 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 of 49 variants validated), and compound heterozygous variants were phased through sequencing cDNA (GFM1), cloned DNA (BCS1L, C1orf31, TYMP, MTHFD1L), familial DNA (GFM1, AGK, EARS2), or by a molecular haplotyping approach (31) (ACAD9, AARS2, POLG) described in Supplementary Methods. Ten of 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 LR-PCR with primers 2F (m.1650_1671) and D1R (m.19_1) and Sanger sequencing with 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 (figs. S2 to S8). Presence and segregation of alleles in family members (when available) were assayed by Sanger sequencing. The effect of mutations on mature RNA transcript splicing and abundance was measured in patient and control fibroblasts by reverse transcription–PCR (RT-PCR) in the presence and absence of cycloheximide (100 ng/μl) for 24 hours to inhibit nonsense-mediated decay (65). Relative abundance of full-length protein products and OXPHOS subunits or assembly factors from all five complexes was measured in patient and control fibroblast lines and skeletal muscle via SDS–polyacrylamide gel electrophoresis (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 decent (P12, P13, P19, P21, P22, P26, P27, P28, and P30) were screened with 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 of 168 Lebanese chromosomes (0.012 allele frequency). 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 with the Illumina HumanCytoSNP-12 array (version 2.1) as previously described (*67*). Automated LCSH detection was performed with the CNVPartition v3.1.6 algorithm in KaryoStudio software. SNP genotypes were generated in GenomeStudio software (Illumina) with data from a set of 102 intra-run samples. Both the proband and the maternal DNA samples yielded an SNP call rate of 99.5%. Identity by state (IBS) analysis, assessing the sharing of alleles between individuals, was performed with the online SNPduo program (*38*) (http://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*) with 5′ Bci I and 3′ Xba I. The pF-5xUAS-*C8orf38*-SV40 puroGEV16-W vector was described previously (*72*). To generate lentiviral particles, we grew human embryonic kidney (HEK) 293T cells on 10-cm plates to 60% confluence and cotransfected them 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 hours after transfection, and after another 24 hours of 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 polybrene (5 µg/ml) (Sigma) for 48 hours. Cells were grown in antibiotic-free medium for 30 hours before application of selection medium containing puromycin (1 µg/ml). We found that our lentivectors were "leaky" in primary human fibroblast lines and that sufficient expression of C8orf38 and NDUFB3 was achieved without tamoxifen induction. After 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 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 with Bonferroni method to determine statistically significant differences.

SUPPLEMENTARY MATERIAL

www.sciencetranslationalmedicine.org/cgi/content/full/4/118/118ra10/DC1
Materials and Methods
Fig. S1. Variant prevalence in cases and controls.
Fig. S2. *ACAD9* variants in patient P1.
Fig. S3. *POLG* mutations in P2.
Fig. S4. *BCS1L* mutations in P11.
Fig. S5. *COX6B1* mutations in P17.
Fig. S6. *GFM1* mutations in P18 and P29.
Fig. S7. *TSFM* mutations in P27 and P28.
Fig. S8. *AARS2* mutations in P40.
Fig. S9. *AGK* mutations in P41 and P42.
Fig. S10. *NDUFB3* mutation in P3.
Fig. S11. Estimated MitoExome diagnostic efficacy in a retrospective cohort.
Table S1. List of MitoExome regions targeted for sequencing.
Table S2. Detailed list of all detected variants, all prioritized variants, and all confirmed pathogenic variants within the mtDNA or coding exons of 1034 mitochondrial genes.
References

REFERENCES AND NOTES

1. M. Choi, U. I. Scholl, W. Ji, T. Liu, I. R. Tikhonova, P. Zumbo, A. Nayir, A. Bakkaloğlu, S. Özen, S. Sanjad, C. Nelson-Williams, A. Farhi, S. Mane, R. P. Lifton, Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19096–19101 (2009).
2. S. B. Ng, E. H. Turner, P. D. Robertson, S. D. Flygare, A. W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E. E. Eichler, M. Bamshad, D. A. Nickerson, J. Shendure, Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
3. D. Skladal, J. Halliday, D. R. Thorburn, Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain* **126**, 1905–1912 (2003).
4. S. DiMauro, M. Hirano, E. A. Schon, *Mitochondrial Medicine* (Informa Healthcare, New York, 2006), p. 368.
5. S. Dimauro, G. Davidzon, Mitochondrial DNA and disease. *Ann. Med.* **37**, 222–232 (2005).
6. P. F. Chinnery, D. M. Turnbull, Epidemiology and treatment of mitochondrial disorders. *Am. J. Med. Genet.* **106**, 94–101 (2001).
7. D. M. Kirby, D. R. Thorburn, Approaches to finding the molecular basis of mitochondrial oxidative phosphorylation disorders. *Twin Res. Hum. Genet.* **11**, 395–411 (2008).
8. E. J. Tucker, A. G. Compton, D. R. Thorburn, Recent advances in the genetics of mitochondrial encephalopathies. *Curr. Neurol. Neurosci. Rep.* **10**, 277–285 (2010).
9. Medical Genetics Laboratories, Baylor College of Medicine (2011).
10. MEDomics, www.medomics.com/MitoNucleomeDx [accessed 30 September 2011].
11. GeneDX, *Next-Generation Sequence Analysis of 24 Nuclear Genes Important for Normal Mitochondrial Function*, www.genedx.com/pdf_files/Nuclear_panel_24_gene_descriptions.pdf [accessed 30 September 2011].
12. C. Scharfe, H. H. Lu, J. K. Neuenburg, E. A. Allen, G. C. Li, T. Klopstock, T. M. Cowan, G. M. Enns, R. W. Davis, Mapping gene associations in human mitochondria using clinical disease phenotypes. *PLoS Comput. Biol.* **5**, e1000374 (2009).
13. S. E. Calvo, E. J. Tucker, A. G. Compton, D. M. Kirby, G. Crawford, N. P. Burtt, M. Rivas, C. Guiducci, D. L. Bruno, O. A. Goldberger, M. C. Redman, E. Wiltshire, C. J. Wilson, D. Altshuler, S. B. Gabriel, M. J. Daly, D. R. Thorburn, V. K. Mootha, High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat. Genet.* **42**, 851–858 (2010).
14. D. J. Pagliarini, S. E. Calvo, B. Chang, S. A. Sheth, S. B. Vafai, S. E. Ong, G. A. Walford, C. Sugiana, A. Boneh, W. K. Chen, D. E. Hill, M. Vidal, J. G. Evans, D. R. Thorburn, S. A. Carr, V. K. Mootha, A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112–123 (2008).
15. P. Shen, W. Wang, S. Krishnakumar, C. Palm, A. K. Chi, G. M. Enns, R. W. Davis, T. P. Speed, M. N. Mindrinos, C. Scharfe, High-quality DNA sequence capture of 524 disease candidate genes. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6549–6554 (2011).
16. V. Vasta, S. B. Ng, E. H. Turner, J. Shendure, S. H. Hahn, Next generation sequence analysis for mitochondrial disorders. *Genome Med.* **1**, 100 (2009).
17. The International HapMap 3 Consortium, D. M. Altshuler, R. A. Gibbs, L. Peltonen, D. M. Altshuler, R. A. Gibbs, L. Peltonen, E. Dermitzakis, S. F. Schaffner, F. Yu, L. Peltonen, E. Dermitzakis, P. E. Bonnen, D. M. Altshuler, R. A. Gibbs, P. I. de Bakker, P. Deloukas, S. B. Gabriel, R. Gwilliam, S. Hunt, M. Inouye, X. Jia, A. Palotie, M. Parkin, P. Whittaker, F. Yu, K. Chang, A. Hawes, L. R. Lewis, Y. Ren, D. Wheeler, R. A. Gibbs, D. M. Muzny, C. Barnes, K. Darvishi, M. Hurles, J. M. Korn, K. Kristiansson, C. Lee, S. A. McCarrol, J. Nemesh, E. Dermitzakis, A. Keinan, S. B. Montgomery, S. Pollack, A. L. Price, N. Soranzo, P. E. Bonnen, R. A. Gibbs, C. Gonzaga-Jauregui, A. Keinan, A. L. Price, F. Yu, V. Anttila, W. Brodeur, M. J. Daly, S. Leslie, G. McVean, L. Moutsianas, H. Nguyen, S. F. Schaffner, Q. Zhang, M. J. Ghori, R. McGinnis, W. McLaren, S. Pollack, A. L. Price, S. F. Schaffner, F. Takeuchi, S. R. Grossman, I. Shlyakhter, E. B. Hostetter, P. C. Sabeti, C. A. Adebamowo, M. W. Foster, D. R. Gordon, J. Licinio, M. C. Manca, P. A. Marshall, I. Matsuda, D. Ngare, V. O. Wang, D. Reddy, C. N. Rotimi, C. D. Royal, R. R. Sharp, C. Zeng, L. D. Brooks, J. E. McEwen, Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).

18. E. Ruiz-Pesini, M. T. Lott, V. Procaccio, J. C. Poole, M. C. Brandon, D. Mishmar, C. Yi, J. Kreuziger, P. Baldi, D. C. Wallace, An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res.* **35**, D823–D828 (2007).

19. M. Li, A. Schonberg, M. Schaefer, R. Schroeder, I. Nasidze, M. Stoneking, Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am. J. Hum. Genet.* **87**, 237–249 (2010).

20. L. Hirt, P. J. Magistretti, L. Hirt, J. Bogousslavsky, O. Boulat, F. X. Borruat, Large deletion (7.2 kb) of mitochondrial DNA with novel boundaries in a case of progressive external ophthalmoplegia. *J. Neurol. Neurosurg. Psychiatry* **61**, 422–423 (1996).

21. J. Nouws, L. Nijtmans, S. M. Houten, M. van den Brand, M. Huynen, H. Venselaar, S. Hoefs, J. Gloerich, J. Kronick, T. Hutchin, P. Willems, R. Rodenburg, R. Wanders, L. van den Heuvel, J. Smeitink, R. O. Vogel, Acyl-CoA dehydrogenase 9 is required for the biogenesis of oxidative phosphorylation complex I. *Cell Metab.* **12**, 283–294 (2010).

22. T. B. Haack, K. Danhauser, B. Haberberger, J. Hoser, V. Strecker, D. Boehm, G. Uziel, E. Lamantea, F. Invernizzi, J. Poulton, B. Rolinski, A. Iuso, S. Biskup, T. Schmidt, H. W. Mewes, I. Wittig, T. Meitinger, M. Zeviani, H. Prokisch, Exome sequencing identifies *ACAD9* mutations as a cause of complex I deficiency. *Nat. Genet.* **42**, 1131–1134 (2010).

23. J. D. Stumpf, W. C. Copeland, Mitochondrial DNA replication and disease: Insights from DNA polymerase γ mutations. *Cell. Mol. Life Sci.* **68**, 219–233 (2011).

24. G. Van Goethem, B. Dermaut, A. Löfgren, J. J. Martin, C. Van Broeckhoven, Mutation of *POLG* is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. *Nat. Genet.* **28**, 211–212 (2001).

25. P. de Lonlay, I. Valnot, A. Barrientos, M. Gorbatyuk, A. Tzagoloff, J. W. Taanman, E. Benayoun, D. Chrétien, N. Kadhom, A. Lombès, H. O. de Baulny, P. Niaudet, A. Munnich, P. Rustin, A. Rötig, A mutant mitochondrial respiratory chain assembly protein causes complex III deficiency in patients with tubulopathy, encephalopathy and liver failure. *Nat. Genet.* **29**, 57–60 (2001).

26. V. Massa, E. Fernandez-Vizarra, S. Alshahwan, E. Bakhsh, P. Goffrini, I. Ferrero, P. Mereghetti, P. D'Adamo, P. Gasparini, M. Zeviani, Severe infantile encephalomyopathy caused by a mutation in COX6B1, a nucleus-encoded subunit of cytochrome c oxidase. *Am. J. Hum. Genet.* **82**, 1281–1289 (2008).

27. M. J. Coenen, H. Antonicka, C. Ugalde, F. Sasarman, R. Rossi, J. G. Heister, R. F. Newbold, F. J. Trijbels, L. P. van den Heuvel, E. A. Shoubridge, J. A. Smeitink, Mutant mitochondrial elongation factor G1 and combined oxidative phosphorylation deficiency. *N. Engl. J. Med.* **351**, 2080–2086 (2004).

28. J. A. Smeitink, O. Elpeleg, H. Antonicka, H. Diepstra, A. Saada, P. Smits, F. Sasarman, G. Vriend, J. Jacob-Hirsch, A. Shaag, G. Rechavi, B. Welling, J. Horst, R. J. Rodenburg, B. van den Heuvel, E. A. Shoubridge, Distinct clinical phenotypes associated with a mutation in the mitochondrial translation elongation factor EFTs. *Am. J. Hum. Genet.* **79**, 869–877 (2006).

29. A. Götz, H. Tyynismaa, L. Euro, P. Ellonen, T. Hyötyläinen, T. Ojala, R. H. Hämäläinen, J. Tommiska, T. Raivio, M. Oresic, R. Karikoski, O. Tammela, K. O. Simola, A. Paetau, T. Tyni, A. Suomalainen, Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. *Am. J. Hum. Genet.* **88**, 635–642 (2011).

30. I. Nishino, A. Spinazzola, M. Hirano, Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science* **283**, 689–692 (1999).

31. B. A. Konfortov, A. T. Bankier, P. H. Dear, An efficient method for multi-locus molecular haplotyping. *Nucleic Acids Res.* **35**, e6 (2007).

32. S. B. Ng, K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. A. Nickerson, J. Shendure, M. J. Bamshad, Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).

33. S. B. Ng, A. W. Bigham, K. J. Buckingham, M. C. Hannibal, M. J. McMillin, H. I. Gildersleeve, A. E. Beck, H. K. Tabor, G. M. Cooper, H. C. Mefford, C. Lee, E. H. Turner, J. D. Smith, M. J. Rieder, K. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D. A. Nickerson, M. J. Bamshad, J. Shendure, Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–793 (2010).

34. M. P. King, Y. Koga, M. Davidson, E. A. Schon, 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.* **12**, 480–490 (1992).

35. M. Bektas, S. G. Payne, H. Liu, S. Goparaju, S. Milstien, S. Spiegel, A novel acylglycerol kinase that produces lysophosphatidic acid modulates cross talk with EGFR in prostate cancer cells. *J. Cell Biol.* **169**, 801–811 (2005).

36. A. Spinazzola, F. Invernizzi, F. Carrara, E. Lamantea, A. Donati, M. Dirocco, I. Giordano, M. Meznaric-Petrusa, E. Baruffini, I. Ferrero, M. Zeviani, Clinical and molecular features of mitochondrial DNA depletion syndromes. *J. Inherit. Metab. Dis.* **32**, 143–158 (2009).

37. A. Saada, A. Shaag, H. Mandel, Y. Nevo, S. Eriksson, O. Elpeleg, Mutant mitochondrial thymidine kinase in mitochondrial DNA depletion myopathy. *Nat. Genet.* **29**, 342–344 (2001).

38. E. D. Roberson, J. Pevsner, Visualization of shared genomic regions and meiotic recombination in high-density SNP data. *PLoS One* **4**, e6711 (2009).

39. N. A. Zaghloul, Y. Liu, J. M. Gerdes, C. Gascue, E. C. Oh, C. C. Leitch, Y. Bromberg, J. Binkley, R. L. Leibel, A. Sidow, J. L. Badano, N. Katsanis, Functional analyses of variants reveal a

40. significant role for dominant negative and common alleles in oligogenic Bardet–Biedl syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10602–10607 (2010).

40. H. A. Tuppen, F. Fattori, R. Carrozzo, M. Zeviani, S. DiMauro, S. Seneca, J. E. Martindale, S. E. Olpin, E. P. Treacy, R. McFarland, F. M. Santorelli, R. W. Taylor, Further pitfalls in the diagnosis of mtDNA mutations: Homoplasmic mt-tRNA mutations. *J. Med. Genet.* **45**, 55–61 (2008).

41. B. J. O'Roak, P. Deriziotis, C. Lee, L. Vives, J. J. Schwartz, S. Girirajan, E. Karakoc, A. P. MacKenzie, S. B. Ng, C. Baker, M. J. Rieder, D. A. Nickerson, R. Bernier, S. E. Fisher, J. Shendure, E. E. Eichler, Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* **43**, 585–589 (2011).

42. C. S. Richards, S. Bale, D. B. Bellissimo, S. Das, W. W. Grody, M. R. Hegde, E. Lyon, B. E. Ward; Molecular Subcommittee of the ACMG Laboratory Quality Assurance Committee, ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet. Med.* **10**, 294–300 (2008).

43. C. J. Bell, D. L. Dinwiddie, N. A. Miller, S. L. Hateley, E. E. Ganusova, J. Mudge, R. J. Langley, L. Zhang, C. C. Lee, F. D. Schilkey, V. Sheth, J. E. Woodward, H. E. Peckham, G. P. Schroth, R. W. Kim, S. F. Kingsmore, Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med.* **3**, 65ra4 (2011).

44. D. R. Thorburn, J. Smeitink, Diagnosis of mitochondrial disorders: Clinical and biochemical approach. *J. Inherit. Metab. Dis.* **24**, 312–316 (2001).

45. M. Ingman, U. Gyllensten, mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res.* **34**, D749–D751 (2006).

46. F. P. Bernier, A. Boneh, X. Dennett, C. W. Chow, M. A. Cleary, D. R. Thorburn, Diagnostic criteria for respiratory chain disorders in adults and children. *Neurology* **59**, 1406–1411 (2002).

47. K. D. Pruitt, T. Tatusova, D. R. Maglott, NCBI reference sequences (RefSeq): A curated nonredundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65 (2007).

48. P. A. Fujita, B. Rhead, A. S. Zweig, A. S. Hinrichs, D. Karolchik, M. S. Cline, M. Goldman, G. P. Barber, H. Clawson, A. Coelho, M. Diekhans, T. R. Dreszer, B. M. Giardine, R. A. Harte, J. Hillman-Jackson, F. Hsu, V. Kirkup, R. M. Kuhn, K. Learned, C. H. Li, L. R. Meyer, A. Pohl, B. J. Raney, K. R. Rosenbloom, K. E. Smith, D. Haussler, W. J. Kent, The UCSC Genome Browser database: Update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2011).

49. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

50. A. Gnirke, A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander, C. Nusbaum, Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).

51. D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, M. C. E. Catenazzi, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. Huw Jones, G. D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. Chris Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy, V. H. Rae, S. R. Rawlings, A. Chiva Rodriguez, P. M. Roe, J. Rogers, M. C. Rogert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. Ernest Sohna Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. Vandevondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, A. J. Smith, Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).

52. S. Fisher, A. Barry, J. Abreu, B. Minie, J. Nolan, T. M. Delorey, G. Young, T. J. Fennell, A. Allen, L. Ambrogio, A. M. Berlin, B. Blumenstiel, K. Cibulskis, D. Friedrich, R. Johnson, F. Juhn,

B. Reilly, R. Shammas, J. Stalker, S. M. Sykes, J. Thompson, J. Walsh, A. Zimmer, Z. Zwirko, S. Gabriel, R. Nicol, C. Nusbaum, A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* **12**, R1 (2011).

53. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

54. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

55. P. D. Stenson, E. V. Ball, K. Howells, A. D. Phillips, M. Mort, D. N. Cooper, The Human Gene Mutation Database: Providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum. Genomics* **4**, 69–72 (2009).

56. S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, K. Sirotkin, dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

57. The 1000 Genomes Project Consortium, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

58. E. Hazkani-Covo, R. M. Zeller, W. Martin, Molecular poltergeists: Mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* **6**, e1000834 (2010).

59. I. Maccallum, D. Przybylski, S. Gnerre, J. Burton, I. Shlyakhter, A. Gnirke, J. Malek, K. McKernan, S. Ranade, T. P. Shea, L. Williams, S. Young, C. Nusbaum, D. B. Jaffe, ALLPATHS 2: Small genomes assembled accurately and with high continuity from short paired reads. *Genome Biol.* **10**, R103 (2009).

60. S. Gnerre, I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, S. Sykes, A. M. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander, D. B. Jaffe, High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1513–1518 (2011).

61. P. A. Pevzner, H. Tang, M. S. Waterman, An Eulerian path approach to DNA fragment assembly. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9748–9753 (2001).

62. P. D. Stenson, M. Mort, E. V. Ball, K. Howells, A. D. Phillips, N. S. Thomas, D. N. Cooper, The Human Gene Mutation Database: 2008 update. *Genome Med.* **1**, 13 (2009).

63. L. He, P. F. Chinnery, S. E. Durham, E. L. Blakely, T. M. Wardell, G. M. Borthwick, R. W. Taylor, D. M. Turnbull, Detection and quantification of mitochondrial DNA deletions in individual cells by real-time PCR. *Nucleic Acids Res.* **30**, e68 (2002).

64. A. T. Pagnamenta, J.-W. Taanman, C. J. Wilson, N. E. Anderson, R. Marotta, A. J. Duncan, M. Bitner-Glindzicz, R. W. Taylor, A. Laskowski, D. R. Thorburn, S. Rahman, Dominant inheritance of premature ovarian failure associated with mutant mitochondrial DNA polymerase gamma. *Hum. Reprod.* **21**, 2467–2473 (2006).

65. S. R. Lamandé, J. F. Bateman, W. Hutchison, R. J. McKinlay Gardner, S. P. Bower, E. Byrne, H. H. Dahl, Reduced collagen VI causes Bethlem myopathy: A heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency. *Hum. Mol. Genet.* **7**, 981–989 (1998).

66. S. T. Cooper, H. P. Lo, K. N. North, Single section Western blot: Improving the molecular diagnosis of the muscular dystrophies. *Neurology* **61**, 93–97 (2003).

67. D. L. Bruno, Z. Stark, D. J. Amor, T. Burgess, K. Butler, S. Corrie, D. Francis, D. Ganesamoorthy, L. Hills, P. A. James, D. O'Rielly, R. Oertel, R. Savarirayan, K. Prabhakara, N. Salce, H. R. Slater, Extending the scope of diagnostic chromosome analysis: Detection of single gene defects using high-resolution SNP microarrays. *Hum. Mutat.* **32**, 1500–1506 (2011).

68. Y. Y. Yeap, I. H. Ng, B. Badrian, T. V. Nguyen, Y. Y. Yip, A. S. Dhillon, S. E. Mutsaers, J. Silke, M. A. Bogoyevitch, D. C. Ng, c-Jun N-terminal kinase/c-Jun inhibits fibroblast proliferation by negatively regulating the levels of stathmin/oncoprotein 18. *Biochem. J.* **430**, 345–354 (2010).

69. E. Lamantea, V. Tiranti, A. Bordoni, A. Toscano, F. Bono, S. Servidei, A. Papadimitriou, H. Spelbrink, L. Silvestri, G. Casari, G. P. Comi, M. Zeviani, Mutations of mitochondrial DNA polymerase γA are a frequent cause of autosomal dominant or recessive progressive external ophthalmoplegia. *Ann. Neurol.* **52**, 211–219 (2002).

70. A. Agostino, L. Valletta, P. F. Chinnery, G. Ferrari, F. Carrara, R. W. Taylor, A. M. Schaefer, D. M. Turnbull, V. Tiranti, M. Zeviani, Mutations of ANT1, Twinkle, and POLG1 in sporadic progressive external ophthalmoplegia (PEO). *Neurology* **60**, 1354–1356 (2003).

71. I. Visapää, V. Fellman, J. Vesa, A. Dasvarma, J. L. Hutton, V. Kumar, G. S. Payne, M. Makarow, R. Van Coster, R. W. Taylor, D. M. Turnbull, A. Suomalainen, L. Peltonen, GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. *Am. J. Hum. Genet.* **71**, 863–876 (2002).

72. M. McKenzie, E. J. Tucker, A. G. Compton, M. Lazarou, C. George, D. R. Thorburn, M. T. Ryan, Mutations in the gene encoding C8orf38 block complex I assembly by inhibiting production of the mitochondria-encoded subunit ND1. *J. Mol. Biol.* **414**, 413–426 (2011).