Supplemental Text: Multi-omic analysis elucidates Complex I deficiency caused by a deep intronic variant in *NDUFB10*.

Guy Helman[1,2], Alison G. Compton[1,3], Daniella H. Hock[4], Marzena Walkiewicz[1], Gemma R. Brett[3,5], Lynn Pais[6], Tiong Y. Tan[1,3,5], Ricardo De Paoli-Iseppi[7], Michael B. Clark[7], John Christodoulou[1,3,5], Susan M. White[3,5], David R. Thorburn[1,3,5], David A. Stroud[4], Zornitza Stark[3,5,#], Cas Simons[1,2,#]

1. Murdoch Children's Research Institute, Royal Children's Hospital, Victoria, 3052, Australia
2. Institute for Molecular Bioscience, The University of Queensland, Queensland, 4072, Australia
3. Department of Paediatrics, University of Melbourne, Melbourne, Victoria, 3052, Australia
4. Department of Biochemistry and Molecular Biology and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Victoria, 3052, Australia
5. Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Royal Children's Hospital, Victoria, 3052, Australia
6. Center for Mendelian Genomics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 02142, USA
7. Centre for Stem Cell Systems, Department of Anatomy and Neuroscience, The University of Melbourne, Victoria, 3052, Australia

[#] To whom correspondence should be addressed

Correspondence to:
Zornitza Stark: zornitza.stark@vcgs.org.au | +61 3 8341 6368
Cas Simons: cas.simons@mcri.edu.au | +61 3 9936 6488

Methods

RNA Preparation and Short Read RNA Sequencing

All research activities were performed under Human Research Ethics Committee approval HREC/16/RCHM/150. RNA was isolated from cultured fibroblasts obtained from a skin biopsy. Cells were cultured both with and without cycloheximide to prevent nonsense-mediated decay (NMD) for 17 hours prior to harvesting for RNA extraction. Human whole transcriptome sequencing was performed by the Genomics Platform at the Broad Institute of MIT and Harvard. The transcriptome product combines poly(A)-selection of mRNA transcripts with a strand-specific cDNA library preparation, with a mean insert size of 550bp. Libraries were sequenced on the HiSeq 2500 platform to a minimum depth of 50-75 million 2 x 100 nt reads.

RNA sequencing reads were processed using a Bpipe (Version 0.9.9.6, release 21/07/2018) (Sadedin, Pope, & Oshlack, 2012) pipeline for quality control checks, trimming, and alignment. FastQC and Trimmomatic (Bolger, Lohse, & Usadel, 2014) were used for sequencing quality checks and trimming of poor-quality reads. Alignment was performed using STAR (Version 2.7.3a, release 08/10/2019)(Dobin et al., 2013), in two-pass mode for read alignment to the Human Reference Genome Build 38 (excluding "ALT" contigs). Duplicate reads were marked with Picard MarkDuplicates and quantification was performed using FeatureCounts from the R Subread package (Version 1.34.7, release 03/01/2019) (Liao, Smyth, & Shi, 2019). Differential expression analysis was performed using the DESeq2 package (Version 1.25.9, release 31/07/2019) comparing expression of the affected individual to 24 unrelated control samples (Love, Huber, & Anders, 2014). Sashimi plots were prepared using ggsashimi (Garrido-Martin, Palumbo, Guigo, & Breschi, 2018).

In silico prediction of alteration in mRNA splicing

The SpliceAI algorithm was used to assess variants found in *NDUFB10* intron 1 from the family under investigation that may impact splicing (Jaganathan et al., 2019). Pre-computed file of SpliceAI scores for all possible single nucleotide substitutions within Gencode genes spliceai_scores.masked.snv.hg38.vcf.gz was downloaded from https://basespace.illumina.com/s/otSPW8hnhaZR on January 17 2020. SpliceAI scores have a range of 0-1, with any score >0.2 predicted to alter splicing (Jaganathan et al., 2019).

RT-PCR and Nanopore Amplicon Sequencing

RNA was isolated using RNeasy Plus Mini Kit (Qiagen) and cDNA was prepared using a High-Capacity cDNA Transcription Kit (Applied Biosystems, cat # 4374966) according to manufacturer's instructions. The primers NDUFB10_E1_FWD and NDUFB10_E4_REV (Table S1) were used to amplify from exon1 to exon 4 of NDUFB10 using Q5 Hot Start High-Fidelity 2X Master Mix, (NEB, cat # M0494) and the

following PCR conditions: 98°C for 30 s, 35 x (98°C for 10 s, 68°C for 30 s, 72°C for 60 s), 72°C for 120 s. PCR products were treated with Exonuclease 1 (NEB, cat# M0293) before AMPure bead clean up (ratio 1:0.8) according to the manufacturer's instructions. 250 ng of each PCR product was end repaired using NEBNext FFPE Repair Mix (NEB, cat# M6630) and NEBNext End repair / dA-tailing Module (NEB, cat# E7546) before being cleaned up with Agencourt AMPure XP beads. Each PCR product was barcoded using the Native Barcoding Expansion 1-12 (ONT, EXPNBD104) before pooling and the ligation of sequencing adapters using the Ligation Sequencing Kit (ONT, SQK-LSK109). A final clean-up was performed using Agencourt AMPure XP beads with Short Fragment Buffer (ONT).

Assuming an average fragment length of 610 nt, 10 fmol of library was loaded on to a Flongle flowcell (FLO-FLG001) and sequenced on a GridION Mk1. Base calling and demultiplexing was performed on the GridION using MinKNOW v3.6.0 and Guppy v3.2.8 in high-accuracy mode.

FLAIR Long Read Analysis

All passing nanopore reads were processed using the FLAIR pipeline v1.5 (https://github.com/BrooksLabUCSC/flair) using the conda environment defined in the FLAIR package using default parameters unless otherwise noted (Tang et al., 2020). Reads were aligned to Human Reference Genome Build 38 (excluding "ALT" contigs) using flair.py align before correction flair.py correct using the SJ.out.tab junctions file generated from the STAR alignments of short read RNAseq data from Individual 1. The corrected reads from all four samples were concatenated and used as input for isoform discovery using flair.py collapse using "—stringent" flag and annotated (-f) using *NDUFB10* gene models from Gencode v32. Per sample read quantification was performed with flair.py quantify then visualised using plot_isoform_usage.py with the minreads parameter set to 75.

Minigene Splicing Assay

A minigene splicing assay was prepared to assess the splicing impact of a rare *NDUFB10* variant found in a region with aberrant splicing. We cloned the genomic DNA region containing exons 1-3 of *NDUFB10* and introduced the variant using three fragment HiFi DNA Assembly. The mutant bearing fragments were isolated using the Monarch DNA Gel Extraction Kit (New England BioLabs, #T1020L) and introduced into a linearized EYFP-C1 vector using the NEBuilder HiFi DNA Assembly Kit (New England BioLabs, #E5520S) (see Table S2 for primers), resulting in an insert size of approximately 2.1 kb and a total vector size of 6.1 kb.

The assembled product was transformed using NEB 5-alpha Competent *E. coli* cells and plated on an LB agar plate containing kanamycin. Plasmids for the mutant and a wild-type construct were Sanger-sequenced to confirm the presence of each variant in mutated constructs and the absence of variants in the wild-type construct (see Table S3 for primers). Wild-type and mutant plasmids were transfected into HEK293T cells (Lipofectamine™ 2000 reagent (Thermo Fisher Scientific, #11668019)) and cultured for 48 hours. RNA was isolated from the transfected cells using the RNeasy Plus Mini Kit (Qiagen) and cDNA was synthesized using a High-Capacity cDNA Transcription Kit (Applied Biosystems) according to the manufacturer's instructions. PCR was performed using *NDUFB10* exonic primers.

Mitochondrial Protein and Enzyme Studies

Five µg of total protein lysate was extracted from cultured fibroblasts and analyzed by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) using Total OXPHOS Human WB Antibody Cocktail (Abcam cat# ab110411; 1:500 dilution) as previously described (Calvo et al., 2012; Cooper, Lo, & North, 2003). This antibody detects subunits NDUFB8, SDHB, UQCRC2, COXII, and ATP5A of mitochondrial OXPHOS complexes I-V, respectively. An antibody against VDAC1 (Calbiochem cat#529534 1:10,000 dilution) was used as a loading control. Blots were incubated with sheep anti-mouse IgG HRP secondary antibody (GE Healthcare cat#NA931) and developed with Clarity Western ECL substrate (Bio-rad cat#170-5060). Separate panels represent images of the same gel captured with exposure times of 30 seconds for Complex V (Top Panel), 60 seconds for Complex I and Complex IV (Middle Panel), and 20 seconds for VDAC (Bottom Panel).

Quantitative Proteomics

A total of 50 µg of protein from Individual 1 (in technical triplicate) and control fibroblasts (three independent controls in singleton) were estimated by the Pierce BCA Assay Kit (ThermoFisher Scientific). Pellets were solubilized in 1% (w/v) SDC, 100mM Tris pH 8.1, 40mM chloroacetamide (Sigma) and 10 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP; BondBreaker, ThermoFisher Scientific) for 5 min at 99°C with 1500 rpm shaking, followed by 15 minutes sonication in a water bath sonicator. Proteins were digested overnight with trypsin (ThermoFisher Scientific) at a 1:50 trypsin:protein ratio at 37°C. The supernatant was transferred to stageTips containing 5x14G plugs of 3M™Empore™ SDB-RPS substrate (Sigma) as described previously (Kulak, Pichler, Paron, Nagaraj, & Mann, 2014; Stroud et al., 2016). Isopropanol 99% (v/v) and 1% (v/v) TFA was added to the tip before centrifugation at 3000 g at room temperature. StageTips were washed first with isopropanol (99%) and TFA (1%) solution and then subjected to an additional wash containing 0.2% (v/v) TFA. Peptides were eluted in 80% (v/v) acetonitrile (ACN) and 1% (w/v) NH_4OH, and then acidified to a final concentration of 1% TFA prior to drying in a CentriVap Benchtop Vacuum Concentrator (Labconco).

Peptides were reconstituted in 0.1% TFA and 2 % ACN for analysis and 1.5 µg from Individual 1 and control samples were combined to generate a library for MS1 spectral matching, which ran for six times in the same conditions. Liquid chromatography (LC) coupled MS/MS was carried out on an Orbitrap Lumos mass spectrometer (ThermoFisher Scientific) with a nanoESI interface in conjunction with an Ultimate 3000 RSLC nanoHPLC (Dionex Ultimate 3000). The LC system was equipped with an Acclaim Pepmap nano-trap column (Dionex-C18, 100 Å, 75 µm x 2 cm) and an Acclaim Pepmap RSLC analytical column (Dionex-C18, 100 Å, 75 µm x 50 cm). The tryptic peptides were injected to the trap column at an isocratic flow of 5 µL/min of 2% (v/v) ACN containing 0.1% (v/v) formic acid for 5 min applied before the trap column was switched in-line with the analytical column. The eluents were 5% DMSO in 0.1% v/v formic acid (solvent A) and 5% DMSO in 100% v/v ACN and 0.1% v/v formic acid (solvent B). The flow gradient was (i) 0-6min at 3% B, (ii) 6-95 min, 3-22% B (iii) 95-105min 22-40% B (iv) 105-110min, 40-80% B (v) 110-115min, 80-80% B (vi) 115-117 min, 80-3% and equilibrated at 3% B for 10 minutes before the next sample injection. The mass spectrometer was operated in positive-ionization mode with spray voltage set at 1.9 kV and the source temperature at 275°C. The mass spectrometer was operated in the data-dependent acquisition mode MS spectra scanning from m/z 350-1550 at 120,000 resolution with AGC target of 4e5. The "top speed" acquisition method mode (3 sec cycle time) on the most intense precursor was used whereby peptide ions with charge states ≥2-5 were isolated with an isolation window of 0.7 m/z and fragmented with high energy collision (HCD) mode with stepped collision energy of 35 ± 5%. Fragment ion spectra were acquired in Orbitrap at 50,000 resolution. Dynamic exclusion was activated for 30s.

Raw files were processed using the MaxQuant platform (Version 1.6.10.43) (Cox & Mann, 2008) and searched against UniProt human database (42,434 entries, June 2019) using default settings for an LFQ experiment with the following modifications: deamination (NQ) added as variable modification, "match between runs" enabled with default settings, and "LFQ min. ratio count" and "label min. ratio count" set to 1. The proteinGroups.txt output from the search was processed in Perseus (Version 1.6.10.43) (Tyanova et al., 2016). Briefly, entries "Only identified by site", "Reverse" and "Potential contaminant" were removed for analysis. Log2-transformed LFQ intensities were grouped in control or patient and filtered to have at least 2 valid values in each group. Two-sample t-tests were performed between groups using p-value for truncation (threshold p-value = 0.05). Volcano plots were generated via scatter plot by selecting "Student's T-test difference" and "-Log Student's T-test p-value". For the topographical heatmap, missing values for complex I subunits were imputed in patient samples using the algorithm built-in to Perseus, with width = 0.2 and down shift = 1.5 from normal distribution. Ratio data was

mapped onto the structure of bovine complex I (PDB: 5LDW (Zhu, Vinothkumar, & Hirst, 2016)) as previously described (Stroud et al., 2016). For the profile plots, LFQ intensities were imported into Prism 8 (Version 8.3.1) and a two-tailed ratio t-test was applied to determine the ratio (shown as $\log_2$ transformed) and significance of OXPHOS complexes in the NDUFB10 affected individual compared to controls.

Clinical Summaries

Individuals 1 and 2 are siblings, born of fourth-degree consanguineous parents of Pakistani origin. There is one unaffected sibling with a history of ventricular septal defect noted at birth, but otherwise no noted family history of neurologic disease although there are two paternal cousins with hearing impairment. Individual 1 is the elder of the affected sibling pair. He presented at birth, with delivery by emergency Cesarean section due to bradycardia after induction and meconium stained amniotic fluid. He stayed in the neonatal intensive care unit for 26 days due to suspected hypoxic ischemic-encephalopathy, low birth weight (2.72kg [$<10^{th}$ percentile]) and sepsis, in addition to poor feeding and gastro-esophageal reflux. MRI revealed a left-sided parietal cephalohematoma due to birth trauma. A skeletal survey revealed shortened first metacarpals bilaterally and gracile osteopenic parietal bones. At five months of age, he was noted to have slow growth, with a weight of 4.74 kg ($<3^{rd}$ percentile) and a head circumference of 40 cm ($<3^{rd}$ percentile). Auditory brainstem response testing diagnosed absent right-sided hearing and profound high frequency hearing loss on the left at seven weeks of age. In the months after discharge he had developed good head control and smiled at three months. On examination he was bright and alert. He had simple low-set ears and bilateral single palmar creases. He had a flexion deformity of the left hand (Type 1 radial dysplasia) and hypoplasia (Type II) of the left thumb, suggestive of a radial ray anomaly (Figure S1:1C). At nine months of age, he was still below the 3^{rd} centile in all growth parameters and had made some development gains in gross and fine motor skills. He made babbling noises and had become attentive to voices. He had multiple admissions due to concerns with growth and feeding, abnormal posturing episodes, and difficulties with respiration. At 11 months of age he was admitted due to increased episodes of apnea requiring intubation and died from complications of his underlying disease.

Laboratory investigations including urine organic acids and glycosaminoglycans, plasma very long chain fatty acids, amino acids and serum transferrin isoforms, were all normal. Plasma lactate levels were elevated at 7 months (3.5 mmol/L [normal range 1.0-1.8 mmol/L]). Pyruvate dehydrogenase complex activity in fibroblasts was normal. Respiratory chain enzyme activities were normal in fibroblasts. In skeletal muscle, when enzymes were expressed relative to protein, Complex I activity was normal but Complex II, III, and IV activities were elevated. Hence, Complex I activity was borderline low (~40% residual activity) relative to citrate synthase and Complex II. Mitochondrial complex I activity was also

borderline low in liver relative to citrate synthase (~50% residual activity while Complex II, III, and IV activities were in the range 87 to 120%). A muscle biopsy showed no significant histological pathology. Electron microscopy from a liver biopsy was unremarkable.

Further MRI studies were performed at nine months in Individual 1. There were signal abnormalities in the brainstem with T_2 hyperintensity and T_1 hypo-intensity suggesting demyelination in the corticonuclear tracts within the cerebral peduncles and affecting the inferior colliculi (Figure S1:1A-1B). The peri-aqueductal midbrain was affected. There was prominence of the ventricles and extra-axial spaces. On repeat studies two months later, the signal abnormalities in the midbrain were more extensive, affecting the substantia nigra and locus ceruleus as well as the superior colliculi (Figure S1:1F) and extending inferiorly into the dorsal pons and medulla. MRS did not reveal elevated lactate in the basal ganglia but technical limitations precluded this from being performed in the brain stem.

Individual 2 is Individual 1's younger brother who was born at term by repeat elective Cesarean section. The pregnancy was complicated by gestational diabetes mellitus requiring insulin in the final weeks of the pregnancy and maternal vitamin D and iron deficiencies and hypothyroidism requiring thyroxine. His birth weight was 2.9 kg (25^{th} percentile) and head circumference was 33 cm ($<10^{th}$ percentile). His newborn screening hearing test and perinatal period was normal. He made developmental gains without regression and at seven months was able to sit independently, roll over, transfer objects, vocalize, and was attentive to objects and sounds. At seven months he was noted to have a unilateral auditory neuropathy on audiologic testing which progressed bilaterally by eight months. Independent sitting was noted to have been lost on follow-up at nine months of age. On examination, he had failure to thrive and laryngomalacia with an inspiratory stridor. Extraocular eye movements were intact. There was mild head lag and truncal hypotonia with increased peripheral tone. He had good strength but reflexes were depressed. There was no clonus. At 14 months of age he was admitted to the intensive care unit due to episodes of bradycardia and persistent apnea, and noted to have increasing dystonic movements. He died at 14 months of age.

Laboratory studies included elevated blood lactate (3.6 mmol/L), normal creatine kinase and normal thyroid stimulating hormone. Plasma amino acids showed a mildly elevated alanine at 552 µmol/L (Reference range:148-475 µmol/L). Urine organic acids showed persistent excretion of fumaric acid. Very long chain fatty acids and serum transferrin isoforms were normal. Cerebrospinal fluid (CSF) pyruvate was mildly elevated (0.16 mmol/L [0.06-0.13 mmol/L]) and CSF alanine was significantly elevated (89 µmol/L [14-33µmol/L]). No mitochondrial respiratory chain studies were performed on this individual.

An MRI at 11 months revealed T_2 and FLAIR signal changes symmetrically within the cerebral peduncles, peri-aqueductal midbrain (Figure S1:2A-2B), and also at the anterolateral aspect of the 4th ventricle. On follow-up studies at 14 months, there was progression of the midbrain findings (Figure S1:2D-2E). There was new involvement of the inferior and superior colliculi and symmetrical small round areas of T_2 hyperintensity within both medial thalami inferiorly (Figure S1:2F). These findings were consistent, albeit less severe than his similarly affected sibling.

A chromosomal microarray was negative and breakage analysis to assess for Fanconi Anemia performed in the elder sibling was normal. Single gene analysis of *GJB2* for a genetic cause of the progressive hearing loss observed in Individual 1 was also normal. Testing for common mitochondrial genome point mutations and exome sequencing of both affected siblings and biological parents did not identify any candidates felt to be relevant to these patient's phenotype upon multidisciplinary review. Genome sequencing was then performed on both affected siblings and their parents by the Broad Center for Mendelian Genomics but was also unable to identify variants in phenotype-compatible genes.

Supplemental Tables:

Table S1: Clinical Summary

	Family 1		Friederich et. al.
	Individual 1	Individual 2	
Variants	NM_004548.3: c.131-442G>C (Hom.)		NM_004548.3: c.206_207insT;p.(Glu70*) NM_004548.3: c.319T>C;p.(Cys107Ser)
Age at Onset	Birth	Birth	*In utero*
Age at Death	12 months	14 months	27 hours
Pregnancy Complications	Fetal bradycardia after induction and meconium stained amniotic fluid	Gestational diabetes	Fetal cardiomyopathy, non-immune hydrops, lung hypoplasia, and symmetric intrauterine growth retardation
Neonatal complications	Hypoxic ischemic-encephalopathy; sepsis; low birth weight, poor feeding and gastro-esophageal reflux; left-sided parietal cephalohematoma	Low birth weight, otherwise normal	Bilateral pneumothoraxes and pulmonary hypertension; hypertrophic cardiomyopathy; lactic acidosis
Hearing Loss	Y	Y	UNK
Limb defects	Y, Radial ray anomaly	Y, Mild degree of hypoplasia	UNK
Relevant Investigations	Peak lactate 3.5 mmol/L (RR: 1.0-1.8 mmol/L).	Peak lactate - 3.6 mmol/L (RR: 0.5-2.0 mmol/L) Plasma amino acids - alanine 552 µmol/L (RR: 148-475 µmol/L) CSF pyruvate - 0.16 mmol/L [0.06-0.13 mmol/L]) CSF amino acids – alanine (89 µmol/L [14-33µmol/L])	Peak lactate - 29.19 mmol/L (RR: 0.5-2.0 mmol/L) Lactate-Pyruvate ratio – 33 Plasma amino acids – alanine 1403 mM (RR 131–710 mM), proline 1034 mM (RR: 110–417 mM) Urine organic acids – elevations of lactate, 3-hydroxybutyric, 2-hydroxybutyric, and glyoxalic species
Mitochondrial Respiratory Chain Studies	Decreased complex I activity in muscle, and liver tissue	Not performed	Decreased complex I activity in heart, muscle, and liver tissue Decreased complex II activity in liver tissue

RR - Reference range; CSF – Cerebrospinal fluid;

Table S2: RT-PCR of Fibroblasts from Individual 1

Primer	**Primer Sequence (5' to 3')**
NDUFB10_E1_FWD	CGGACGGAGGTAGAGGCCAG
NDUFB10_E4_REV	GGCGGCCTCTTTTGCAGC

Table S3: Minigene Splicing Assay Primers

Primer	Primer Sequence (5' to 3')
EYFP-C1: NDUFB10_ex1_fwd	gtcagatccgctagcgctaccgCGGACGGAGGTAGAGGCCAG
EYFP-C1: NDUFB10_ex3_rev	gtttcaggttcaggggggaggtgTGGACAGGCTTTGAGCCGATCCTGCATAATG
NDUFB10_131-442G-C_fwd	CGAGCTCTCCTTTTA**c**CTCGTTCATTTCCCAGTGAGG
NDUFB10_131-442G-C _rev	GGAAATGAACGAG**g**TAAAAGGAGAGCTCGAACATC
EYFP-C1_INV_FWD	CGGTAGCGCTAGCGGATCTGAC
EYFP-C1_INV_REV	CACACCTCCCCCTGAACCTGAAAC

Table S4: Variant validation Sanger sequencing primers

Primer	Primer Sequence (5' to 3')
NDUFB10_E1_Seq_FWD	GACGGAGGTAGAGGCCAG
NDUFB10_E3_Seq_REV	GAGGTGTGGACAGGCTTTG

Table S5: RT-PCR of Minigene splicing Assay

Primer	Primer Sequence (5' to 3')
EYFP-C1_PCR_FWD	CGTCAGATCCGCTAGCGCTAC
NDUFB10_E2_PCR_REV	CCTCTTCCACTGCATTTCGGCTTC

Table S6: Differentially expressed genes in RNA sequencing analysis of Individual 1

Gene	log2FoldChange	p-value	p-adj
NDUFB10	-1.85783491795367	2.83069399550321e-10	9.05878692440936e-06
LGALS3BP	-4.72382844693388	1.871415667893e-07	0.00199630147346373
NA	2.28781328847404	9.06006446718501e-07	0.00724850457697137
CHST11	-2.67239656885776	2.0423150117439e-06	0.0130716330011657
PCDHGA7	-3.69746963019335	2.96462093055958e-06	0.0136354290901365
PDE3A	-5.90041320363948	2.98256370323591e-06	0.0136354290901365
XIST	-25.6365774400112	4.42803225033785e-06	0.0157450986750347
ACO1	-1.57034767854533	6.35221725023665e-06	0.0189263481029283
LAMC1	-2.72746488269619	1.08624037120918e-05	0.0289682202995302
LPAR4	2.83017412331205	1.23989724978633e-05	0.0305224552212786
NID2	-2.9860547704538	2.24459370741041e-05	0.0478876585496986
LINC00601	3.08506862913851	2.54337677466509e-05	0.0488195248679934
APOL1	-5.29796059919297	3.21863174937555e-05	0.0542119227597454
ATP10D	-1.69120331712977	4.21225479324852e-05	0.0545113330475257
NA	-21.9206115964851	3.6660872514261e-05	0.0545113330475257
SLC39A4	-2.44579080607088	4.25843174235405e-05	0.0545113330475257
NA	2.45491357326114	4.0173961915846e-05	0.0545113330475257
ADA2	-5.15909131089536	4.44615740480234e-05	0.0547253574109555
NEXN	-2.32800507748542	5.57794254035082e-05	0.0637518989915382
PM20D2	1.95478042270083	5.51860115437479e-05	0.0637518989915382
COL4A2	-3.58952227203714	6.73922626415134e-05	0.0743685237604728
LMOD1	-4.10222942413689	7.80253562645034e-05	0.0805473371347302
COL4A1	-4.4161926391176	7.55978035300404e-05	0.0805473371347302
MYO1E	-2.11374197088706	9.30992016196889e-05	0.0931050203197901
NA	2.9782348438013	9.61560497692339e-05	0.0932480577186371

Table S7: Allele frequency of shared *NDUFB10* variants in Individuals 1 and 2

Variant	1KG	gnomAD Genomes	gnomAD Exomes	Topmed	CADD	SpliceAI
ENST00000268668.10:c.130+97G>C	0.318	0.163743	0	0.13873	6.034	
ENST00000268668.10:c.130+113dupC	0	0.163083	0	0.137694	2.383	
ENST00000268668.10:c.130+115C>G	0.318	0.162144	0	0.138618	6.822	
ENST00000268668.10:c.130+190C>A	0.318	0.162808	0	0.13857	0.356	
ENST00000268668.10:c.130+331C>A	0.318	0.163735	0	0.138674	5.449	
ENST00000268668.10:c.130+383G>A	0.9576	0.924884	0	0.84277	6.224	
ENST00000268668.10:c.130+426G>C	0.966	0.933653	0	0.83456	2.928	
ENST00000268668.10:c.130+482dupT	0.9145	0.933658	0	0.778104	3.527	

ENST00000268668.10:c.131-676C>G	0.2761	0.138235	0	0.089641	0.694	
ENST00000268668.10:c.131-577G>A	0.2812	0.161387	0	0.137264	0.02	
ENST00000268668.10:c.131-442G>C	0	0	0	0	2.464	DS_AG=0.085; DS_DG=0.145
ENST00000268668.10:c.131-297T>A	0.7428	0.720973	0	0.544892	1.719	
ENST00000268668.10:c.131-262T>C	0.3006	0.161153	0	0.137463	5.233	
ENST00000268668.10:c.131-147C>T	0.3006	0.161062	0	0.137479	1.757	
ENST00000268668.10:c.131-90T>C	0.3006	0.161308	0	0.137455	1.823	
ENST00000268668.10:c.131-72T>C	0.3006	0.161172	0	0.137519	3.884	
ENST00000268668.10:c.131-28A>G	0.9932	0.960692	0.966427	0.857312	0.426	

Table S8: Proteomics results for all peptides in *NDUFB10*-affected individual versus controls.

Table S9: Proteomics results for all peptides with minimum of two supporting samples in *NDUFB10*-affected individual versus controls.

Supplemental Figures:

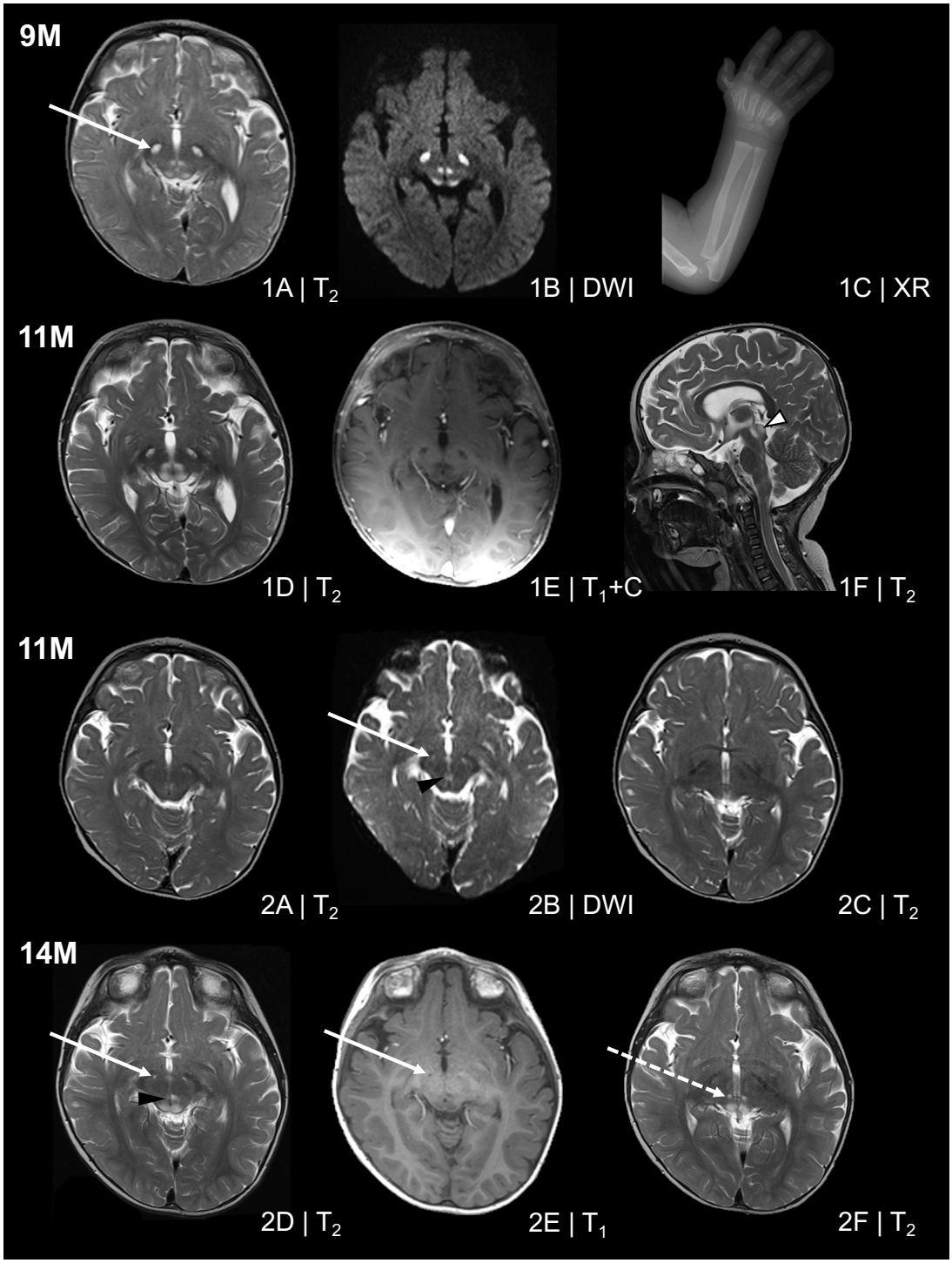

Figure S1. Imaging of NDUFB10 affected individuals. Magnetic resonance imaging was performed according to standard clinical protocols at nine and eleven months in Individual 1 and eleven and fourteen months in Individual 2. (1A-1B) Imaging at nine months of age in Individual 1 revealed T2 hyperintensity in the cerebral peduncles predominantly affecting the corticonuclear (white arrow) and frontopontine tracts, the peri-aqueductal midbrain and the inferior colliculi. Hypo-intensity on T1 (not shown) and bright regions on diffusion weighted imaging suggested a demyelinating process. There was no restricted

diffusion confirmed on the apparent diffusion coefficient (ADC) maps (not shown). Myelination was appropriate for age. There was mild enlargement of the ventricles and extra-axial spaces suggestive of atrophy. (1C) A left forearm X-ray performed at the same age shows hypoplasia of thumb and a mild flexion deformity. (1D-1E) Progression of the neuroimaging features observed in Individual 1 was observed on follow-up imaging at 11 months of age. There is continued hyperintensity on T2, and diffusion weighted imaging and hypo-intensity on T1 with (1E) and without contrast. There was extension superiorly to the superior colliculi (white arrowhead with black outline, 1F) and inferiorly into the dorsal pons and medulla. (2A-2C) Neuroimaging of the similarly affected sibling reveals similar features albeit less significant involvement at matched time points. (2B) The corticonuclear fibres (white arrow) and periaqueductal midbrain (black arrowhead) are mildly affected although not to the degree of the similarly affected sibling. Follow-up imaging shows a progression of these features, with extension into the inferior colliculi of the midbrain. (2E) A demyelinating process was suspected based on T1 images showing hypo-intensity within the corticonuclear fibres, peri-aqueductal midbrain, and the inferior colliculi. There are isolated lesions of the inferior thalami (white dashed arrow, 2F) and superior colliculi that were not apparent on imaging three months prior (2C).

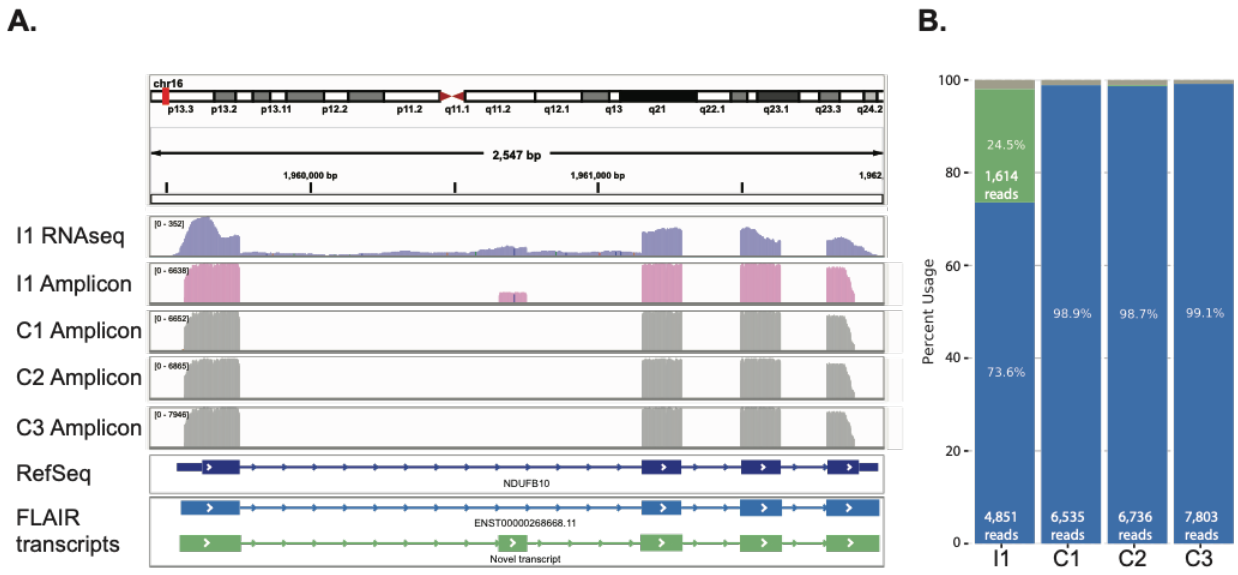

Figure S2. Long read sequencing of RT-PCR amplicons confirms the presence of a cryptic exon in intron 1 of *NDUFB10*. RT-PCR of *NDUFB10* was performed on cycloheximide treated fibroblast RNA using the same primers and PCR conditions as used in Figure 2C (see methods). An Oxford Nanopore Flongle was used to sequence full length amplicon reads yielding over 5,000 reads per sample. The Full-Length Alternative Isoform analysis of RNA (FLAIR) analysis tool(Tang et al., 2020) was used to identify and quantitate the isoforms present in each RT-PCR sample. (A.) Integrative Genomics Viewer (IGV) image comparing the base level read coverage of Illumina short read RNAseq from Individual 1 (I1) (purple, not cycloheximide treated) to nanopore sequencing of RT-PCR amplicons from cycloheximide treated fibroblast RNA also from Individual 1 (pink) and three control lines (C1, C2 and C3, shown in grey). The bottom panel depicts all isoforms identified by FLAIR from the nanopore data that account for a minimum of 1.5% of reads in at least one sample. The first transcript is identical to the RefSeq NDUFB10 transcript while the second transcript includes the addition of the cryptic exon in intron 1. (B.) Relative quantitation of isoform usage in each sample. The canonical NDUFB10 isoform (blue) accounts for 73.6% of transcripts in I1 and > 98.7 transcripts in each of the three controls. In Individual 1 the novel isoform including the cryptic exon (green) accounts for 24.5% of transcripts.

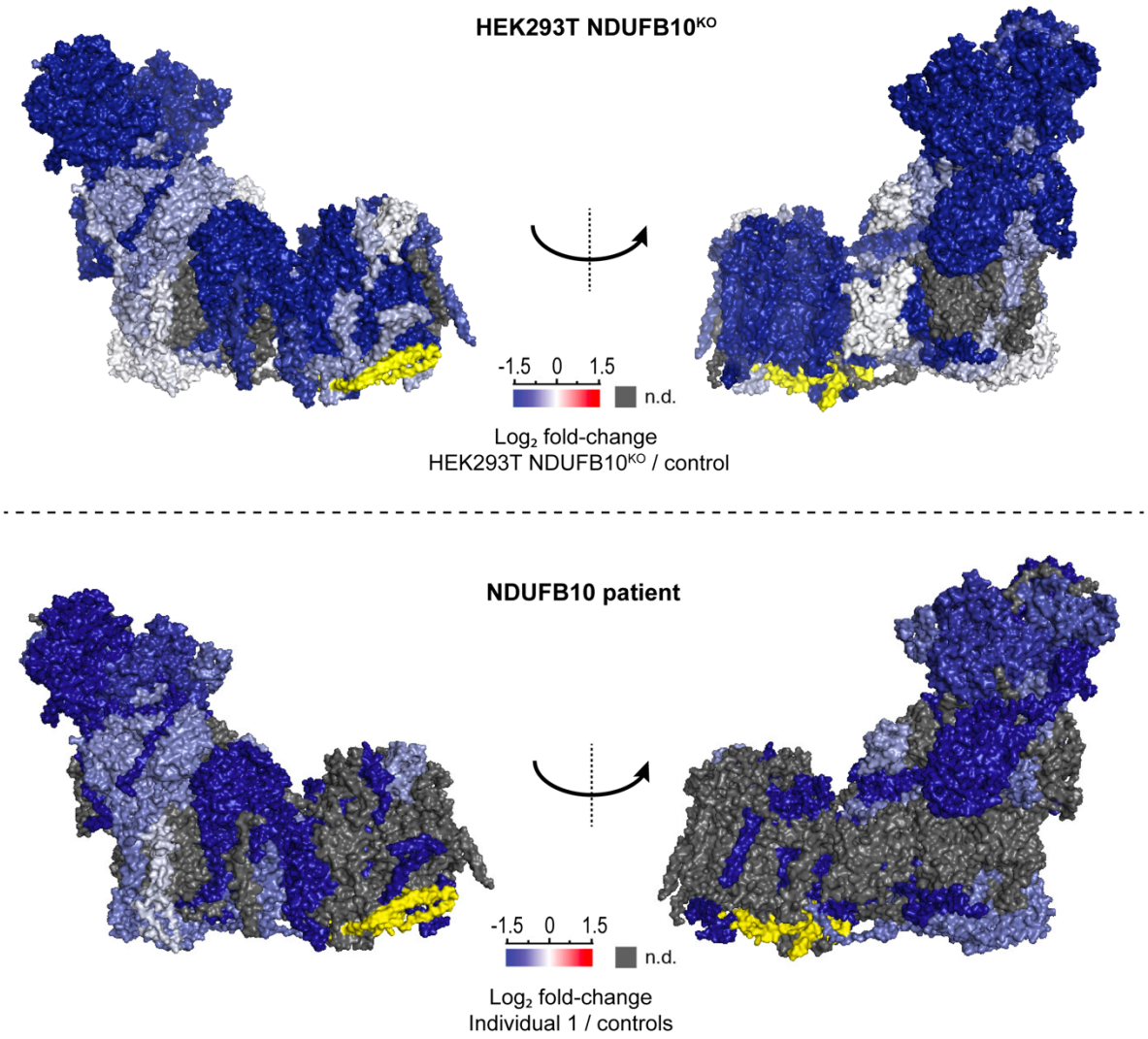

Figure S3. Topographical heatmap comparing the effects of the variant in Individual 1 on complex I subunits and gene-edited loss of NDUFB10 in HEK293T cells. Protein ratios for the HEK293T cell line were taken from a previously published dataset.(Stroud et al., 2016) Protein ratios for the NDUFB10 affected individual were taken from a two-sample t-test comparing Individual 1 to controls. Ratios were topographically mapped to the structure of bovine complex I (PDB: 5LDW). Proteins colored in blue are decreased, and those colored in red are increased. Grey indicates not detected (n.d.). and yellow indicates the NDUFB10 subunit.

References:

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*. doi:10.1093/bioinformatics/btu170

Calvo, S. E., Compton, A. G., Hershman, S. G., Lim, S. C., Lieber, D. S., Tucker, E. J., . . . Mootha, V. K. (2012). Molecular Diagnosis of Infantile Mitochondrial Disease with Targeted Next-Generation Sequencing. *Sci Transl Med, 4*(118), 118ra110-118ra110. doi:10.1126/scitranslmed.3003310

Cooper, S. T., Lo, H. P., & North, K. N. (2003). Single section Western blot. *Improving the molecular diagnosis of the muscular dystrophies, 61*(1), 93-97. doi:10.1212/01.Wnl.0000069460.53438.38

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol, 26*(12), 1367-1372. doi:10.1038/nbt.1511

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., & Jha, S. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*. doi:10.1093/bioinformatics/bts635

Garrido-Martin, D., Palumbo, E., Guigo, R., & Breschi, A. (2018). ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Comput Biol, 14*(8), e1006360. doi:10.1371/journal.pcbi.1006360

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., . . . Farh, K. K.-H. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell, 176*(3), 535-548.e524. doi:https://doi.org/10.1016/j.cell.2018.12.015

Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., & Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods, 11*(3), 319-324. doi:10.1038/nmeth.2834

Liao, Y., Smyth, G. K., & Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res, 47*(8), e47. doi:10.1093/nar/gkz114

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*. doi:10.1186/s13059-014-0550-8

Sadedin, S. P., Pope, B., & Oshlack, A. (2012). Bpipe: a tool for running and managing bioinformatics pipelines. *Bioinformatics, 28*(11), 1525-1526. doi:10.1093/bioinformatics/bts167

Stroud, D. A., Surgenor, E. E., Formosa, L. E., Reljic, B., Frazier, A. E., Dibley, M. G., . . . Ryan, M. T. (2016). Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature, 538*(7623), 123-126. doi:10.1038/nature19754

Tang, A. D., Soulette, C. M., van Baren, M. J., Hart, K., Hrabeta-Robinson, E., Wu, C. J., & Brooks, A. N. (2020). Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nature Communications, 11*(1), 1438. doi:10.1038/s41467-020-15171-6

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., . . . Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods, 13*(9), 731-740. doi:10.1038/nmeth.3901

Zhu, J., Vinothkumar, K. R., & Hirst, J. (2016). Structure of mammalian respiratory complex I. *Nature, 536*(7616), 354-358. doi:10.1038/nature19095