Supplementary Materials

Supplement to Lieber et al., Targeted exome sequencing of suspected mitochondrial disorders

Appendix e-1: Supplementary Methods

MitoExome sequencing

Targeted gene transcripts from the RefSeq[1] and UCSC known gene[2] collections were downloaded from the UCSC genome browser[3] assembly hg19 (February 2009) corresponding to all genes listed in table e-1. Single 120bp baits were synthesized on a chip (Agilent) for each target region, or tiled across targets that exceeded 120bp. Two Agilent bait sets were synthesized: design1 contained 42,119 baits targeted to mtDNA and 1,560 nuclear genes (coding regions and 5' UTRs, single bait per target), and design2 contained 25,963 baits targeted to mtDNA and a superset of 1,598 nuclear genes (coding regions only, two baits per nuclear target and single bait per mtDNA target) (table e-1). 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.[4, 5] Index tags were added to 60 samples to facilitate "barcoded" multiplex sequencing. All samples were sequenced at the MGH Next-Generation Sequencing Core (http://nextgen.mgh.harvard.edu). 42 samples were sequenced in singleplex (one sample per lane, Agilent bait design1) and 60 barcoded samples were multiplexed (4-12 samples per lane, Agilent bait design2).

Variant detection

Sequence alignment, variant detection, and variant annotation for the nuclear and mitochondrial genomes were performed as described previously,[6] with modifications described here. Illumina reads were aligned to the GRCh37 reference human genome assembly using BWA[7] and variants were detected and annotated using the Genome Analysis Toolkit (GATK) software package version 1.0.5973 [8] and custom scripts. The ReadBackedPhasing algorithm was used to determine the phase of heterozygous variants. For the analysis of mtDNA, all reads were separately aligned to the mtDNA revised Cambridge Reference Sequence (NC_012920) and heteroplasmy level for SNVs was estimated based on the GATK pooled variant caller (Unified Genotyper v1.0.2986) assuming a mixture of 500 mtDNA chromosomes (-poolSize 500). mtDNA indels were detected using the standard Unified Genotyper and heteroplasmy was estimated as the percentage of aligned reads containing the variant. mtDNA variants supported by ≥1% of aligned reads were detected. Variants present at <20% heteroplasmy were deemed tentative until confirmed at higher heteroplasmy in an independent DNA sample or a maternal relative. We developed an algorithm to exclude mtDNA variants derived from sequence artifacts or nuclear sequences of mitochondrial origin (NuMTs)[9], based on filtering out low-heteroplasmy variants (<10%) that met the following criteria: (i) were present in at least 10% of samples, (ii) showed apparent paternal transmission based on sequencing of five trios, or (iii) exhibited correlation with the nuclear:mitochondrial DNA content across the 102 patients ($p<0.01$ Bonferroni-corrected, t-test of Spearman correlation). The nuclear:mitochondrial DNA ratio was estimated by the ratio of reads aligning to chromosome 1 to reads aligning to mtDNA, respectively. Large mtDNA deletions or rearrangements were detected through *de novo* mtDNA assembly using a modified ALLPATHS assembly approach, as previously described.[6]

Confirmatory genotyping and phasing

All variants underlying molecular diagnoses were independently validated by Sanger sequencing, and compound heterozygous variants were phased through read-backed phasing (GATK version 1.0.5973), sequencing familial DNA, or haplotype inference using genotype data from 1000 Genomes. For haplotype inference, two rare variants in the same gene observed together in multiple individuals were

considered unlikely to be compound heterozygous and thus excluded, whereas two rare variants in the same gene observed to be mutually exclusive across multiple individuals were considered to be likely compound heterozygous.

Heteroplasmy levels for three known pathogenic mtDNA mutations (MELAS, MERRF, NARP) were quantified at in DNA derived from whole blood and/or patient lymphoblastoid cell lines (LBLs) using ARMS quantitative PCR (Medical Genetics Laboratories, Baylor College of Medicine, Test ID 3006).[10] Dihydropyrimidine dehydrogenase (DPD) deficiency was confirmed by a purine and pyrimidine panel (Mayo Clinic, Test ID 81420).

Modeling the ATP5A1 p.Y321C mutation in yeast

An *atp1* null mutant was created in the W303-1A *S. cerevisiae* yeast strain (*a, ade2-1, his3-1, 15, leu2–3, 112, trp1-1, ura3-1*) by homologous recombination of PCR products using the KanMX resistance marker, as previously described.[11] pRS305 plasmids expressing the wildtype or p.Y315C allele of *ATP1* were created via Quikchange (Agilent). Residue numbering is based on the precursor peptide. Multiple sequence alignment was performed using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2) with the following UniProt IDs: P25705, Q03265, P35381, P07251, and P0ABB0.

dbGaP accessions

Patient identifiers from this study are listed in dbGaP phs000339 with the prefix "MGO".

Appendix e-2: Supplementary Notes

MitoExome sensitivity and specificity

For each individual, we detected SNVs and indels compared to the reference human genome. On average, individuals had 24 homoplasmic mtDNA variants, 1470 nuclear SNVs, and 56 nuclear indels. Over 95% of nuclear SNVs in each individual represented common polymorphisms in existing databases.

To assess sensitivity and specificity, we performed MitoExome sequencing on two HapMap individuals (NA12878, NA12891) with publicly available genotype data (HapMap Phase 3 release 2) and deep-coverage whole-genome data (1000 Genomes Project pilot 1[12]). Nuclear SNVs were estimated to be 97.9% sensitive and 99.9% specific, based on 797 and 1655 sites genotyped using independent technology. We detected 78/82 (95%) of indels ranging between 1 and 13bp in size in our HapMap samples, based on comparisons using the same GATK indel detection algorithm on deep-coverage whole genome sequence data, although we note that this algorithm may have a significant false negative rate. Reproducibility was assessed by independent hybrid selection and sequencing of HapMap sample NA12878, and showed 98% genotype concordance at heterozygous sites, consistent with our reported sensitivity.

The particularly deep coverage of the mtDNA enabled us to detect mtDNA variants with both high and extremely low (≥1%) levels of heteroplasmy. mtDNA SNV detection was estimated at 98% sensitivity and 99.5% specificity, based on reports for 29 individuals with independent whole-mtDNA sequence data (Medical Genetics Laboratories, Baylor College of Medicine, Test ID 3055). Additionally, we detected 11/12 sites independently assessed by quantitative PCR (qPCR) from the same tissue sample, including 5/6 variants present at <10% heteroplasmy (table e-4). At one site (m.3243A>G) assessed across seven individuals with MELAS, we observed strong correlation (R^2=0.98) between the levels of heteroplasmy in LBLs as estimated by NGS read count versus qPCR (figure e-6 panel A). However, the estimates of heteroplasmy consistently differed by two-fold between these two methods. We note that qPCR validation was performed at only three mtDNA sites and that other study designs, such as sequencing pre-defined mixtures of mtDNA species,[13] are required to reliably assess the accuracy of NGS-based quantitation of heteroplasmy.

Heteroplasmy in LBL versus whole blood

Previous studies have shown that the levels of heteroplasmy depends on the source of the DNA.[14] Comparison of m.3243A>G heteroplasmy levels by qPCR in LBL versus whole blood revealed consistently higher heteroplasmy levels in whole blood (figure e-6 panel B). This emphasizes the importance of tissue selection in future NGS studies of mitochondrial disease.

Loss-of-function variants in LBLs from MELAS patients

We observed that lymphoblastoid cell lines derived from MELAS patients carrying the m.3243A>G mutation harbored a large number of low-heteroplasmy loss-of-function (LOF) variants (i.e. frameshift indels and nonsense mutations). Specifically, we observed a tenfold enrichment of LOF mutations: 5 such mutations in samples from 7 MELAS patients (table e-4) versus 8 in all 102 patients ($p<1e-3$, binomial exact test). No enrichment was observed for missense or synonymous variants. The five LOF mutations in MELAS patients included four in complex I subunit and two that were previously reported as somatic mutations (m.5260G>A and m.4788G>A).[13, 14] Interestingly, somatic LOF mutations in Complex I subunits, including the m.5260G>A mutation, have been previously associated with oncocytic thyroid tumors.[15] Further studies are required to evaluate whether the m.3243A>G MELAS mutation causes a preferential accumulation of somatic loss-of-function mtDNA mutations either *in vivo* or in cell culture.

Prioritized variants of unknown significance (pVUS) in mtDNA

Six patients harbored pVUS in mtDNA that did not meet the requirements for molecular diagnosis, as described here:

- Patients 1008 and 1038 harbored pVUS in *RNR1* (m.1555A>G, m.1494C>T) that were not consistent with the previously reported phenotype of aminoglycoside-induced hearing loss. [16, 17]
- Patient 1011 harbored two heteroplasmic pVUS (*COX1*:p.G140X 42% heteroplasmy, *COX2*:p.W106X, 16% heteroplasmy) that were Sanger-confirmed in DNA from LBLs but were not detected in an independent blood sample from the patient or from an affected daughter, likely representing cell-line mutations.
- Patients 1051, 1085, and 1014 harbored low-heteroplasmy variants that have not yet been confirmed in an independent DNA sample (table e-5). These include two mutations previously reported in patients with MELAS [18, 19] and a *COX2* nonsense mutation also detected in Patient 1011.

Prioritized variants of unknown significance (pVUS) in nuclear disease genes

11 patients harbored pVUS in established nuclear disease loci that did not meet the requirements for molecular diagnosis, as summarized here:

- Eight patients harbored pVUS that were not consistent with previously reported phenotypes of the relevant genes. Five of these patients harbored rare recessive variants in established autosomal recessive disease genes (*DNAI1*, *DPYD*, *ETFB*, *SECISBP2*, *WFS1*), and four patients harbored rare heterozygous variants in dominant-acting disease genes (*POLG, POLG2*).
- Three patients harbored X-linked pVUS that did not have enough support to warrant diagnosis:
    - A hemizygous missense *AIFM1*:p.G53A variant was detected in male patient 1104. The variant was not present in the Exome Variant Server (EVS) nor 1000 Genomes. Previously, a hemizygous frameshift mutation in *AIFM1* was identified in a male with severe encephalopathy.[20] More recently, a hemizygous p.G308E missense mutation was identified in a patient with early prenatal ventriculomegaly and postnatal developmental delay.[21] The patient phenotype was not similar enough to either previous report to warrant molecular diagnosis.

- A hemizygous *HCCS*:p.A174V variant located near a splice site was detected in male patient 1002. The variant was present in 0/2443 male and 1/4060 female X chromosomes in EVS. Previous reports indicate that deletions, nonsense mutations, and *de novo* missense mutations in *HCCS* can cause male-lethality and X-linked dominant microphthalmia in females (OMIM #309801).[22, 23, 24] The patient phenotype was not similar enough to previous reports to warrant molecular diagnosis
- A heterozygous *NDUFA1*:p.G32R variant was detected in female patient 1100 with complex I deficiency (22% normalized complex I+III activity). The variant was found to be hemizygous in an affected brother and DNA from a second affected brother was unavailable. This mutation has previously been reported in the heterozygous state in a single female with complex I deficiency and in the hemizygous state in multiple males with complex I deficiency.[25, 26] However this allele is present at 1% allele frequency (AF) in European Americans in the Exome Variant Server and therefore may represent a polymorphism.

Recessive KARS mutations

Patient 1098 harbored compound heterozygous missense mutations in *KARS* (figure e-2), which were phased through parental genotyping. The p.T587M variant was absent from dbSNP, 1000 Genomes, and Exome Variant Server (EVS)[27] and altered a residue conserved to *E. coli.* The p.P228L variant was positioned within the anti-codon binding domain and altered a residue conserved to *C. elegans*, similar to the p.L133H mutation previously shown to impair aminoacylation *in vitro*.[28] The p.P228L variant was present in 2/12994 chromosomes in EVS. Both patient mutations were predicted to be "probably damaging" by PolyPhen2.[29]

Recessive *KARS* mutations have been associated with peripheral neuropathy (OMIM #613641) based on a single patient with peripheral neuropathy, developmental delay, dysmorphic features, and vestibular Schwannoma.[28] Patient 1098, who died by age 4, had no reports of peripheral neuropathy and no EMG or nerve conduction studies were performed to our knowledge. In addition to the single published case, compound heterozygous *KARS* mutations were recently identified in two siblings who presented in infancy with microcephaly, cortical blindness, developmental delay, and seizures (Abstract #701F, 62nd Annual Meeting of the American Society of Human Genetics).[30] We hypothesize that peripheral neuropathy could be a later-onset feature of recessive *KARS* mutations, just as recessive mutations in *DARS2* have been associated with childhood- or adolescent-onset peripheral neuropathy in addition to early-onset neurological symptoms such as tremors and ataxia.[31, 32] Further investigation of patients with recessive *KARS* mutations will likely expand the phenotypic spectrum associated with this disorder.

Hypothesized mechanism of ATP5A1 mutation causing combined respiratory chain deficiency

How might a mutation in the complex V subunit *ATP5A1* lead to combined respiratory chain deficiency and mtDNA depletion? Yeast experiments have shown that nearby mutations in this protein lead to loss of membrane potential and to mtDNA instability.[11] We hypothesize that human *ATP5A1* mutations may act in a similar manner. The mutation may cause uncoupling of complex V, which could lead to decreased membrane potential and subsequent loss of mtDNA by an undefined mechanism. The mtDNA loss would lead to decreased synthesis of mtDNA-encoded respiratory chain subunits and thus deficiencies of multiple respiratory chain complexes, as observed in the patient and her affected sister. Future experiments in mammalian cells are needed to explore this hypothesis.

Table e-1. Targeted genes

[see Excel file]

Table e-2. Variant details for molecular diagnoses and pVUS

[see Excel file]

Targeted DNA	Nuclear DNA	mtDNA
# gene loci	1589	37
target size (bp)	2,297,893	16,569
coverage		
mean coverage	226 [105-401]	12,680 [3,805-19,253]
% targeted bp covered ≥1X	97 [95-98]	100 [100]
% targeted bp covered ≥10X	95 [90-97]	100 [100]
% targeted bp covered ≥20X	93 [88-96]	100 [100]
# variants compared to reference DNA*	1521 [1427-1822]	55 [21-291]
# rare variants	40 [23-94]	9 [0-51]
# rare, likely deleterious variants	21 [11-40]	2 [0-10]
# genes with 2 rare, likely deleterious variants	0 [0-3]	NA

* mtDNA variants: ≥ 1% heteroplasmy

Table e-3: MitoExome sequencing statistics

Median values per patient are shown with range across patients indicated in brackets (N=102).

Patient ID	Patient Dx	OMIM #	mtDNA mutation (Prior Dx)	Observed heteroplasmy qPCR (Blood)	qPCR (LBL)	NGS (LBL)	Other prioritized variants
1107	MELAS	540000	m.3243A>G	37%	7%	2%	-
1110	MELAS	540000	m.3243A>G	47%	9%	4%	m.3659G>A 2% (*ND1*:p.W118X); **m.5260G>A** 1% (*ND2*:p.W264X)
1111	MELAS	540000	m.3243A>G	72%	52%	26%	-
1119*	MELAS	540000	m.3243A>G	27%	3%	0.7%	m.11866A>AC 16% (*ND4*:p.P370PfsX12)
1120	MELAS	540000	m.3243A>G	37%	35%	14%	**m.4788G>A** 2% (*ND2*:p.G107X)
1122	MELAS	540000	m.3243A>G	39%	9%	3%	m.15683C>T 1% (*CYTB*:p.Q313X)
1123	MELAS	540000	m.3243A>G	14%	4%	2%	-
1108	MERRF	545000	m.8344A>G	14%	0%	1%	-
1116	MERRF	545000	m.8344A>G	53%	67%	91%	-
1121	MERRF	545000	m.8344A>G	Hom	nd	71%	-
1112	NARP	551500	m.8993T>G	Hom	nd	84%	-
1124	NARP	551500	m.8993T>C	97%	nd	99%	-
1012	KSS	530000	m.3643_15569del	Het	nd	18%	-

* Variant not called by NGS pipeline due to heteroplasmy <1%

Table e-4. Quantification of mtDNA heteroplasmy in patients with prior mtDNA diagnoses

Percentages represent heteroplasmy levels derived from qPCR or from next-generation sequencing (NGS) read counts. Previously reported somatic mutations are indicated in bold (MITOMAP[33]). Abbreviations: LBL, lymphoblastoid cell line; MELAS, mitochondrial encephalomyopathy lactic acidosis and stroke-like episodes; MERRF, myoclonic epilepsy with ragged red fibers; hom., homoplasmic; het., heteroplasmic; nd, no data; KSS, Kearns-Sayre Syndrome; NARP, Neuropathy, ataxia, and retinitis pigmentosa;

Patient ID	Clinical description	Prioritized variants of unknown significance (pVUS)
1002	57 yo M w/ PEO, ptosis, dysphagia, muscle weakness, exercise intolerance, mild HL	*HCCS*:p.A174V (hem.)
1003	17 yo F w/ LD, myopathy, migraine, exercise intolerance, SS, GI dysmotility	*ACACB*:p.R830W+p.D1481V
1007	58 yo M w/ myopathy, exercise intolerance, sleep disorder, GI dysmotility	*DNAI1*:p.T48I+p.I682T
1008	4 yo M w/ GI dysmotility, dyspnea, exercise intolerance, dysphagia, SS	*RNR1* m.1555A>G (3% hp)[a]
1011	61 yo F w/ leukodystrophy, progressive dementia, gait difficulties, incontinent bladder and bowel; deceased; family history (sister and two daughters with leukodystrophy)	*COX1*:p.G140X (42% hp)[b] *COX2*:p.W106X (16% hp)[b]
1014	6 yo M w/ loss of skills, seizures, GI dysmotility, autonomic dysfunction	*COX2*:p.W106X (1% hp)
1015	5 yo F w/ PDD/Autism, GI dysmotility	*SECISBP2*: p.M233I,p.N102MfsX23
1019[c]	53 yo F w/ bilateral ptosis, PEO, ataxia, fatigue, swallowing dysfunction;family history (3 siblings with PEO)	*RNASEH1*:p.V142I+p.C148R
1020	3 mo F w/ FTT, microcephaly, encephalopathy, IUGR, hypotonia, pulmonary hypertension, heart failure; d. 3 mo; consanguinity (parents first cousins); family history (affected sister)	*ATP5A1*:p.Y321C (hom.) *HAO2*:p.D259E (hom.)
1025	17 yo F w/ myoclonus, myopathy, migraines, exercise intolerance, GI dysmotility	*POLG2*:p.Q422P (het.)
1026	33 yo F w/ myoclonic epilepsy, migraines, exercise intolerance, encephalopathy	*ANGEL2*:p.P80S+p.Y68Xfs; *MAN2B1*:p.R613Q,p.D74E; *POLG*:p.G609S (het.)
1032	23 yo M w/ brainstem stroke, lactic acidemia, SNHL, optic neuropathy, exercise intolerance	*NDUFAB1*:p.D132Y (hom.) *DPYD*:p.T768K (hom.)
1038	56 yo M w/ spastic paraparesis, optic atrophy, HL, bladder/bowel dysfunction, dysphagia	*RNR1* m.1494C>T (98% hp)
1040	31 yo F w/ visual loss, myopathy, GI dysmotility, tachycardia, exercise intolerance	*AK8*:p.N454S+p.R339C
1044	51 yo F w/ myopathy, SNHL, sleep disorder, GI dysmotility/reflux, DM, autonomic dysfunction	*ACACA*:p.E227K+c.1009-3T>G
1046	37 yo F w/ retinopathy, SNHL, exercise intolerance, GI dysmotility, headache	*THG1L*:p.A130T (hom.)
1051	2 yo F w/ GI dysmotility, fatigue, dystonia, DD, hypotonia, stroke-like events, migraines	*ND1* m.3949T>C (3% hp)
1067	30 yo F w/ ataxia, hypogonadotropic hypogonadism, hyperhidrosis, cognitive decline, migraine	*ETFB*:p.T285M+p.V7F *ABCB1*:p.K624R+p.I221V *POLG*:p.L392V (het.)

1072	20 yo F w/ migraines, exercise intolerance, GI dysmotility, autonomic dysfunction, myoglobinuria	*CERK*:p.V338G (hom.) *WFS1*:p.D389E+p.C426Y
1083	3 yo M w/ mild language delay, acute hypoxic encephalopathy	*APEX2*:p.R141C (hem.)
1085	53 yo F w/ DM, muscle weakness, exercise intolerance	*TRNL1* m.3271T>C (2% hp)
1088	6 yo M w/ pseudo obstruction, glucose intolerance, B cell dysfunction, DD, SS	*MAOA*:p.K520R (hem.)
1091	19 yo F w/ GI dysmotility, migraine, hypoplasia cerebellum, vision loss, LD, seizures, balance problems, autonomic dysfunction, hypotonia, immune deficiencies	*POLG*:p.P163L (het.)
1092	1 yo F w/ encephalopathy, vision loss, HL, myoclonus, dystonia, seizures, GI dysmotility; d. 1 yo	*ECH1*:p.P269L,p.Y55C
1100	15 yo F w/ GI dysmotility, cyclical vomiting, fatigue, exercise intolerance, headaches; family history (two affected brothers)	*NDUFA1*:p.G32R (het.)[d]
1104	1 yo M w/ global DD, dystonia, GI dysmotility, migraines, seizures	*ACSL4*:p.T227M (hem.); *AIFM1*:p.G53A (hem.); *ASMTL*:c.52-3G>C (het.)

[a] Variant present at 28% in maternal DNA
[b] Not detected by Sanger sequencing in whole blood of patient or affected daughter
[c] Patient harbored variant in *POLG2* underlying diagnosis of PEO, see Table 1 in Main Text
[d] Variant found to be hemizygous in an affected brother; other brother untested

Table e-5. Prioritized variants of unknown significance (pVUS) in 84 patients without prior molecular diagnoses

Prioritized variants (defined in Methods) are listed using Human Genome Variation Society (HGVS) nomenclature along with the patient genotype (hom., homozygous; het., heterozygous; hem., hemizygous; hp, heteroplasmy) and brief clinical description. Age represents age at diagnosis. Underline denotes established disease gene as in Figure 2 (Main Text). Variants separated by "+" are compound heterozygous. Variants separated by "," are potential compound heterozygous (unphased). Genomic coordinates available in table e-2.

Abbreviations: DD, developmental delay; DM, diabetes mellitus; F, female; FTT, failure to thrive; GI, gastrointestinal; HL, hearing loss; HTN, hypertension; IUGR, intrauterine growth restriction; LD, Leigh's disease; M, male; mo, month-old; PDD, pervasive developmental delay; PEO, progressive external ophthalmoplegia; SNHL, sensoneural hearing loss; SS, short stature.

Patient ID	Previous diagnosis	OMIM	Gene	Type	Recovered mutations underlying Mol. Dx	Other prioritized variants
1117	Mitochondrial Trifunctional Protein Deficiency	609015	*HADHA*	hom.	c.180+3A>G	-
1118	SANDO	607459	*POLG*	cmpd het.	p.R852C + [p.R627Q; p.G11D]	*RSAD1*:p.R9C (hom.)
1113	SANDO	607459	*POLG*	cmpd het.	p.G848S + p.A143V	-
1115	MIRAS	607459	*POLG*	hom.	p.A467T	-
1114	PDHAD	312170	*PDHA1*	hemizygous	c.963_977dup21	-

Table e-6. Prior molecular diagnoses in nuclear DNA recovered by targeted exome sequencing

Abbreviations: SANDO, sensory ataxia neuropathy dysarthria and ophthalmoplegia ; MIRAS, mitochondrial recessive ataxia syndrome; PDHAD, Pyruvate dehydrogenase E1 alpha deficiency; het., heterozygous; hom., homozygous; cmpd het., compound heterozygous; Mol. Dx, molecular diagnosis

	Patients without prior Mol. Dx (N=84)			All patients (N=102)		
Category	**# Mol. Dx/ # in category**	**% Mol. Dx**	**P-value**	**# Mol. Dx/ # in category**	**% Mol. Dx**	**P-value**
All	6/84	7%		24/102	24%	
ETC≤30%	3/52	6%	ns	0/0	-	-
Definite	4/24	17%	ns	8/28	29%	ns
Infantile	2/11	22%	ns	0/0	-	-
Family history	3/6	50%	<0.01	8/11	73%	<0.01
Deceased	2/5	40%	ns	3/7	43%	ns

Table e-7. Diagnostic yield in patient subsets

Rate of molecular diagnosis (Mol. Dx) within five subsets of patients: (i) ETC activity ≤30% in any complex, normalized to citrate synthase, (ii) "Definite" Morava score, (iii) Infantile age of diagnosis ≤2 years, (iv) Family history, based on first-degree relative with same phenotype, and (v) Deceased status. Patients with molecular diagnoses included all those listed in table 1, table e-4, and table e-6. P-value calculated with one-tailed binomial exact test relative to category "All" (first-row). Abbreviations: ns, not significant at $p<0.01$

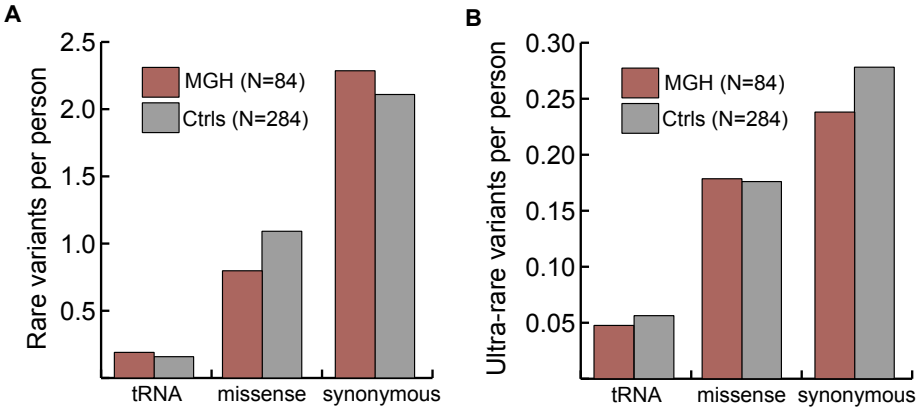

Figure e-1. Number of rare mtDNA variants in cases and controls

(A) Mean number of rare mtDNA variants, defined as allele frequency (AF) <0.3% in mtDB and AF<10% in sequenced cases.

(B) Mean number of ultra-rare mDNA variants, defined as not present in mtDB and AF<2% in sequenced cases.

MGH refers to 84 cases lacking previous molecular diagnoses. Ctrls refers to 284 European controls from 1000 Genomes Project. Analysis limited to high heteroplasmy variants (at least 50% heteroplasmy based on read count) that are not known to cause disease (MITOMAP).

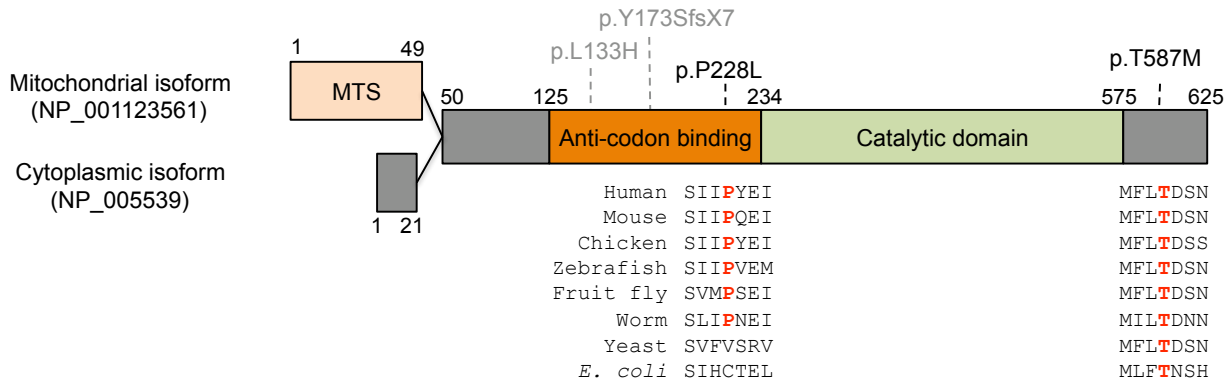

Figure e-2. Mutations in the KARS protein

Schematic diagram depicts two splice isoforms of KARS protein that differ at the N-terminus, with coordinates above listed relative to mitochondrial isoform. Black text indicates mutations detected in patient 1098, gray text indicates previously reported mutations.[28] ClustalW2 multiple sequence alignment shown for each patient mutation; UniProt identifiers: Q15046, Q99MN1, Q5ZKP8, F1QSF6, Q8SXM8, Q22099, P15180, P0A8N5. Abbreviations: MTS, mitochondrial targeting sequence

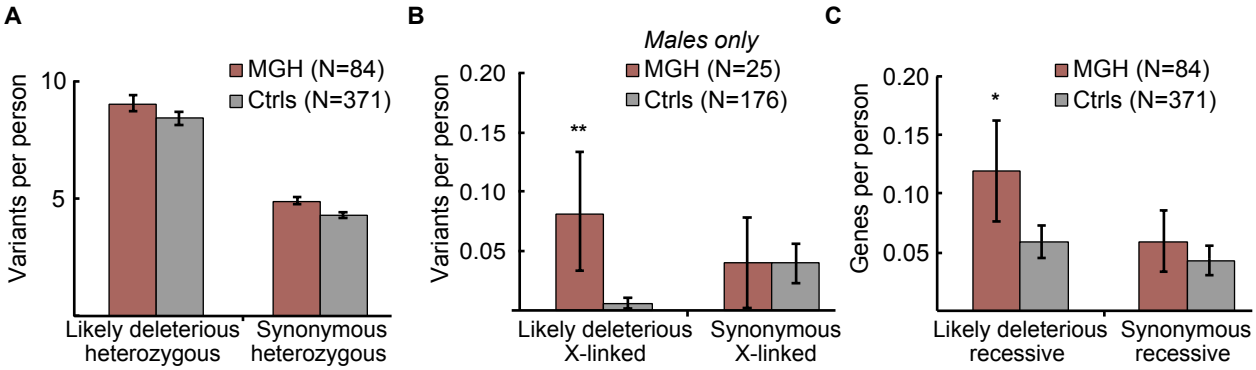

Figure e-3. Number of variants in genes not previously linked to mitochondrial disease

(A) Mean number of rare, likely deleterious, heterozygous variants.

(B) Mean number of rare, likely deleterious, hemizygous, X-linked variants in males.

(C) Mean number of autosomal genes containing rare, likely deleterious, recessive (i.e. homozygous or two heterozygous) variants.

MGH refers to 84 cases lacking previous molecular diagnoses. Ctrls refers to 371 healthy individuals from the NIMH Control repository. Rare and likely deleterious are as defined in Methods. The relevant number of rare, synonymous variants is shown in each panel as a control. Analysis limited to MitoCarta genes not previously linked to mitochondrial disease. Asterisks indicate significance based on binomial exact test (*$p<0.05$, **$p<0.01$). Error bars represent standard error of the mean.

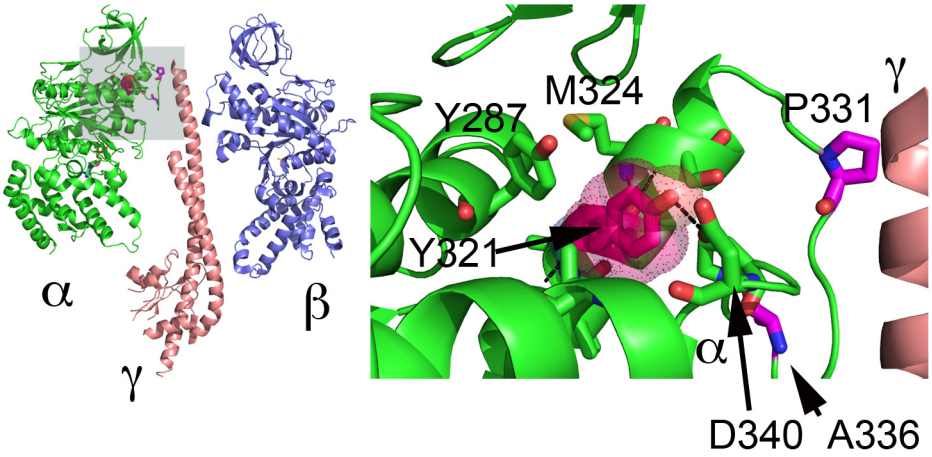

Figure e-4. Position of ATP5A1 patient mutation relative to yeast *mgi* mutations

Schematic of F_1 ATPase 3D structure based on x-ray crystal structures of the bovine[34] and yeast[35] complexes. Red indicates position of the ATP5A1:p.Y321 residue in the alpha subunit (coordinates based on human precursor peptide). Zoomed inset shows that patient mutation (αY321) forms a H-bond with αD340 and is within 4Å of αY287, αQ317-αS325, αR330, αP338, αD340. Note αP331 and αA336 are known *mgi* mutation sites.[11] These residues lie near the α–γ interface and are thought to be important for coupling of the F_1 and F_o sub-complexes.[36] Visualization was performed using Pymol[37] with coordinates of the bovine and yeast F_1 ATPase (pdb, 1W0J and 2HLD).

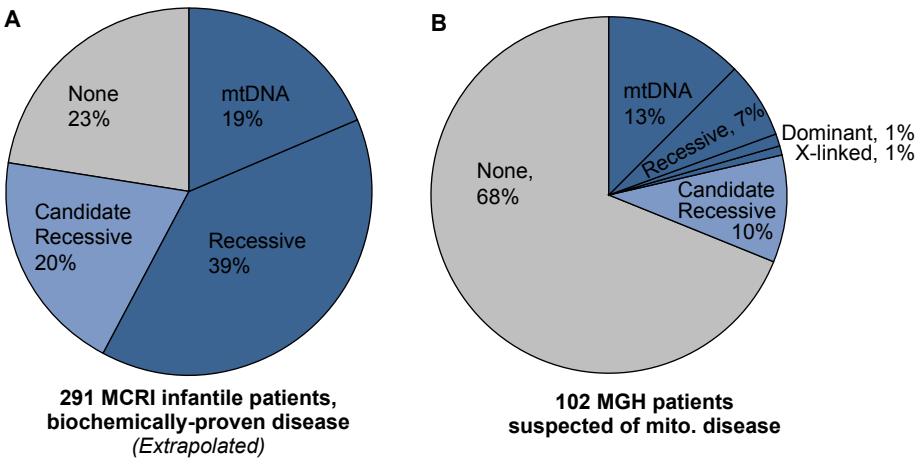

Figure e-5. Molecular diagnoses and recessive candidates in two patient groups

Comparison of MitoExome sequence results from our previous study[6] and current study shows the fraction of patients with firm molecular diagnoses (dark blue) and autosomal recessive candidate genes (light blue). Comparison was limited to the 1,034 genes sequenced in both studies (thus excludes 212 DDx genes). Unlike Figure 4 from the main text, this comparison shows all cases, including those with prior molecular diagnoses from traditional genetic testing.

(A) Murdoch Childrens Research Institute (MCRI) cohort of 291 infantile patients with severe ETC defects[6]: 124 with molecular diagnosis based on traditional sequencing, 38 with MitoExome sequencing, and 129 without MitoExome sequencing. MitoExome sequence results were extrapolated from the 38 patients with MitoExome sequencing to the 129 patients without MitoExome sequencing. New diagnoses from the extrapolation were then added to 124 prior diagnoses.

(B) 102 MGH patients: 18 with molecular diagnosis based on traditional sequencing (11 of which had diagnoses prior to referral to MGH), and 84 patients without prior molecular diagnosis (selected from a larger set of 159 patients as described in Methods).

As a caveat, we note that different methods of ascertainment were used for each group. Future prospective studies will be required to explore differences in the genetic architecture between these two groups.

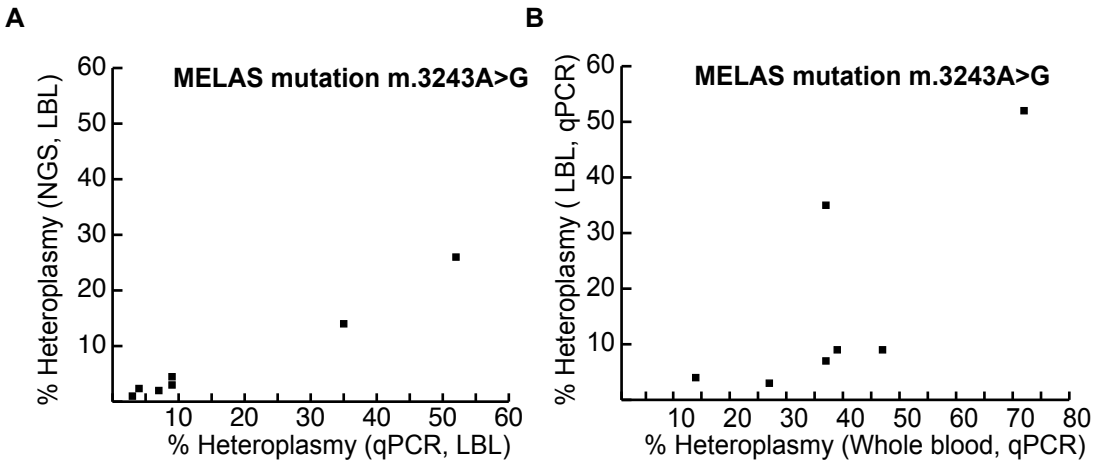

Figure e-6. Heteroplasmy of m.3243A>G across 7 patients with MELAS

(A) Quantification of heteroplasmy based on qPCR versus next-generation sequencing (NGS) read-counts in the same LBL-derived DNA sample.

(B) Quantification of heteroplasmy based on qPCR of DNA from whole blood versus lymphoblastoid cell line (LBL).

References

1. 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:D61-65.
2. Fujita PA, Rhead B, Zweig AS, et al. The UCSC Genome Browser database: update 2011. Nucleic Acids Res 2011;39:D876-882.
3. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res 2002;12:996-1006.
4. Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009;27:182-189.
5. Fisher S, Barry A, Abreu J, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol 2011;12:R1.
6. Calvo SE, Compton AG, Hershman SG, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. Science translational medicine 2012;4:118ra110.
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-1760.
8. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-1303.
9. Tourmen Y, Baris O, Dessen P, Jacques C, Malthiery Y, Reynier P. Structure and chromosomal distribution of human mitochondrial pseudogenes. Genomics 2002;80:71-77.
10. Wang J, Venegas V, Li F, Wong LJ. Analysis of mitochondrial DNA point mutation heteroplasmy by ARMS quantitative PCR. Curr Protoc Hum Genet 2011;Chapter 19:Unit 19 16.
11. Wang Y, Singh U, Mueller DM. Mitochondrial genome integrity mutations uncouple the yeast Saccharomyces cerevisiae ATP synthase. The Journal of biological chemistry 2007;282:8228-8236.
12. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature;467:1061-1073.
13. Porcelli AM, Ghelli A, Ceccarelli C, et al. The genetic and metabolic signature of oncocytic transformation implicates HIF1alpha destabilization. Human molecular genetics 2010;19:1019-1032.
14. Taylor RW, Barron MJ, Borthwick GM, et al. Mitochondrial DNA mutations in human colonic crypt stem cells. J Clin Invest 2003;112:1351-1360.
15. Gasparre G, Porcelli AM, Bonora E, et al. Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors. Proceedings of the National Academy of Sciences of the United States of America 2007;104:9001-9006.
16. Estivill X, Govea N, Barcelo E, et al. Familial progressive sensorineural deafness is mainly due to the mtDNA A1555G mutation and is enhanced by treatment of aminoglycosides. American journal of human genetics 1998;62:27-35.
17. Zhao H, Li R, Wang Q, et al. Maternally inherited aminoglycoside-induced and nonsyndromic deafness is associated with the novel C1494T mutation in the mitochondrial 12S rRNA gene in a large Chinese family. American journal of human genetics 2004;74:139-152.
18. Goto Y, Nonaka I, Horai S. A new mtDNA mutation associated with mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes (MELAS). Biochimica et biophysica acta 1991;1097:238-240.
19. Kirby DM, McFarland R, Ohtake A, et al. Mutations of the mitochondrial ND1 gene as a cause of MELAS. Journal of medical genetics 2004;41:784-789.
20. Ghezzi D, Sevrioukova I, Invernizzi F, et al. Severe X-linked mitochondrial encephalomyopathy associated with a mutation in apoptosis-inducing factor. American journal of human genetics 2010;86:639-649.
21. Berger I, Ben-Neriah Z, Dor-Wolman T, et al. Early prenatal ventriculomegaly due to an AIFM1 mutation identified by linkage analysis and whole exome sequencing. Molecular genetics and metabolism 2011;104:517-520.

22. Prakash SK, Cormier TA, McCall AE, et al. Loss of holocytochrome c-type synthetase causes the male lethality of X-linked dominant microphthalmia with linear skin defects (MLS) syndrome. Human molecular genetics 2002;11:3237-3248.
23. Wimplinger I, Morleo M, Rosenberger G, et al. Mutations of the mitochondrial holocytochrome c-type synthase in X-linked dominant microphthalmia with linear skin defects syndrome. American journal of human genetics 2006;79:878-889.
24. Wimplinger I, Shaw GM, Kutsche K. HCCS loss-of-function missense mutation in a female with bilateral microphthalmia and sclerocornea: a novel gene for severe ocular malformations? Molecular vision 2007;13:1475-1482.
25. Potluri P, Davila A, Ruiz-Pesini E, et al. A novel NDUFA1 mutation leads to a progressive mitochondrial complex I-specific neurodegenerative disease. Molecular genetics and metabolism 2009;96:189-195.
26. Mayr JA, Bodamer O, Haack TB, et al. Heterozygous mutation in the X chromosomal NDUFA1 gene in a girl with complex I deficiency. Molecular genetics and metabolism 2011;103:358-361.
27. Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [January 2012]. [online]. Available.
28. McLaughlin HM, Sakaguchi R, Liu C, et al. Compound heterozygosity for loss-of-function lysyl-tRNA synthetase mutations in a patient with peripheral neuropathy. American journal of human genetics 2010;87:560-566.
29. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nature methods 2010;7:248-249.
30. Geraghty M, Nowaczyk M, Humphreys P, et al. Exome sequencing reveals that mutations in the genes encoding aminoacyl tRNA synthetases (ARS) cause a variety of clinical syndromes. In: 62nd Annual Meeting of the American Society of Human Genetics (Abstract #701F). San Francisco, CA.
31. Yamashita S, Miyake N, Matsumoto N, et al. Neuropathology of leukoencephalopathy with brainstem and spinal cord involvement and high lactate caused by a homozygous mutation of DARS2. Brain & development 2012.
32. Isohanni P, Linnankivi T, Buzkova J, et al. DARS2 mutations in mitochondrial leucoencephalopathy and multiple sclerosis. Journal of medical genetics 2010;47:66-70.
33. Ruiz-Pesini E, Lott MT, Procaccio V, et al. An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res 2007;35:D823-828.
34. Kagawa R, Montgomery MG, Braig K, Leslie AG, Walker JE. The structure of bovine F1-ATPase inhibited by ADP and beryllium fluoride. EMBO J 2004;23:2734-2744.
35. Kabaleeswaran V, Puri N, Walker JE, Leslie AG, Mueller DM. Novel features of the rotary catalytic mechanism revealed in the structure of yeast F1 ATPase. EMBO J 2006;25:5433-5442.
36. Arsenieva D, Symersky J, Wang Y, Pagadala V, Mueller DM. Crystal structures of mutant forms of the yeast F1 ATPase reveal two modes of uncoupling. The Journal of biological chemistry 2010;285:36561-36569.
37. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. In, 2010.