Advancing the diagnosis of mitochondrial diseases with genomic sequencing

Rocio Rius Dominguez

ORCID: 0000-0002-9871-3126

Doctor of Philosophy

August 2020

Department of Paediatrics

Faculty of Medicine, Dentistry and Health Sciences

University of Melbourne

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

Abstract

Mitochondrial diseases are the most common metabolic disease, caused by pathogenic variants in both nuclear and mitochondrial genomes. The molecular diagnosis can be challenging, sometimes taking years and requiring multiple invasive and risky diagnostic techniques. Genome sequencing (GS) is an emerging technology that allows the analysis of nuclear and mitochondrial genomes simultaneously. However, the interpretation of data can be complex, and the use of this technology in mitochondrial diseases is limited.

This PhD thesis explores the diagnostic utility of GS by investigating retrospective and prospective cohorts of patients with suspected (but molecularly undiagnosed) mitochondrial disease.

The thesis is divided into eight chapters. Chapter 1 begins with a general introduction to mitochondrial diseases and associated diagnostic challenges. Chapter 2 describes the subjects and methods used throughout the PhD. Chapters 3 to 6 use trio genome sequencing to analyse a retrospective cohort of patients with paediatric onset mitochondrial disease. In Chapter 3, trio GS uncovers *TEFM* and *COX11* as new candidate mitochondrial disease genes. In Chapter 4, trio GS was useful to identify a deep intronic variant in *NBAS*, enabling diagnosis of a patient who showed clinical overlap with mitochondrial disease. In Chapter 5, trio GS was used to identify a novel apparently synonymous variant in *PNPT1* where *in silico* and functional studies revealed a splicing defect instead. Clinical information and functional evidence from other patients with *PNPT1* variants were studied through an international collaboration to gain a better molecular and clinical understanding of *PNPT1*-related mitochondrial disease. In Chapter 6, trio GS data was interrogated to study mtDNA inheritance, showing an absence of biparental mitochondrial DNA transmission in the cohort.

In Chapter 7, the use of singleton GS data was studied in a prospective cohort of patients with suspected mitochondrial disease. A molecular diagnosis was made in 37% of cases. Examples that show the utility of GS include the identification of a mtDNA variant with low levels of heteroplasmy in blood (NC_012920.1(*MT-TL1*):m.3243A>T with 2% heteroplasmy), deletions in mtDNA (9kb deletion with 50% heteroplasmy) and nDNA (heterozygous 4.1kb

intragenic deletion in *AARS2*), and the use of GS data to aid in paternity confirmation and variant phasing in a patient with suspected Perrault syndrome and *PEX6* variants.

Finally, Chapter 8 concludes that GS is an effective tool for the diagnosis of mitochondrial diseases. The diagnostic potential is increased by combining GS with other techniques such as RNA sequencing. Further research is being conducted to assess the cost-benefit of this technology.

Declaration

This declaration certifies that:

1. This thesis contains only my original work towards the degree of Doctor of Philosophy except where indicated in the preface;
2. Due acknowledgement has been made in the text to all other material used;
3. This thesis is fewer than 100,000 words in length exclusive of tables, figures, bibliographies and appendices.

Rocio Rius Dominguez

09/08/2020

Preface

1. Contributions

The work towards this thesis was carried out in collaboration with others:

Genome sequencing (GS) was performed at the Kinghorn Centre for Clinical Genomics (KCCG, Garvan Institute, Sydney).

RNA sequencing, Sanger sequencing, and relatedness testing was performed at the Victorian Clinical Genetics Services (VCGS, Melbourne, Australia).

RNA sequencing data was analysed by Guy Helman and Dr Cas Simons from the translational bioinformatics group at the Murdoch Children's Research Institute (MCRI, Melbourne, Australia) for section 3.2.1

The Exome sequencing (ES) for Section 3.2.2 was performed in Bioscientia Center for Human Genetics and analysed by Dr Zafer Yueksel.

Blue-Native (BN)-PAGE for Section 3.2.2 was performed by Dr Luke Formosa (Monash University, Melbourne).

The Real-time ATP assay for Section 3.2.2 was conducted by Dr Neal K. Bennett (Prof. Nakamura's group, Gladstone Institute of Neurological Disease, USA).

Exome sequencing and variant curation for the patients in the ES arm in Section 7.2 was performed by Victorian Clinical Genetics Service (VCGS, Melbourne, Australia) staff members.

Diagnostic enzymology assays were performed by Tegan Stait and Simone Tregoning from the mitochondrial diagnostic service (MCRI/VCGS Melbourne, Australia) at the Brain and Mitochondrial Research Group (MCRI, Melbourne, Australia).

2. Publication status

a) Publications contained in this thesis where I was the primary author and contributed to more than half of the content include:

Chapter 4- Rius, R., L. G. Riley, Y. Guo, M. Menezes, A. G. Compton, N. J. Van Bergen, V. Gayevskiy, M. J. Cowley, B. B. Cummings, L. Adams, C. Ellaway, D. R. Thorburn, H. Hakonarson and J. Christodoulou (2019). "Cryptic intronic NBAS variant reveals the genetic basis of recurrent liver failure in a child." Mol Genet Metab **126**(1): 77-82.

Contribution: First author.
I contributed to the trio GS analysis, variant curation, performed the experimental work, data analysis and wrote the first draft of the manuscript included in Chapter 4.

Chapter 5- Rius, R., N. J. Van Bergen, A. G. Compton, L. G. Riley, M. P. Kava, S. Balasubramaniam, D. J. Amor, M. Fanjul-Fernandez, M. J. Cowley, M. C. Fahey, M. K. Koenig, G. M. Enns, S. Sadedin, M. J. Wilson, T. Y. Tan, D. R. Thorburn and J. Christodoulou (2019). "Clinical Spectrum and Functional Consequences Associated with Bi-Allelic Pathogenic PNPT1 Variants." J Clin Med **8**(11).

Contribution: First author.

I contributed to trio GS analysis of P2, variant curation, collected clinical information, performed the experimental work, data analysis and wrote the first draft of the manuscript included in Chapter 5.

Chapter 6- Rius, R., M. J. Cowley, L. Riley, C. Puttick, D. R. Thorburn and J. Christodoulou (2019). "Biparental inheritance of mitochondrial DNA in humans is not a common phenomenon." Genet Med **21**(12): 2823-2826.

Contribution: First author

I contributed to the trio GS data analysis for the cohort, and wrote the first draft of the manuscript.

b) Contributions to manuscripts in which I am not the primary author and are cited in this thesis include:

Chapter 3- Riley, L. G., M. J. Cowley, V. Gayevskiy, A. E. Minoche, C. Puttick, D. R. Thorburn, R. Rius, A. G. Compton, M. J. Menezes, K. Bhattacharya, D. Coman, C. Ellaway, I. E. Alexander, L. Adams, M. Kava, J. Robinson, C. M. Sue, S. Balasubramaniam and J. Christodoulou (2020). "The diagnostic utility of genome sequencing in a pediatric cohort with suspected mitochondrial disease." Genet Med **22**(7): 1254-1261.

In this chapter I included only my contributions to the article which involved reanalysis of GS data, clinical data collection, variant curation, and functional studies of the patients described in Chapter 3.

Chapter 7- Tucker, E. J., R. Rius, S. Jaillard, K. Bell, P. J. Lamont, A. Travessa, J. Dupont, L. Sampaio, J. Dulon, S. Vuillaumier-Barrot, S. Whalen, A. Isapof, T. Stojkovic, S. Quijano-Roy, G. Robevska, J. van den Bergen, C. Hanna, A. Simpson, K. Ayers, D. R. Thorburn, J. Christodoulou, P. Touraine and A. H. Sinclair (2020). "Genomic sequencing highlights the diverse molecular causes of Perrault syndrome: a peroxisomal disorder (PEX6), metabolic disorders (CLPP, GGPS1), and mtDNA maintenance/translation disorders (LARS2, TFAM)." Hum Genet.**139**(10): 1325-1343.

In this chapter I included only my contributions to the article which involved the analysis of GS data and experimental follow up of patient "P1" described in Section 7.3.1.

3. Funding

To undertake my PhD studies, I've received support from the CONACYT (Mexican National Council for Science and Technology) Postgraduate Research Scholarship, a Studentship from the Department of Paediatrics, University of Melbourne, and the Professor David Danks PhD Top Up Scholarship from Murdoch Children's Research Institute.

To attend conferences, I was awarded with the Henry and Rachel Ackman travelling scholarship from the Department of Paediatrics, University of Melbourne, the Student Conference Support from Murdoch Children's Research Institute, the Human Genetics Society of Australasia (HGSA) Vic/Tas branch Student Travel Award, and the Mito Foundation Award for Excellence in Mitochondrial Research.

Laboratory research activities were supported by grants from the Australian Genomics Health Alliance (National Health & Medical Research Council (NHMRC) GNT1113531), NHMRC GNT 1164479 and the Mito Foundation. In addition, research reported in this thesis was supported by donations from the Crane and Perkins families.

Acknowledgements

To my supervisors: Prof. John Christodoulou, thank you for the opportunity to do a PhD under your supervision, for opening the doors to the Brain and Mitochondrial Research Group that became my house for all these years. Thank you for the constant motivation and trust in this project.

Prof. David Thorburn, it has been a privilege to learn from you, thank you for sharing your expert knowledge during these years. Dr Nicole Van Bergen, thank you for teaching me the different lab skills involved in this project. Dr Alison Compton, I am grateful for all the guidance with variant curation and for sharing your genomic wisdom.

To my advisory committee: A/Prof. Shireen Lamandé and A/Prof. Tiong Tan for making the time for meetings, valuable feedback, and help to keep the direction of the project.

To all members of the brain and mitochondrial research group: Ann, Simone, Tegan, Cameron, Sumudu, Yau, AnneMarie, Carolyn, Nat, Sim, Sean, Xiaomin, and Nicole A., thank you for your support and friendship. I was fortunate to share this PhD with you. It has been extraordinary to be part of such a cohesive team.

To all members of the mitochondrial flagship: this work could not have been done without the fantastic clinicians, genetic counsellors, scientists, curators, and managers of the mitochondrial flagship who were involved in every step of the process. I am especially grateful to the members involved in the weekly "mini-MDTs", Prof. John Christodoulou, Prof. David Thorburn, Dr Alison Compton, Dr Naomi Becker, Dr Belinda Chong, Ms Shannon Cowie, Dr Michelle de Silva, and Dr Chern Lim.

To the financial sponsors: CONACYT (Mexican National Council for Science and Technology), University of Melbourne, Murdoch Children's Research Institute, Human Genetics Society of Australasia Vic/Tas branch, Brain and Mitochondrial Research Group, Australian Genomics Health Alliance and the Mito Foundation, thank you for funding and supporting this PhD.

***To my family*:** Ma y Fran, gracias por todo el apoyo, nunca lo hubiera logrado sin ustedes. Iris y Paco gracias por siempre estar conmigo "ni lejos…ni fronteras, ni lejos…ni barreras."

To my partner: Sophia, thank you for all the emotional support, encouragement, listening to my presentations, and celebrating every PhD milestone with me.

***To patients living with mitochondrial disease*:** you are the inspiration and drive of this work.

Abbreviations

3-methylglutaconic aciduria	3MGA
Allelic expression imbalance	AEI
American College of Medical Genetics and Genomics	ACMG
Australian Genomics Health Alliance	AGHA
Bicinchoninic acid	BCA
Blue native PAGE	BN-PAGE
Coenzyme Q_{10}	CoQ_{10}
Complementary DNA	cDNA
Complex I	CI
Complex II	CII
Complex III	CIII
Complex IV	CIV
Complex V	CV
Copy number variant	CNV
Cerebrospinal fluid	CSF
Citrate synthase	CS
Clustered regularly interspaced short palindromic repeats interference	CRISPRi
Cycloheximide	CHX
Cytochrome *c* oxidase	COX
Dulbecco's Modified Eagle's Medium	DMEM
Electromyography	EMG
Electron microscopy	EM
Exome sequencing	ES
Fluorescence resonance energy transfer	FRET
Gastrointestinal	GI
Genome sequencing	GS
Human Splice Finder	HSF
Interquartile range	IQR
Integrative Genome Viewer	IGV

Kinghorn Centre for Clinical Genomics	KCCG
Leber Hereditary Optic Neuropathy	LHON
Multidisciplinary Team	MDT
Melbourne Genomics Health Alliance	MGHA
Minor Allelic Frequency	MAF
Mitochondrial DNA	mtDNA
Mitochondrial Encephalopathy, Lactic acidosis, and Stroke-like episodes	MELAS
Massively parallel sequencing	MPS
Murdoch Children's Research Institute	MCRI
National Center for Biotechnology Information	NCBI
Neonatal intensive care unit	NICU
New South Wales	NSW
Nuclear DNA	nDNA
Nuclear-mitochondrial DNA segments	NUMT
Oxidative phosphorylation	OXPHOS
Polyacrylamide gel electrophoresis	PAGE
Quantitative PCR	qPCR
Queensland	QLD
Respiratory Chain	RC
Reverse transcription PCR	RT-PCR
RNA sequencing	RNAseq
Short tandem repeats	STR
Single Nucleotide Polymorphism	SNP
Single Nucleotide Variants	SNV
South Australia	SA
Structural Variants	SV
Succinate dehydrogenase	SDH
Primer Melting Temperature	Tm
Polyvinylidene fluoride	PVDF
Tasmania	TAS
Translocase of the inner membrane	TIM
Translocase of the outer membrane	TOM

Tricarboxylic acid cycle	TCA
University of California, Santa Cruz	UCSC
Phosphate Buffer Solution	PBS
Polyvinylidene fluoride	PVDF
Very-long-chain fatty acids	VLCFA
Victoria	VIC
Victorian Clinical Genetic Services	VCGS
Western Australia	WA
Wild Type	WT

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Mitochondria

Mitochondria are organelles present in almost all cells in the human body. The most accepted hypothesis argues that they evolved from a symbiotic event 1.5 billion years ago when an alpha-proteobacterium was engulfed by an ancestral eukaryote (Sagan 1967, Gray 2014).

This "endosymbiotic hypothesis" explains why, like prokaryotes, mitochondria have their own circular DNA, divide independently of the nucleus and initiate protein synthesis with formylmethionine (Suomalainen and Battersby 2018).

With the exception of mature red blood cells which lack mitochondria, every other cell in the human body can contain hundreds to thousands of mitochondria. The mitochondrion structure contains two specialized membranes. The outer membrane separates mitochondria from the cytosol, and contains protein pores (porin) to allow the transport of small proteins and ions.

The inner membrane, by folding into cristae has a large surface area, and contains the protein complexes involved in the respiratory chain (RC). Encapsulated by this membrane is the mitochondrial matrix, where the multiple copies of mtDNA are located and also where metabolic pathways such as the tricarboxylic acid cycle (TCA or Krebs cycle) produce metabolic intermediates that are then transported through the RC protein complexes (I, II, III, IV) in the inner membrane. During this process, a proton gradient in the intermembrane space is created and used by ATP synthase (Complex V) to form adenosine triphosphate (ATP). This metabolic pathway, whereby the mitochondria oxidize nutrients to synthesise ATP, is also called oxidative phosphorylation (OXPHOS), and constitutes the major source of energy that the human body needs (Figure 1) (Kuhlbrandt 2015, Gorman, Chinnery et al. 2016). Thirteen of the subunits that compose the enzyme complexes are encoded by the mtDNA, while the rest are encoded in the nuclear genome (nDNA) and need to be imported to the mitochondria (Table 1).

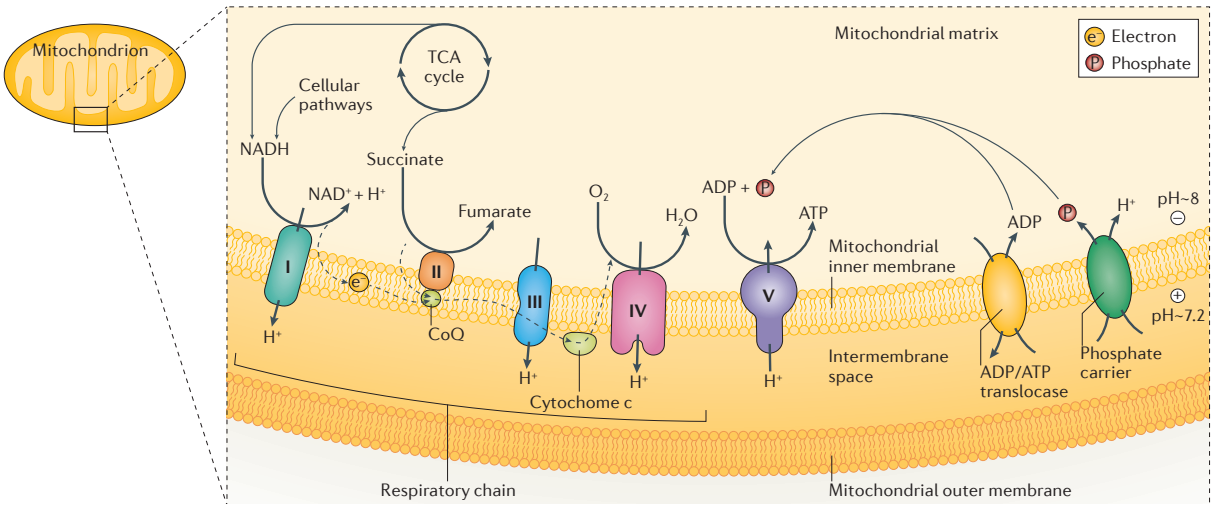

Figure 1 Oxidative phosphorylation.

Complex I (NADH-ubiquinone oxidoreductase), complex II (succinate-ubiquinone oxidoreductase), complex III (ubiquinol-cytochrome *c* oxidoreductase), and complex IV (cytochrome *c* oxidase, COX) are the mitochondrial RC enzymes located in the inner membrane and transfer electrons from substrates to molecular oxygen, generating an electrochemical transmembrane gradient. In the inner membrane, complex V (ATP synthase) uses the gradient energy to synthesise ATP. Figure from (Gorman, Chinnery et al. 2016).

Table 1 Human OXPHOS subunit composition

Complex	Name	Number of unique subunits	Number of subunits in the mtDNA
I	NADH ubiquinone oxidoreductase	44	7 MT-ND1, MT-ND2, MT-ND3, MT-ND4, MT-ND4L, MT-ND5 and MT-ND6
II	Succinate ubiquinone oxidoreductase	4	0
III	Ubiquinone-cytochrome *c* oxidoreductase	11	1 MT-CYB
IV	Cytochrome *c* oxidase	14	3 MT-CO1, MT-CO2, and MT-CO3
V	ATP synthase	16	2 MT-ATP6 and MT-ATP8

(Balsa, Marco et al. 2012, Alston, Rocha et al. 2017, Zong, Wu et al. 2018)

The human mitochondrial genome (mtDNA) is circular, double-stranded, and contains 16,569 base pairs. It comprises only 37 genes that encode for 13 proteins, two rRNAs (12S and 16S), and 22 tRNAs (Figure 2). However, ~1500 proteins are required for mitochondrial function, the majority of which are encoded by the nDNA and imported to mitochondria through the TIM/TOM pathway (Kunze and Berger 2015, Calvo, Clauser et al. 2016, Stenton and Prokisch 2018). These include proteins required for the generation of ATP by OXPHOS, and also in the numerous metabolic pathways in which mitochondria are involved, such as calcium homeostasis, hormone metabolism, metabolism of toxic compounds, non-shivering thermogenesis, regulation of membrane potential, apoptosis, fatty acid oxidation, TCA cycle, urea cycle, amino acid metabolism, lipid homeostasis, gluconeogenesis, ketogenesis, and biosynthesis of haem, steroids, and iron– sulfur clusters (Gorman, Chinnery et al. 2016, Frazier, Thorburn et al. 2019).

In humans, mtDNA is inherited from the mother and is present in multiple copies, mammalian oocytes carry ~4000-fold more copies of mtDNA than sperm (Hecht, Liem et al. 1984, Zhang, Burr et al. 2018, Latorre-Pellicer, Lechuga-Vieco et al. 2019); after fertilization the already low copy number of paternal mtDNA copies is selectively degraded (Sutovsky, Moreno et al. 1999, Sutovsky, Van Leyen et al. 2004).

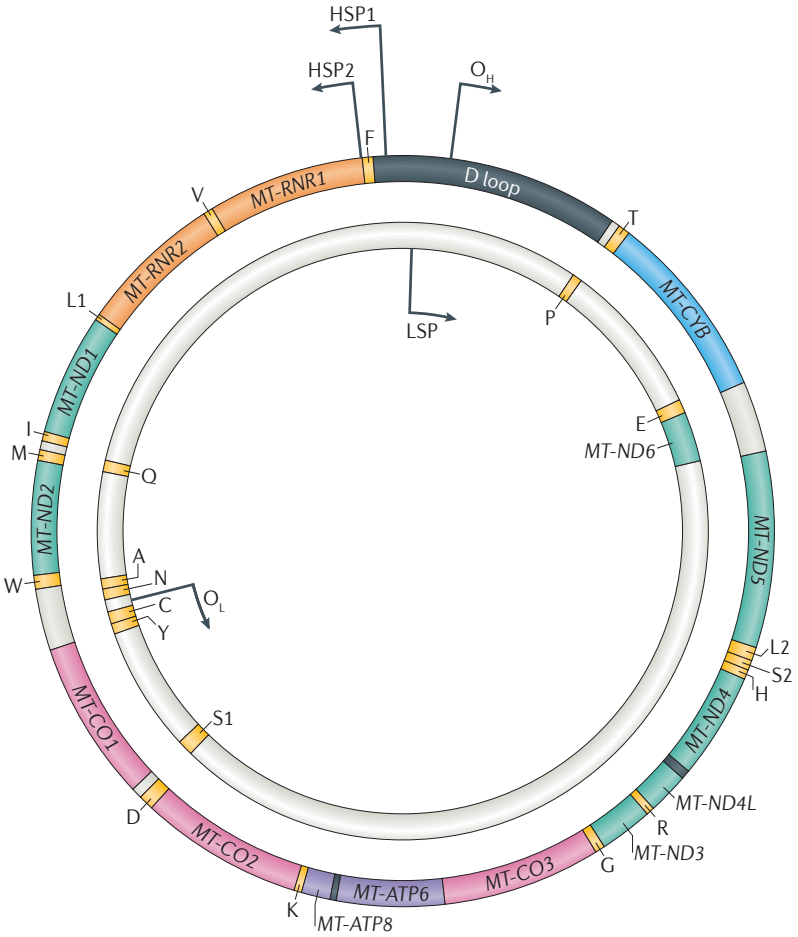

Figure 2 Human mitochondrial genome.

The mitochondrial genome is circular, double-stranded and contains 37 genes. tRNA genes are indicated in yellow. Origins of heavy-strand and light-strand mtDNA replication are indicated by O_H and O_L respectively. HSP, heavy-strand promoter; LSP, light-strand promoter. Figure from (Gorman, Chinnery et al. 2016).

1.2 Mitochondrial diseases

Mutations in more than 300 different genes can cause mitochondrial disorders (Stenton and Prokisch 2018, Frazier, Thorburn et al. 2019), and while each is individually rare, they are the most common group of inborn errors of metabolism, affecting at least 1 in 5,000 live births (Skladal, Halliday et al. 2003). They are caused by pathogenic variants that can impair mitochondrial energy generation by directly or indirectly impacting OXPHOS and other cellular functions (Gorman, Chinnery et al. 2016, Stenton and Prokisch 2018, Frazier, Thorburn et al. 2019). Symptoms can present at any age and generally involve organs with high ATP requirements. The dependence of neurons on OXPHOS to generate ATP results in the frequent presentation of neurological symptoms such as encephalopathy, seizures, and stroke-like episodes. Common non-neurological presentations include ptosis,

ophthalmoplegia, optic atrophy, cardiomyopathy, myopathy, exercise intolerance, hyperglycaemia, renal dysfunction and liver failure (Gorman, Chinnery et al. 2016).
Some patients present with a phenotype that can fit into an identifiable syndrome (Table 2). Yet, there is extensive clinical variability, and most patients cannot be grouped into any specific clinical phenotype.

Table 2 Examples of common clinical syndromes of mitochondrial disease and their clinical features

Syndrome	Causative Gene(s)	Common phenotype	Reference
Leigh syndrome	>90 genes in the nDNA and mtDNA	Encephalopathy, regression, basal ganglia and/or brainstem dysfunction with symmetric lesions.	(Lake, Compton et al. 2016, Rahman, Noronha et al. 2017)
Alpers syndrome and Childhood myocerebrohepatopathy spectrum	*POLG*	Refractory seizures, regression, liver failure, neuropathy, renal tubulopathy.	(Huttenlocher, Solitare et al. 1976, Hikmat, Tzoulis et al. 2017)
Kearns–Sayre syndrome (KSS)	Multiple deleted genes due to a single mtDNA deletion	Progressive external ophthalmoplegia, pigmentary retinopathy, cerebellar ataxia, cardiac conduction abnormalities, deafness, myopathy, diabetes mellitus.	(Kearns and Sayre 1958, Moraes, DiMauro et al. 1989)
Mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS)	mtDNA variants; m.3243A>G in *MT-TL1* is the most common	Stroke-like episodes, seizures, dementia, headache, hearing impairment, lactic acidosis, ragged-red fibres.	(El-Hattab, Adesina et al. 2015)

1.2.1 Inheritance of mitochondrial diseases

Since the proteins required for mitochondrial function are encoded by both nDNA and mtDNA, mitochondrial diseases can be dominant, recessive, X-linked, maternally inherited, or *de novo*. It is estimated that ~75% of primary paediatric mitochondrial diseases with onset in childhood result from variants in the nuclear genome. Pathogenic variants in nuclear genes can be inherited in any Mendelian pattern, but are most commonly autosomal recessive (Gorman, Chinnery et al. 2016).

mtDNA is maternally transmitted, and is present in many copies within a cell, however not all mtDNA copies carry the same variants; this mixture of mutant and wild-type mtDNA is known as "heteroplasmy". The degree of heteroplasmy may be different in the offspring than in their mothers, and the difference is explained by a restriction-amplification event called the "bottleneck effect", which occurs during oocyte maturation. During oogenesis, a small proportion of the mitochondria within the primordial germ cell is segregated to each primary oocyte. During maturation there is replication of the mtDNA, and therefore the mature oocyte may contain a different proportion (mutant load) of mutant mitochondrial DNA compared to the proportion found in the primordial germ cell (Figure 3) (Taylor and Turnbull 2005, Gorman, Chinnery et al. 2016).

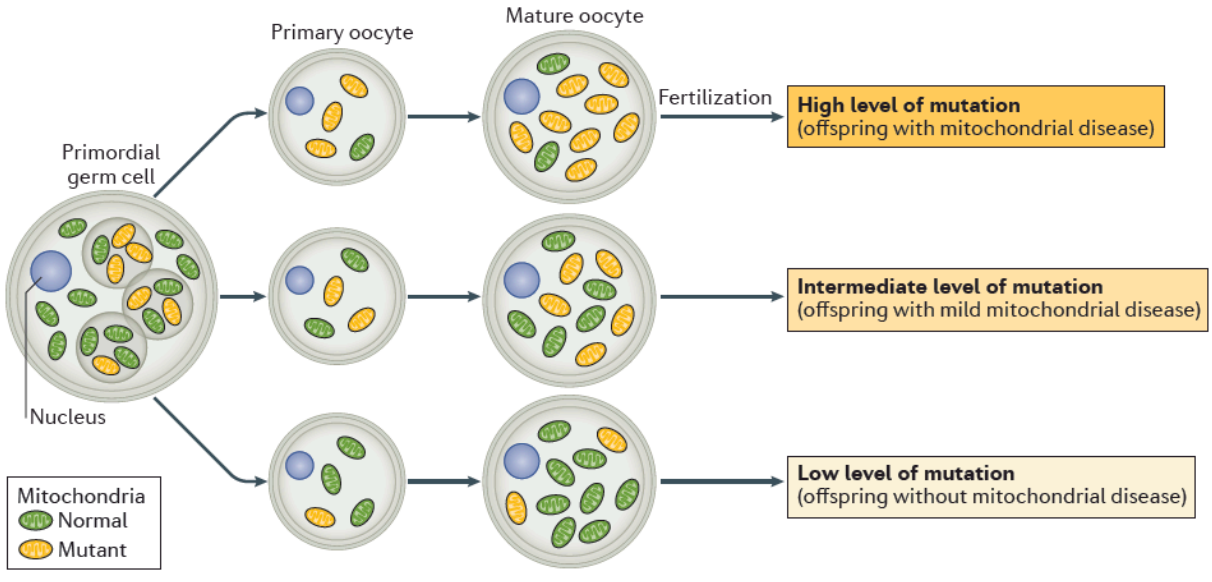

Figure 3 Mitochondrial bottleneck during oogenesis.

During oogenesis a small number of mtDNA molecules are segregated into each primary oocyte, followed by rapid replication leading to variable levels of heteroplasmy in the mature oocyte. Figure from (Gorman, Chinnery et al. 2016)

Furthermore, the heteroplasmy level can vary between cell types and tissues, and consequently heteroplasmy levels must reach a specific threshold before a clinical or biochemical phenotype is present, which depends on the type of variant and the cell type. The phenotypic severity of the disease can thus differ based on the proportion of mitochondria with wild type and mutant genomes (Stewart and Chinnery 2015).

1.2.2 Diagnosis of mitochondrial diseases

The diagnosis of mitochondrial diseases can be challenging. A survey conducted by Grier and colleagues showed that on average patients visit 8 different specialists before receiving a diagnosis, 55% are initially misdiagnosed, and 32% were misdiagnosed more than once (Grier, Hirano et al. 2018). Numerous factors contribute to the difficulty in establishing a mitochondrial diagnosis:

- They can be caused by pathogenic variants in the nDNA or in the mtDNA
- Approximately 300 genes have been involved in mitochondrial diseases
- There are different modes of inheritance
- Different levels of mtDNA heteroplasmy contribute to phenotypic variability
- Pleiotropy contributes to phenotypic variability
- There are different diagnostic classification systems (clinical, enzymology, molecular)
- There is no standardised diagnostic pathway

The approach for diagnosing mitochondrial diseases often varies between clinical presentations and is not standardised. Several diagnostic criteria have been created in an attempt to harmonise the clinical diagnosis of mitochondrial disorders, such as the Bernier and Nijmegen criteria, which use a combination of clinical, pathological, biochemical, and molecular analyses (Bernier, Boneh et al. 2002, Morava, van den Heuvel et al. 2006).

The Nijmegen criteria scoring system classifies the probability of mitochondrial disease as definite (score 8–12), probable (score 5–7), possible (score 2–4) or unlikely (score 1) (Morava, van den Heuvel et al. 2006). To incorporate additional currently available clinical, metabolic, and biochemical evidence of respiratory chain disorders, a modified version was used for scoring patients in this research project, and is presented in Table 3. The modified criteria include more specific signs/symptoms such as optic neuropathy and retinitis pigmentosa. It adds metabolic and biochemical biomarkers that have been identified in recent years such as

FGF21 and GDF15 (Lehtonen, Auranen et al. 2020), and it also includes well established enzymological evidence such as abnormal respiratory chain enzyme activity (Frazier, Vincent et al. 2020).

Table 3 Modified Nijmegen criteria

I. Clinical criteria (max. 4 points)			II. Metabolic and imaging studies (max. 4 points)	III. Morphology and biochemical (max. 4 points)
A. Muscular presentation (max. 2 points)	**B. CNS presentation (max. 2 points)**	**C. Multisystem disease (max. 3 points)**		
▪ Progressive external ophthalmoplegia [a] ▪ Facies myopathica ▪ Ptosis* ▪ Exercise intolerance ▪ Muscle weakness ▪ Rhabdomyolysis ▪ Abnormal EMG	▪ Intellectual disability* ▪ Loss of skills / regression ▪ Stroke-like episode ▪ Migraine ▪ Seizures ▪ Myoclonus ▪ Cortical blindness ▪ Optic neuropathy* ▪ Retinitis pigmentosa* ▪ Pyramidal signs ▪ Extrapyramidal signs ▪ Brainstem involvement	▪ Haematology ▪ GI tract ▪ Endocrine / growth ▪ Diabetes mellitus* ▪ Cardiomyopathy ▪ Hypertension* ▪ Kidney ▪ Sensorineural deafness* ▪ Neuropathy ▪ Recurrent /familial	▪ Elevated blood lactate [b*] ▪ Elevated L/P ratio ▪ Elevated blood alanine ▪ Elevated serum FGF21 [a*] ▪ Elevated serum GDF15 [a*] ▪ Elevated CSF lactate ▪ Elevated CSF protein ▪ Elevated CSF alanine ▪ Urinary TA excretion ▪ Ethylmalonic aciduria ▪ 2-ethylhydracylic aciduria* ▪ 3-methylglutaconic aciduria* ▪ Stroke-like picture/MRI [a] ▪ Leigh syndrome /MRI [a] ▪ Elevated lactate/MRS [a]	▪ Ragged red/blue fibres [c] ▪ COX-negative fibres ▪ Reduced COX staining ▪ SDH positive blood vessels [a] ▪ Abnormal mitochondria/EM [a] ▪ Abnormal RC enzymology [d*] ▪ mtDNA depletion or multiple mtDNA deletions [a*]

[a] this specific symptom scores 2 points, [b*] if blood lactate is elevated once scores 1 point; if elevated thrice scores 2 points,[c] 2 points if present; 4 points if >2%

[d] score 2 points if <20% residual activity (relative to marker enzymes such as citrate synthase or RC complex II) of any RC complex in a tissue or <30% residual activity of any RC complex in a cell line or <30% residual activity of any RC complex in two or more tissues. Score 1 point if 20–30% residual activity of any RC complex in a tissue or 30–40% residual activity of any RC complex in a cell line or 30–40% residual activity of any RC complex in two or more tissues. * Modified/new criteria.

Table modified from (Morava, van den Heuvel et al. 2006, Riley, Cowley et al. 2020).

Before the availability of Massively Parallel Sequencing (MPS), the traditional diagnostic pathway for diagnosis of mitochondrial disease would commonly require extensive evaluations, often including invasive studies such as liver and/or muscle biopsies to perform biochemical, histological and enzyme evaluations that could guide the molecular genetic testing of candidate genes by Sanger sequencing. If negative, this would often lead to more investigations and further genetic evaluations. With this approach, in addition to performing invasive testing in patients with an increased risk of anaesthetic complications, most patients with suspected mitochondrial disease remained genetically undiagnosed (Parikh, Goldstein et al. 2015, Wortmann, Mayr et al. 2017).

Over the past two decades, technical improvements and cost reductions in sequencing technologies have revolutionized the pathways to achieve a genetic diagnosis. Massively Parallel Sequencing (MPS) also known as Next-Generation Sequencing is technology that allows multiple sequencing reactions to occur simultaneously (Jarvik and Evans 2017).

With this technology multiple genes can be sequenced at the same time at relatively low cost, giving an opportunity to yield a more rapid and less invasive diagnosis where a known mitochondrial disease gene is involved, and to uncover novel candidate disease genes (Menezes, Riley et al. 2014, Topol 2014). Using MPS, the diagnostic yield has reached 35-60% in different mitochondrial disease cohort studies (Stenton and Prokisch 2018), and this so call “genetics-first” approach is increasingly becoming a routine diagnostic approach to use in patients with the suspicion of mitochondrial disease (Wortmann, Koolen et al. 2015).

Many MPS platforms are available, which differ in the methods generating sequencing data from DNA. The most commonly used platform is Illumina short-read sequencing by synthesis (SBS) (Bentley, Balasubramanian et al. 2008), which provides high coverage and accuracy to detect DNA variants. In brief, we can divide the sequencing process into four stages (Metzker 2010):

1. Library preparation: the DNA is fragmented and the regions of DNA that will be sequenced are isolated. The target regions can range from a single gene, a panel of genes, most of the exome (ES) or virtually all the genome (GS) (Jarvik and Evans 2017). Adapter oligonucleotide sequences at the ends are added as well as barcodes specific to the sample.

2. Isolation: the library is loaded into a flow cell where the individual DNA fragments are captured and immobilized by oligonucleotides complementary to the library adapters allowing the DNA molecules to be physically separated from each other.

3. Parallel sequencing: the DNA molecules are simultaneously sequenced, where each amplified molecule corresponds to its sequence or "read". Sequencing by synthesis uses a reversible terminator that detects single fluorescently labelled nucleotides as they are incorporated into DNA template strands.

4. Alignment and data analysis: the sequence is aligned to the reference genome, the variants are identified and exported for analysis.

The bioinformatic pipelines and variant filtering process vary between centres. More than 1500 genes encode the known mitochondrial proteome, and the interpretation of DNA variants in individual patients can be complex, especially if a novel variant is found in a known disease gene or in a gene not previously associated with mitochondrial disease (Calvo, Clauser et al. 2016, Wortmann, Mayr et al. 2017, Stenton and Prokisch 2018).

To interpret the pathogenicity of the variants, the American College of Medical Genetics and Genomics (ACMG) and the Association of Molecular Pathology (AMP) have published recommendations to guide the interpretation of variants and provide criteria to classify variants as benign, likely benign, likely pathogenic, or pathogenic based on the supporting evidence such as: frequency in population databases, segregation data, *in silico* predictions, functional studies and conservation among species (Richards, Aziz et al. 2015). More recently they have also included guidelines to interpret copy number variants (CNV) (Riggs, Andersen et al. 2020). However, none of these widely used guidelines consider the complexity of curating mtDNA variants. Approaches that address the challenges of interpreting some mtDNA variants include bioinformatic tools and online databases specific for mtDNA variants (Bris, Goudenege et al. 2018), and classification guidelines for mitochondrial DNA variants have been proposed by single groups (Wang, Schmitt et al. 2012, Wong, Chen et al. 2020). However, at the time of this thesis, consensus recommendations for the interpretation of mtDNA variants were not yet available, further complicating the diagnosis of mitochondrial diseases.

1.2.3 Genome sequencing for mitochondrial diseases

Exome sequencing is increasingly being implemented to diagnose patients with suspected mitochondrial disease (Stenton and Prokisch 2018). Advantages over GS include lower costs, less data storage, and easier bioinformatic processing, while still detecting most variants in coding regions in the nuclear DNA. However, to detect variants in the mitochondrial DNA, specific capture systems are often needed to isolate the mtDNA. Alternatively, some groups use off-target ES data for mtDNA analysis, with a major limitation of this approach being the lower depth and an inferior detection of heteroplasmy compared to using targeted mtDNA sequencing (Wagner, Berutti et al. 2019). Another option is to perform ES and separate targeted mtDNA MPS (Theunissen, Nguyen et al. 2018).

GS is a more comprehensive DNA test that allows detection of most types of genomic variants in both the mitochondrial and nuclear genomes. Because of associated costs and difficulties in data interpretation, GS has not yet been widely used in mitochondrial diseases but it has several advantages that make it a potentially optimal testing strategy. Firstly, GS can provide an unbiased approach useful in diseases that are genetically and clinically heterogenous such as mitochondrial diseases (Stenton and Prokisch 2018). Secondly, GS has the advantage of giving, with a single test, a good coverage of both the nuclear and mitochondrial genomes (Riley, Cowley et al. 2020). Finally, the data can be interrogated to detect single nucleotide variants (SNV), structural variants (SV), and more recently even to identify short tandem repeats (STR) (Tankard, Bennett et al. 2018).

1.2.4 Diagnostic yield using ES and GS in mitochondrial diseases

In recent years, there has been an increasing interest in evaluating the diagnostic utility of exome sequencing for mitochondrial diseases. However, before the submission of this thesis only Riley and colleagues have attempted to investigate the diagnostic utility of genome sequencing. This work comprises a cohort of patients analysed through trio GS and will be further discussed in Chapters 3-6 (Riley, Cowley et al. 2020).

The molecular diagnostic yields of cohorts studied with ES or GS ranged from 35-70% and are summarised in Figure 4. The disparate diagnostic rate between cohorts could be in part due to the differences in inclusion criteria. For instance, some groups performed genetic testing such

as mtDNA sequencing or gene panels prior to ES (Wortmann, Koolen et al. 2015, Legati, Reyes et al. 2016, Theunissen, Nguyen et al. 2018). Also, some of the groups included both adults and children (Legati, Reyes et al. 2016, Theunissen, Nguyen et al. 2018) while others were specific to childhood onset diseases (Haack, Haberberger et al. 2012, Taylor, Pyle et al. 2014, Wortmann, Koolen et al. 2015, Kohda, Tokuzawa et al. 2016, Pronicka, Piekutowska-Abramczuk et al. 2016, Puusepp, Reinson et al. 2018, Riley, Cowley et al. 2020). Finally, some cohorts included stringent biochemical evidence such as isolated Complex I deficiency (Haack, Haberberger et al. 2012) while others were broader such as the suspicion of a mitochondrial disease by a referring physician, resulting in a more heterogenous population (Wortmann, Koolen et al. 2015, Puusepp, Reinson et al. 2018).

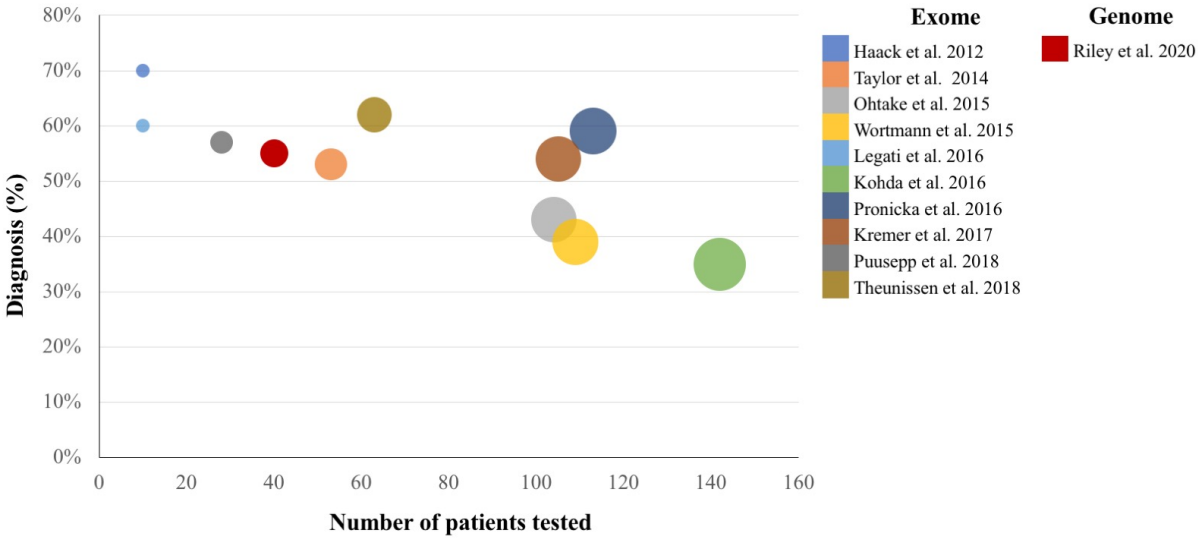

Figure 4 Diagnostic yield of ES and GS in mitochondrial disease cohorts.

The number of patients tested in each cohort is indicated on the X-axis and reflected in the relative size of each bubble. The percentage of patients with a suspected mitochondrial disease that achieved a molecular diagnosis by ES or GS is shown on the Y-axis. Cohorts using MPS panels were not included. Data from: (Haack, Haberberger et al. 2012, Ohtake, Murayama et al. 2014, Taylor, Pyle et al. 2014, Wortmann, Koolen et al. 2015, Kohda, Tokuzawa et al. 2016, Legati, Reyes et al. 2016, Pronicka, Piekutowska-Abramczuk et al. 2016, Kremer, Bader et al. 2017, Puusepp, Reinson et al. 2018, Theunissen, Nguyen et al. 2018, Riley, Cowley et al. 2020).

1.2.5 RNA sequencing for mitochondrial diseases

RNAseq is a method of massively parallel transcriptome sequencing which in recent years has been introduced as a potential tool in the diagnosis of mitochondrial diseases (Kremer, Bader et al. 2017). Kremer and colleagues, analysed 48 patients undiagnosed by ES and performed RNAseq using fibroblasts. The RNAseq analysis pipeline included detection of aberrant expression, splicing, and monoallelic expression; with this approach, a molecular diagnosis was achieved in 5 (10%) patients and detected candidate variants in the rest (Kremer, Bader et al. 2017).

These findings suggest that RNAseq could be a valuable tool to increase the diagnostic yield in patients with mitochondrial diseases when integrating to ES/GS. Nevertheless, an important limitation for RNAseq is the need for clinically relevant tissue as gene expression and splicing can be tissue-specific (Cummings, Marshall et al. 2017). However, 68% of genes listed in the online mendelian inheritance in man (OMIM) catalogue are expressed in fibroblasts (Kremer, Wortmann et al. 2018). Therefore, this cell line could be considered as an alternative when clinically-relevant tissue is not available.

1.3 Research Aims and Hypothesis

The central aim of this thesis is to study the use of genome sequencing to identify the molecular aetiology of patients with suspected mitochondrial disease.

To achieve this, two cohorts of patients with suspected mitochondrial disease but for whom a genetic diagnosis remained unknown were analysed through GS. These cohorts were a retrospective paediatric trio GS cohort, and a prospective “mitochondrial flagship” cohort, which included adults and children as outlined in Section 2.1. The patients were recruited from referring clinicians in Australia, were not previously studied through ES or GS and satisfied the inclusion/exclusion criteria outlined in Section 2.1.

Specific aims of this project are to:

a) Identify the genetic aetiology of patients suspected of mitochondrial disease through GS.

b) Identify novel genes and variants causative of mitochondrial disease using GS.

c) Determine the pathogenicity of the novel variants identified by GS in known or novel disease genes through an array of functional assays using patient material and cell lines.

d) Evaluate the utility of trio GS in the analysis of mtDNA inheritance in patients with suspected mitochondrial disease.
e) Compare the detection of causative variants in coding regions and known disease genes with singleton ES+mtDNAseq versus singleton GS.

In this project, it is hypothesised that GS is a useful tool to identify the genetic diagnosis of patients with suspected mitochondrial disease.

Chapter 2 Materials and Methods

2.1 Patients and cohorts

The work undertaken in this research project included patients from 2 cohort studies:

2.1.1 Trio GS cohort:

This study was funded by an NSW Ministry of Health Office for Health Research Genomics Collaborative Research grant.

- Inclusion criteria: Paediatric patients with a suspected mitochondrial disease.
- Exclusion criteria: Existing molecular diagnosis, previous exome or genome sequencing.
- Number of patients: 40 family trios.
- Molecular studies: DNA extracted from blood was studied through trio genome sequencing (GS) conducted in the KCCG (Garvan Institute, Sydney) as described in Section 2.3.
- Studies undertaken for this research project: re-curation of trio GS data from unsolved patients, performance of functional studies for novel variants, and analysis of mtDNA inheritance.

2.1.2 Prospective AGHA cohort:

This cohort is part of the Australian Genomics Health Alliance (AGHA) national health research project for the implementation of genomic testing.

- Inclusion criteria: Prospectively identified patients with probable (score 5–7) or definite (score 8–12) diagnosis of mitochondrial disease based on modified Nijmegen criteria Table 3 (Morava, van den Heuvel et al. 2006).
- Exclusion criteria: Existence of a molecular diagnosis, previous testing through exome, genome, panel or mtDNA sequencing, an indication that there is another non-mitochondrial disease diagnosis from other investigatory testing as defined by an intake review committee.
- Number of patients: 135 recruited patients. To date, 132 paediatric (n=78) and adult (n=54) patients have been sequenced.
- Molecular studies: patients were randomised to be studied through singleton ES+mtDNA MPS performed and analysed by VCGS as described in Section 2.2 or singleton GS conducted in the KCCG (Garvan Institute, Sydney) as described in Section 2.3.
- Studies undertaken for this research project: curation of singleton GS data from the patients randomised to GS and performance of functional studies for presumed novel pathogenic variants.

2.2 Exome sequencing

Exome sequencing was performed and analysed by VCGS. Samples were sequenced using Agilent SureSelect QXT Clinical Research Exome libraries and sequenced in an Illumina instrument, with a minimum of 90% of bases sequenced to 15x quality and target mean coverage of 100x.

Reads were aligned to the GRCh37 genome and variants were called within the target region (exons +/- 2bp) using using Cpipe (Sadedin, Dashnow et al. 2015).

2.3 Genome sequencing

Genome sequencing was performed at the KCCG (Garvan Institute, Sydney). Samples were sequenced on the Illumina HiSeq X Ten, with more than 75% of bases having Q30 base quality and mean coverage of > 30x. Reads were aligned to the b37d5 genome using Burrows-Wheeler Aligner (BWA) (Li and Durbin 2010), sorted by coordinate using Novosort (Novocraft Technologies), realigned and recalibrated using GATK to generate BAM files. BAM files were examined using the Integrated Genome Viewer 2.3 software (Broad Institute).

Variants were called using GATK HaplotypeCaller (DePristo, Banks et al. 2011), annotated using ENSEMBL's Variant Effect Predictor (McLaren, Gil et al. 2016) (v74) and converted to an SQLite database using Gemini (Paila, Chapman et al. 2013)(v0.17.2). Databases were imported into SEAVE (Gayevskiy, Roscioli et al. 2019) which was used to perform variant filtration and prioritization of the nDNA variants, while mitochondrial SNV and indels were identified using mity (Puttick, Kumar et al. 2019).

SV were studied using ClinSV (Minoche et al., in preparation) which detects SV using Lumpy (Layer, Chiang et al. 2014) and CNVnator (Abyzov, Urban et al. 2011) and filters rare SV against population allele frequencies from 500 healthy individuals from the Medical Genome Reference Bank (MGRB) cohort (Pinese, Lacaze et al. 2020). After SVs were prioritized population frequency was further analysed using population databases such as GnomAD-SV(Collins, Brand et al. 2020).

2.4 Variant prioritisation and classification

SEAVE was used to filter the GS data with a phenotype driven approach, first using the mitochondrial disease gene list (https://panelapp.agha.umccr.org/panels/203/) and then, when no causative variants were identified, then the analysis was expanded to a Mendeliome gene list that includes genes known to cause monogenic disease

(https://panelapp.agha.umccr.org/panels/137/). For the trio GS cohort in Chapter 3, the analysis included non-coding regions, and was also expanded to include the Mitocarta 2.0 gene list which includes known and candidate genes encoding mitochondrial proteins (Calvo, Clauser et al. 2016). For the singleton prospective cohort in Chapter 7 the analysis was limited to coding regions and known disease genes.

Candidate variants were discussed in a multidisciplinary team (MDT) meeting and classified as shown in Table 4 based on the American College of Medical Genetics and Genomics (ACMG) standards expanded to three categories of Class 3 variants (Richards, Aziz et al. 2015). For mtDNA variants, the classification was based on in-house mitochondrial DNA curation guidelines from the AGHA Mitochondrial Flagship. Online databases and resources used to classify the variants included in Table 10.

Table 4 Variant classification based on ACMG guidelines

Classification	Interpretation
Class 5	Pathogenic variant
Class 4	Likely pathogenic variant
Class 3a	Variant of uncertain significance with potential clinical relevance
Class 3b	Variant of uncertain significance
Class 3c	Variant of uncertain significance with low clinical relevance
Class 2	Likely Benign
Class 1	Benign

(Richards, Aziz et al. 2015)

2.5 Tissue culture

All fibroblast cell lines were tested for *Mycoplasma*. The control and patient cell lines were maintained in Dulbecco's Modified Eagle's Medium (DMEM) 3.7g/L NaHCO3 (HyClone #SH30022) supplemented with 10% Fetal Bovine Serum (FBS) and 1% Penicillin/Streptomycin, incubated at 37 °C in 5% CO_2.

Depending on the growth rate, semi-confluent cultures were passaged by washing the cells with Phosphate Buffer Solution (1X PBS, pH 7.4) and then incubating with 2ml of Trypsin/EDTA (0.025%) at 37 °C for ~2 min until cells were detached from the flask. DMEM

with FBS was used to neutralise the trypsin. Then, the cells were resuspended and seeded into new flasks.

Cell harvesting was performed by trypsinisation, the cell suspensions were centrifuged for 5 min at 300 g, washed twice with PBS and cell pellets were stored at -80 °C.

2.6 Cycloheximide treatment

To inhibit nonsense-mediated decay, fibroblasts at ~70% confluency were treated with 100 ng/µL cycloheximide (CHX, Sigma-Aldrich, #C6255) for 24 hrs before harvesting by trypsinisation (Lamande, Bateman et al. 1998, Hillman, Green et al. 2004).

2.7 Interferon treatment

To stimulate interferon stimulated genes, fibroblasts at ~70% confluency were incubated in the presence of 100 U/mL interferon α-2a (Roferon-A, Roche, Australia) in DMEM at 37 °C and 5% CO_2 for 24 hrs before harvesting by trypsinisation as per Section 2.5 (Meuwissen, Schot et al. 2016).

2.8 DNA extraction

DNA was extracted from cultured fibroblasts using QIAamp DNA Mini kit (Qiagen, #51304) following the manufacturer's instructions. The DNA was eluted in 100 µl of buffer AE and stored at -30°C. DNA was quantified using a NanoDrop 2000 spectrophotometer (ThermoFisher) and purity was regarded as acceptable if the absorbance ratio of A60/A280 was between 1.7 and 1.9.

2.9 RNA extraction

RNA was extracted from cultured fibroblasts using the RNeasy Plus Mini Kit (Qiagen, #74134). Peripheral blood samples were collected in PAXgene RNA tubes (Qiagen, #762165) and RNA was extracted using the PAXgene blood RNA kit (Qiagen, #762164) following the manufacturer's protocol.

RNA was quantified using a NanoDrop 2000 spectrophotometer (ThermoFisher) and purity was regarded as acceptable if the absorbance ratio of A260/A280 was between 1.9 and 2.1; RNA was stored at -80°C.

2.10 Complementary DNA (cDNA) synthesis

cDNA was synthesised by reverse transcription using the Superscript III first strand supermix kit (ThermoFisher, #18080400) following the manufacturer's instructions using random hexamers and RNA templates from 200-800 ng.

2.11 Primer design and Polymerase Chain Reaction (PCR)

Target specific primers for DNA and cDNA amplifications were designed using Primer-BLAST (Ye, Coulouris et al. 2012). De-salted 100 µM oligonucleotides were synthesised by Macrogen Inc. (Macrogen, South Korea).

PCRs were performed in 25 µl reactions using Platinum Hot Start PCR 2X Master Mix (ThermoFisher, #203203; Table 5). The thermal cycler reaction conditions were optimised for each experiment. The standard conditions included an initial denaturation at 94 °C for 2 min, followed by 25-35 cycles of denaturation at 94 °C for 30 seconds, annealing at ~55-60 °C (depending on primer Tm) for 30 seconds, and extension at 72°C for 1 min/kb, a final extension step at 72 °C was carried out for 10 min before terminating the reaction at 4 °C for indefinitely time.

Table 5 Standard PCR using Platinum Hot Start PCR 2X Master Mix

Component	For 25 µL reaction
Platinum Hot Start PCR 2X Master Mix	12.5 µL
Primer F (10 uM)	1 µL
Primer R (10 uM)	1 µL
Template	varies
Nuclease free water	to 25 µl

2.12 Agarose gel electrophoresis

0.5 – 2 % (w/v) agarose gels were prepared by dissolving agarose (Sigma, #A9539) in 1x Tris Acetate EDTA 1x buffer. SYBR Safe DNA stain (ThermoFisher, #S33102) was added at a 1:10,000 dilution. The gels were cast in BioRad agarose gel electrophoresis trays.

Eight microlitres of PCR reaction was mixed with 2 µL of loading dye (New England Biolabs, # B7025S) and loaded into each well, and in a parallel lane either 100 bp DNA molecular

weight ladder (ThermoFisher, #SM0243) or 1 kb DNA molecular weight ladder (ThermoFisher, #SM0313). Electrophoresis was conducted in a Sub-Cell GT cell (BioRad, #170-4401) at ~100 volts for 20-50 min or until the DNA ladder was sufficiently separated. PCR amplicons were visualised using UV light.

2.13 Purification and Sanger sequencing

To remove unused primers and nucleotides, the PCR products were purified using ExoSAP-IT (ThermoFisher, #78201) containing Exonuclease 1 and Shrimp Alkaline Phosphatase enzymes. In brief, 5 µl of the PCR reaction product was mixed with 2 µl of ExoSAP-IT, followed by 15 min of 37 °C incubation and 15 min at 80 °C to deactivate the reagent.

When more than one PCR product were amplified, the desired amplicon was cut out with a blade under UV light (Dark Reader DR-45M) and extracted from the agarose gel using the QIAquick Gel Extraction Kit (Qiagen, #28704) following the manufacturer's protocol.

The samples were then sequenced by the Sequencing Service and Development Platform at the Victorian Clinical Genetics Services (VCGS) using the 96-capillary ABI 3730 sequencer (Applied Biosystems). Data were analysed using the demo version of CodonCode Aligner (v6.0.2).

2.14 Quantitative PCR (qPCR)

qPCR was performed with AccuPower 2X Greenstar qPCR Master Mix (Bioneer Pacific, # K-6251-C) or with TaqMan Fast advanced master mix (ThermoFisher, #4444556), using the primers or TaqMan probes listed in Table 6 and Table 7 accordingly. Cycling was conducted on a LightCycler 480 (Roche) as per the manufacturer's protocol. The relative mRNA levels were calculated using the $2^{(-\Delta\Delta Ct)}$ method (Schmittgen and Livak 2008).

Table 6 TaqMan probes used for qPCR

Gene	TaqMan Gene Expression Assay ID
IFI27	Hs01086370_m1
IFI44L	Hs00199115_m1
IFIT1	Hs00356631_g1
ISG15	Hs00192713_m1
RSAD2	Hs01057264_m1
SIGLEC1	Hs00988063_m1
COX11	Hs01680112_mH
HPRT1	Hs03929096_g1
18S	Hs999999001_s1

Table 7 Primer pairs used for qPCR

Transcript ID	Forward	Reverse
ND6-1	AACCCTACTCCTAATCACATAACCT	CTGGTTGAACATTGTTTGTTGG
ND6-2	GGCTTAGAAGAAAACCCCACA	TAGTCCGTGCGAGAATAATGATG
COXI +COXII	GCTCATTCATTTCTCTAACAGCAG	GGCGTGATCATGAAAGGTG
COXIII +ATP6	TCGCCTTAATCCAAGCCTAC	CCTTTTTGGACAGGTGGTGT
GAPDH	CGCTCTCTGCTCCTCCTGTT	CCATGGTGTCTGAGCGATGT
ACTB	CCTGGCACCCAGCACAAT	GGGCCGGACTCGTCATAC

These primers were previously published in (Matilainen, Carroll et al. 2017).

2.15 RNA sequencing

RNA sequencing was performed by VCGS. In brief, libraries were prepared using a TruSeq Stranded mRNA library prep kit (Illumina, #RS-122-2101, RS-122-2102, and RS-122-2103).

Paired-end sequencing was performed on an Illumina HiSeq 2000 system (Illumina) with a minimum coverage of 100 million reads.

Raw sequencing data were processed by Guy Helman at the Translational Bioinformatics Group (MCRI) using Bpipe pipeline, FastQC and Trimmomatic for QC, trimming, and alignment (Sadedin, Pope et al. 2012, Bolger, Lohse et al. 2014). The sequences were aligned using STAR (v2.7.3a) (Dobin, Davis et al. 2013) to the b38 human reference genome. Quantification was performed using FeatureCounts from the Rsubread package (v1.34.7) (Liao, Smyth et al. 2019). Differential expression analysis was performed using DESeq2 (v1.25.9)(Love, Huber et al. 2014). Sashimi plots to visualize splicing events were constructed with ggsashimi (Garrido-Martin, Palumbo et al. 2018).

2.16 Protein extraction and quantification

Protein was extracted from cultured fibroblasts by resuspending pellets in radioimmunoprecipitation assay buffer (10 mM, Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS and 140 mM NaCl) with 1X protease inhibitor cocktail (ThermoFisher, #88266); lysates were then sonicated on ice with 3 pulses (0.3 sec/ cycle at 20-25% amplitude) using a Sonifier (Branson), and incubated on ice for 30 min. Subsequently lysates were centrifuged at 18,000 x g for 20 min at 4 °C. The supernatant was then transferred to a fresh tube and stored at -80 °C.

The total protein concentration was determined with the Pierce Bicinchoninic acid (BCA) kit (ThermoFisher, #23225). A set of protein standards was produced by diluting Pierce Bovine Serum Albumin Standard Ampules (ThermoFisher, #23209); each protein standard and unknown sample was added in duplicate to a microplate with the prepared BCA working reagent, and after a 30 min incubation at 37 °C, the absorbance was measured at 562 nm using the FLUOstar Optima Microplate Reader (BMG Labtech). A standard curve was used to determine the protein concentration of each sample.

2.17 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot

Protein lysates containing 10-40 μg of protein were dissolved in a small volume of 2X SDS-PAGE loading buffer (3.55 ml deionized water,1.25 ml 0.5 M Tris-HCl, pH 6.8, 2.5 ml glycerol, 2.0 ml 10% SDS, 0.2 ml 0.5% Bromophenol Blue) with 5% β- mercaptoethanol and heated at 95 °C for 5 min loaded onto 10-12 % handcast SDS-PAGE gels (BioRad system) alongside a Precision Plus Protein Standard (BioRad, #161-0374) and separated at 150 V for 50-90 min in running buffer (25 mM Tris-HCl, 192 mM glycine, 0.1% SDS). Then, proteins were transferred to polyvinylidene fluoride (PVDF) membranes (GE Healthcare, #10600061) using transfer buffer (25 mM Tris-HCl, 192 mM glycine, 20% methanol) at constant 350 mA for 90 min on ice.

Membranes were blocked with 5% skim milk in 1× PBS, 0.05% Tween 20, for 60 min at room temperature, then probed with the primary antibodies diluted in blocking solution overnight at 4 °C (Table 8).

Table 8 Primary antibodies and dilutions used for SDS-PAGE western blot

Antibody	ID	Dilution
GAPDH	Sigma #G9545	1:10,000
OXPHOS Human Cocktail (ATP5A, COXII, NDUFB8, SDHB, and UQCRC2)	Abcam #ab110411	1:500
PNPT1	Abcam #ab96176	1:500
VDAC1	Abcam #ab14734	1:5,000

Following antibody incubation, membranes were washed in 1× PBS, 0.05% Tween 20 five times for 5 min, then probed with the corresponding anti-mouse (GE Healthcare, #NA931) or anti-rabbit (Cell Signaling Technology, #7074s) IgG secondary-horseradish peroxidase conjugated antibody (1:5000) for 1 hour at room temperature. Membranes were then washed in 1× PBS, 0.05% Tween 20 five times for 5 min. The membranes were developed using ECL detection reagents (GE Healthcare, #GEHERPN2209) and exposed to Hyperfilm ECL (GE Healthcare, #28906836).

2.18 Blue Native (BN)-PAGE immunoblotting

BN-PAGE and immunoblot analysis were performed by Dr Luke Formosa (Monash University, Melbourne) using 30 µg of mitochondria isolated from fibroblasts and solubilized in 1% Digitonin or 1% Triton X-100 as previously described (Formosa, Muellner-Wong et al. 2020). The primary antibodies and dilutions used in the immunoblot analysis are listed in Table 9.

Table 9 Primary antibodies and dilutions used for BN-PAGE immunoblot

Antibody	ID	Dilution
Core-1 (UQCRC1)	ThermoFisher #459140	1:1000
ATP5A	Abcam #ab14748	1:1000
COX1	ThermoFisher #459600	1:1000
COX2	ThermoFisher #A-6404	1:1000
COX4	Abcam #ab110261	1:1000
NDUFA9	In house (Prof. Mike Ryan lab, Monash University)	1:500
SDHA	Abcam #ab14715	1:1000

2.19 Relative protein quantification

Protein bands intensities from Western blot films were quantitated using ImageJ software (National Institutes of Health), levels were normalised to the lane's loading control and expressed as percentage of controls.

2.20 Dipstick activity assays

Protein lysate samples were prepared using the extraction buffer from the enzyme activity dipstick assay kit (Abcam #ab109720) following the manufacturer's protocol. The protein concentration was measured using the Pierce BCA) kit (ThermoFisher, #23225) as previously described. Dipstick assays for complex I (Abcam, #ab109720), and complex IV (Abcam, #ab109876) enzyme activities were performed on 15-20 µg of protein following the manufacturer's protocol. Relative band intensities were quantified with a MS1000 Dipstick

Reader (Abcam, #ab117066) or scanned with Canon CS9000 flatbed scanner then analysed using ImageJ software (Schneider, Rasband et al. 2012).

2.21 Real-time ATP assay

The fluorescence resonance energy transfer (FRET)-based ATP assay was conducted by Dr Neal K. Bennett (Prof Nakamura's group, Gladstone Institute of Neurological Disease, USA) as previously published (Mendelsohn, Bennett et al. 2018). In brief, patient and control fibroblasts were immortalized using a lenti-hTERT-Neo Virus (ABMgood #G204) following the manufacturer's protocol. A FRET-based ATP sensor was transduced into fibroblasts, which were treated with either 50 μM CoQ10 or dimethylformamide-vehicle control for 5 days. The fibroblasts were then subjected to metabolic perturbations: restricted metabolism to respiration-only (PBS + 2% FBS + 10 mM pyruvate + 10 mM 2-deoxy-D-glucose), ATP depletion (PBS + 2% FBS + 5 μM oligomycin + 10 mM 2-deoxy-D-glucose), or unrestricted basal metabolism (PBS + 2% FBS + 5 mM pyruvate + 10 mM glucose) for 6 hours before measuring their ATP level by fluorescence-activated cell sorting.

2.22 Web resources

The following online resources were used for this PhD project:

Table 10 Web resources

Resource	Website
CADD (Combined Annotation-Dependent Depletion) score	https://cadd.gs.washington.edu/snv
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/
Genome Aggregation Database (gnomAD)	https://gnomad.broadinstitute.org
Genomizer	https://genomizer.com
Haplogrep	https://haplogrep.i- med.ac.at
Human splice finder	http://www.umd.be/HSF3
Matchmaker Exchange	https://www.matchmakerexchange.org
MitoBreak	http://mitobreak.portugene.com/

Mitochondrial Disease Sequence Data Resource MSeqDR	https://mseqdr.org
MitoFates	http://mitf.cbrc.jp/MitoFates/cgi-bin/top.cgi
MitoMap	https://www.mitomap.org/MITOMAP
Mutation Taster	https://www.mutationtaster.org
OMIM	https://www.omim.org
OMNI	https://ai-omni.com
Panelapp	https://panelapp.agha.umccr.org
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2
Primer-BLAST	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
PROVEAN (Protein Variation Effect Analyzer)	http://provean.jcvi.org/index.php
Seave	https://seave.bio
SIFT	https://sift.bii.a-star.edu.sg
SROOGLE	http://sroogle. tau.ac.il
TargetP 2.0	http://www.cbs.dtu.dk/services/TargetP/
TPpred 3.0	https://tppred3.biocomp.unibo.it/welcome/default/index
UCSC Genome Browser	https://genome.ucsc.edu/

Chapter 3 Use of Trio GS in patients with suspected mitochondrial diseases

3.1 Introduction

Studies of aggregate cohorts of patients with monogenic diseases suggest that trio sequencing has a higher diagnostic yield than singleton sequencing (Smith, Swint et al. 2019). One of the advantages of using trio-based GS testing is that it enables rapid segregation analysis, which is particularly advantageous for conditions that are known to be recessive or to identify *de novo* variants in one single test. The results of the first paediatric cohort of patients with suspected mitochondrial disease analysed by trio GS has been recently published by Riley and colleagues (Riley, Cowley et al. 2020). Cases from the cohort that were analysed as part of this PhD project were classified in 4 groups:

1. Diagnosis in novel candidate mitochondrial disease genes: *TEFM, COX11,* which will be discussed in this chapter.
2. Diagnosis in known non-mitochondrial disease genes: *NBAS*, discussed in Chapter 4.
3. Diagnosis in known mitochondrial disease genes: *PNPT1*, discussed in Chapter 5.
4. Unsolved cases: There were 3 patients with a negative GS, or where the functional studies did not support pathogenicity of the variants. We extracted RNA from fibroblasts to perform RNA sequencing (RNAseq). The RNAseq analysis is currently being conducted by Guy Helman and Dr Cas Simons (Translational bioinformatics, MCRI). In addition, GS data from an additional unsolved patient with epileptic encephalopathy and a movement disorder was sent to Dr Haloom Rafehi (Prof. Melanie Bahlo's lab, Walter and Eliza Hall Institute) for interrogation for STR variants.

RNA testing was used as a complementary tool in several cases from the trio GS cohort and the method of choice was individualized. For instance, a targeted approach using RT-PCR and cDNA sequencing was selected in the cases where a "single hit" in a gene with a strong phenotype match was identified, for instance the patients that will be described in Chapter 4 (*NBAS*) and in Chapter 5 (*PNPT1*). However, if there were zero, or multiple candidate variants without a strong phenotype match, then whole transcriptome RNAseq was preferred as in patient T1 in Section 3.2.1.

3.2 Trio GS aids in the diagnosis of mitochondrial diseases in novel disease genes.

3.2.1 Candidate gene: *TEFM*

A female patient (T1) from the trio GS cohort presented from the newborn period with hypertonia, vomiting, poor feeding and failure to thrive. At 4 weeks she was admitted to a neonatal intensive care unit (NICU) with myoclonic jerking. She later developed severe refractory myoclonic seizures, and severe global developmental delay. She also had ptosis, nystagmus, small optic discs and left esotropia. She died at 17 months of age.

The initial trio GS was inconclusive. To detect potential splicing defects or differences in gene expression levels, RNA sequencing (RNAseq) was performed on RNA extracted from fibroblasts as outlined in Section 2.9 and analysed by Guy Helman (Translational Bioinformatics research group, MCRI, Melbourne). RNAseq revealed an exon skipping event in *TEFM* that led to the exclusion of exon 2 in 59% of the transcripts. This was not observed in 24 unrelated in-house control fibroblasts (Figure 5). Re-analysis of the trio GS data, targeting *TEFM*, identified a paternally inherited intronic variant NM_024683.4:c.32-14A>G (
Figure 6). This variant occurs at a highly conserved base, and is absent from the gnomAD population allele frequency database. Consistent with the RNA sequencing data, the *in silico* splicing tools human splicing finder (HSF) (Desmet, Hamroun et al. 2009) and spliceAI (Jaganathan, Kyriazopoulou Panagiotopoulou et al. 2019) suggested a loss of the acceptor splice site at exon 2 in *TEFM* that would result in exon exclusion. To validate the RNAseq results, cDNA studies were conducted in fibroblasts with and without cycloheximide treatment as described in Section 2.6 and 2.10. By sequencing cDNA, exon 2 skipping was confirmed in the paternal allele as shown in Figure 7.

We also identified a maternally inherited 6 nucleotide deletion in *trans*, NM_024683.4:c.484_489delGAAAGA p.(Glu162_Arg163del), which was present in approximately 70% of reads covering exon 2, suggesting allelic expression imbalance (AEI) favouring the allele with the deletion. These residues have moderate conservation and are located in a connecting loop within the C-terminal domain of the protein (Hillen, Temiakov et al. 2018).

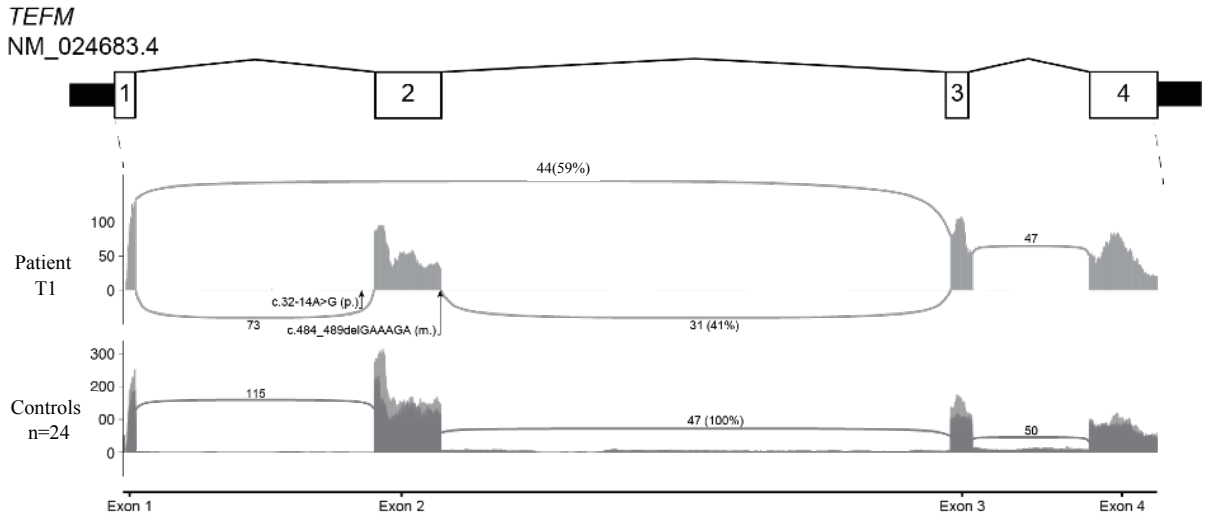

Figure 5 RNA sequencing detects a splicing defect in *TEFM*.

RNA sequencing data in the patient shows *TEFM* exon 2 skipping in 59% of the transcripts with the c.32-14A>G paternally inherited variant. RNAseq data analysis and Figure performed by Guy Helman (Translational Bioinformatics research group, MCRI, Melbourne) using ggsashimi (Garrido-Martin, Palumbo et al. 2018) .

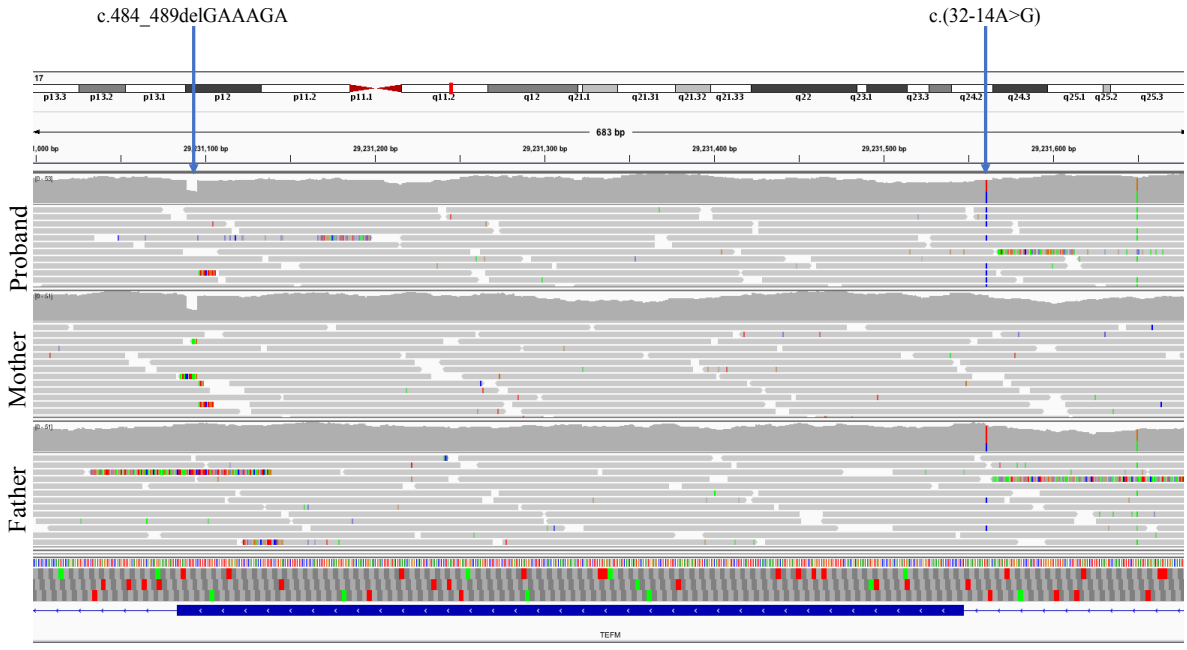

Figure 6 Trio GS detects compound heterozygous variants in *TEFM*.

IGV screenshots depict the NM_024683.4:c.484_489delGAAAGA variant inherited from the mother and the paternally inherited intronic variant NM_024683.4:c.32-14A>G.

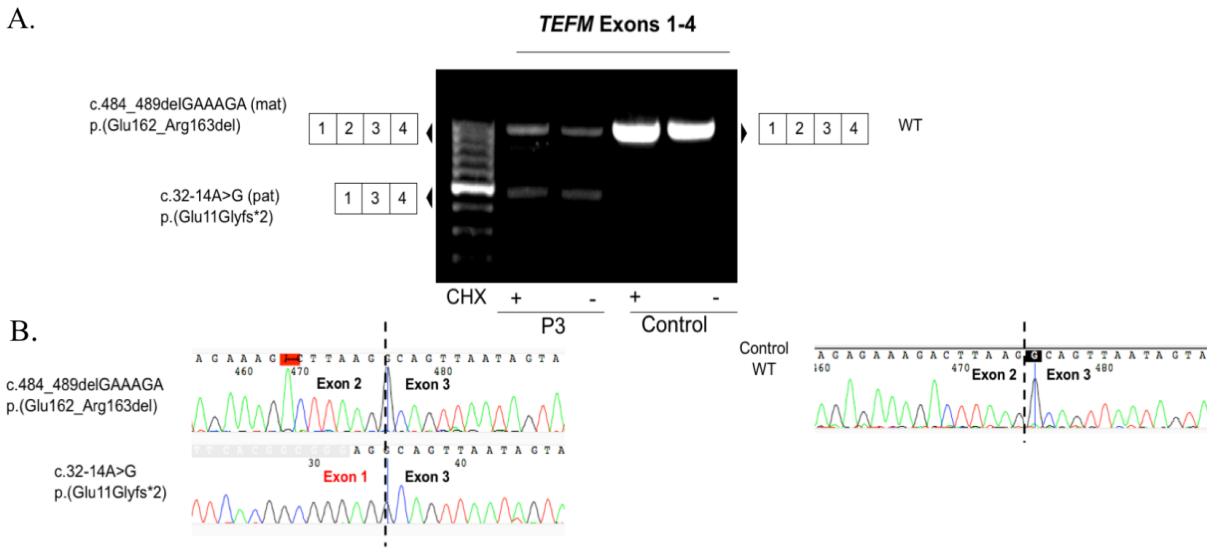

Figure 7 Characterisation of *TEFM* variants in cDNA.

(A) Gel electrophoresis of reverse transcription PCR (RT-PCR) products containing exons 1-4 shows a fragment of the expected size (956 bp) in both control and patient fibroblast cells, as well as an additional smaller fragment (~500 bp) consistent with skipping of exon 2, which is only present in the patient. (B) Sanger sequencing of the cDNA PCR amplicons confirmed exon 2 skipping in the paternal allele, and the 6nt deletion in the maternal allele.

TEFM encodes a mitochondrial transcription elongation factor, which has not been previously associated with human disease (Jiang, Koolmeister et al. 2019).

In order to assess the effect of the *TEFM* variants on OXPHOS proteins, Western blotting was performed using a cocktail of antibodies specific for one protein subunit of each OXPHOS complex. The results suggested moderately decreased levels of NDUFB8 (complex I) with normal levels of other OXPHOS subunits in fibroblasts (Figure 8 A and B), which is consistent with preliminary quantitative proteomic analyses performed by Daniella Hock (Bio 21, Melbourne) (Appendix A Figure T1). Activities of OXPHOS enzyme complexes I and IV were measured in fibroblasts by dipstick assays as described in Section 2.18. The residual activity of OXPHOS complex I was 65% of control mean, with no difference in the activity of complex IV (Figure 8 B).

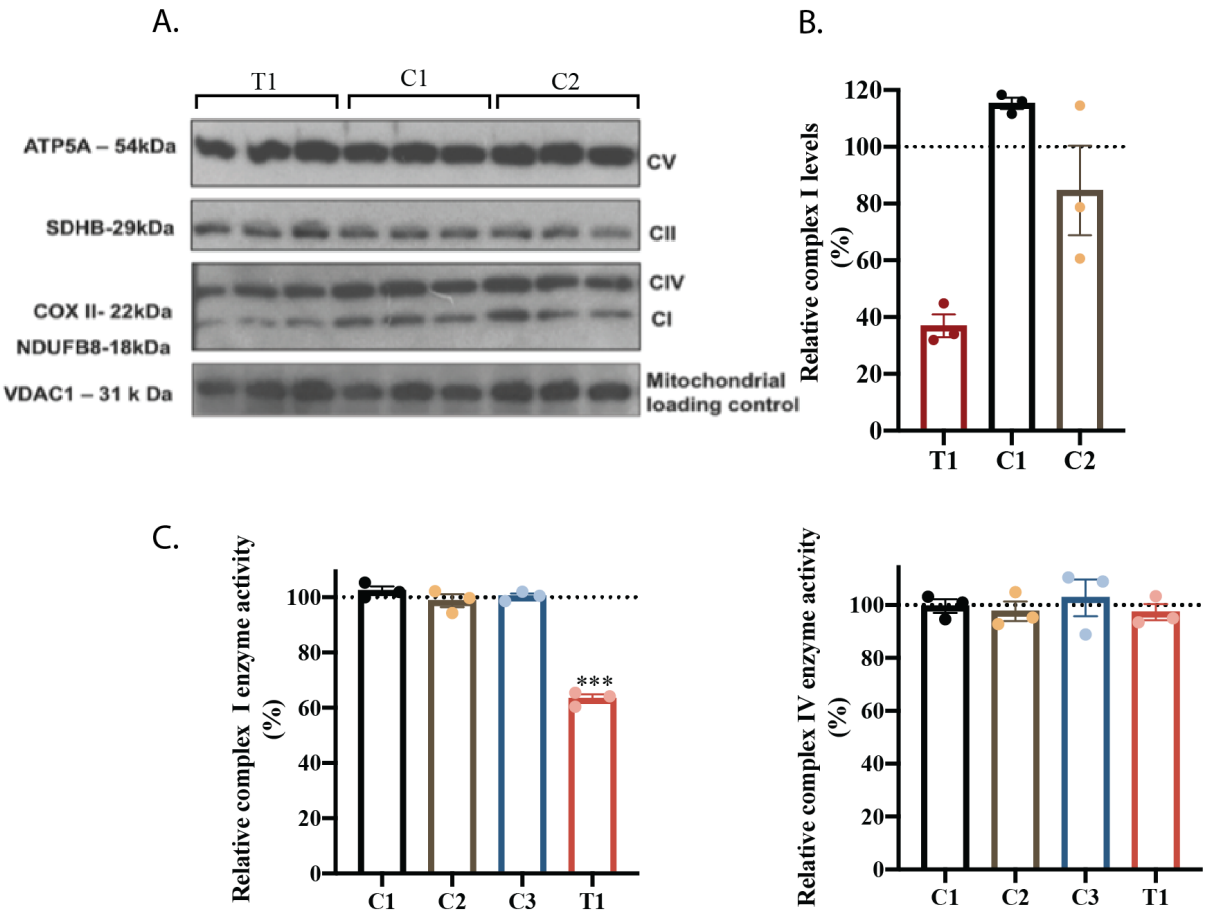

Figure 8 OXPHOS subunit and complex I and IV activities in fibroblasts from a patient with *TEFM* variants.

(A) Representative Western blot suggests moderately lower protein levels of NDUFB8 (CI) in the patient, with porin (VDAC1) used as a mitochondrial loading control. The experiment was performed twice with three cell pellets per cell line. (B) Densitometry analysis suggests 60% lower levels of NDUFB8 (CI) relative to controls. (C) Enzyme activity was analysed using immunocapture dipstick assays. Compared to controls T1 had a 65% residual complex I activity (left) and normal complex IV enzyme activity (right). The mean and variation (SEM) between three independent experiments are shown (*** $p < 0.001$).

To further characterize the functional consequences of the *TEFM* variants, proteomic studies are currently being conducted by Daniella Hock and Dr David Stroud (Bio21, Melbourne).

3.2.2 Candidate gene: *COX11*

A proband (X1) from the trio GS cohort was the first child of consanguineous parents. She presented with congenital sensorineural deafness and tonic–clonic seizures in the newborn period. By 5 months of age she regressed and had worsening of the seizure disorder. Brain MRI showed prominence of the middle cerebellar peduncles bilaterally, cerebral volume loss and hyperintense signals in the subcortical, deep and periventricular white matter in both cerebral hemispheres. Further investigations showed elevated CSF and blood lactate. She died at 9 months of age.

Spectrophotometric assay of OXPHOS enzyme activity in muscle measured by the Mitochondrial Diagnostic Service (MCRI/VCGS) showed all complexes were normal to elevated. In liver, complexes I, II and citrate synthase (CS) were elevated, so when the activities of complexes III and IV were expressed relative to Complex II or citrate synthase they were borderline low.

Trio GS identified a homozygous missense variant in *COX11*, NM_004375.4(*COX11*):c.730G>C p.(Ala244Pro), with both parents being heterozygous. Sanger sequencing in the patient's unaffected siblings showed they were both heterozygous for the NM_004375.4(*COX11*):c.730G>C p.(Ala244Pro) variant (Figure 9).

The alanine at position 244 has high conservation and is situated within the Cytochrome *c* oxidase assembly functional domain. *In silico* software predictions are conflicting (CADD, Polyphen2, SIFT, Mutation Taster) and the variant has not been observed in the gnomAD population database. *COX11* encodes a copper chaperone that is thought to participate in the assembly of complex IV by mediating copper delivery. *COX11* has not been previously associated with human mitochondrial disease (Bourens and Barrientos 2017).

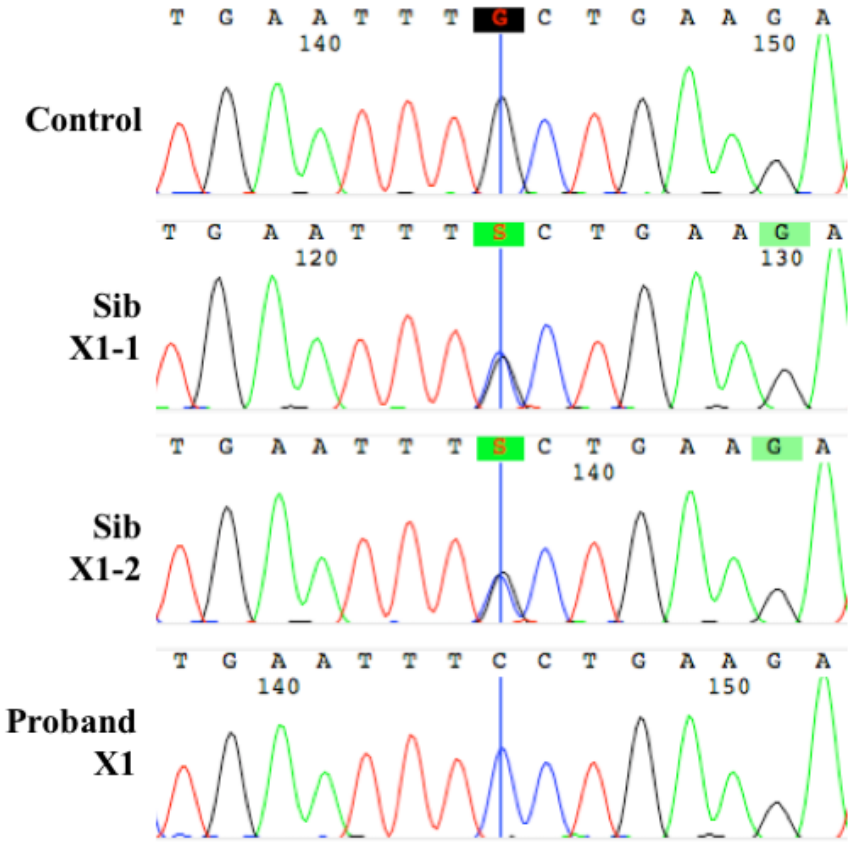

Figure 9 *COX11* Sanger sequencing in patient X1 and sibs.

Partial electropherogram showing Sanger sequencing results from DNA extracted from blood. The homozygous NM_004375.4(*COX11*):c.730G>C p.(Ala244Pro) variant identified in the proband X1 was heterozygous in the unaffected siblings (X1-1 and X1-2) and not present in control DNA.

To analyse the effect of this variant on CIV assembly, BN-PAGE immunoblot analysis was performed by Dr Luke Formosa (Monash University, Melbourne) in mitochondria isolated from fibroblasts from proband X1 and controls, solubilised in digitonin (to preserve the supercomplex form) or Triton X100 detergent (to separate the individual OXPHOS complexes into holoenzymes) (Formosa, Muellner-Wong et al. 2020). The results suggested a lower level of assembled monomeric CIV was more evident in the Triton X-100 eluates, consistent between COX2, COX4 (Figure 10 A) and COX1 antibodies (Figure 10 B), with normal levels of other OXPHOS subunits in isolated mitochondria compared to controls (Figure 10 C). The modest reduction in complex IV levels was consistent with the reduced levels of complex IV from analysis of whole cell lysates by SDS-PAGE western with COXII antibody (Figure 10 D).

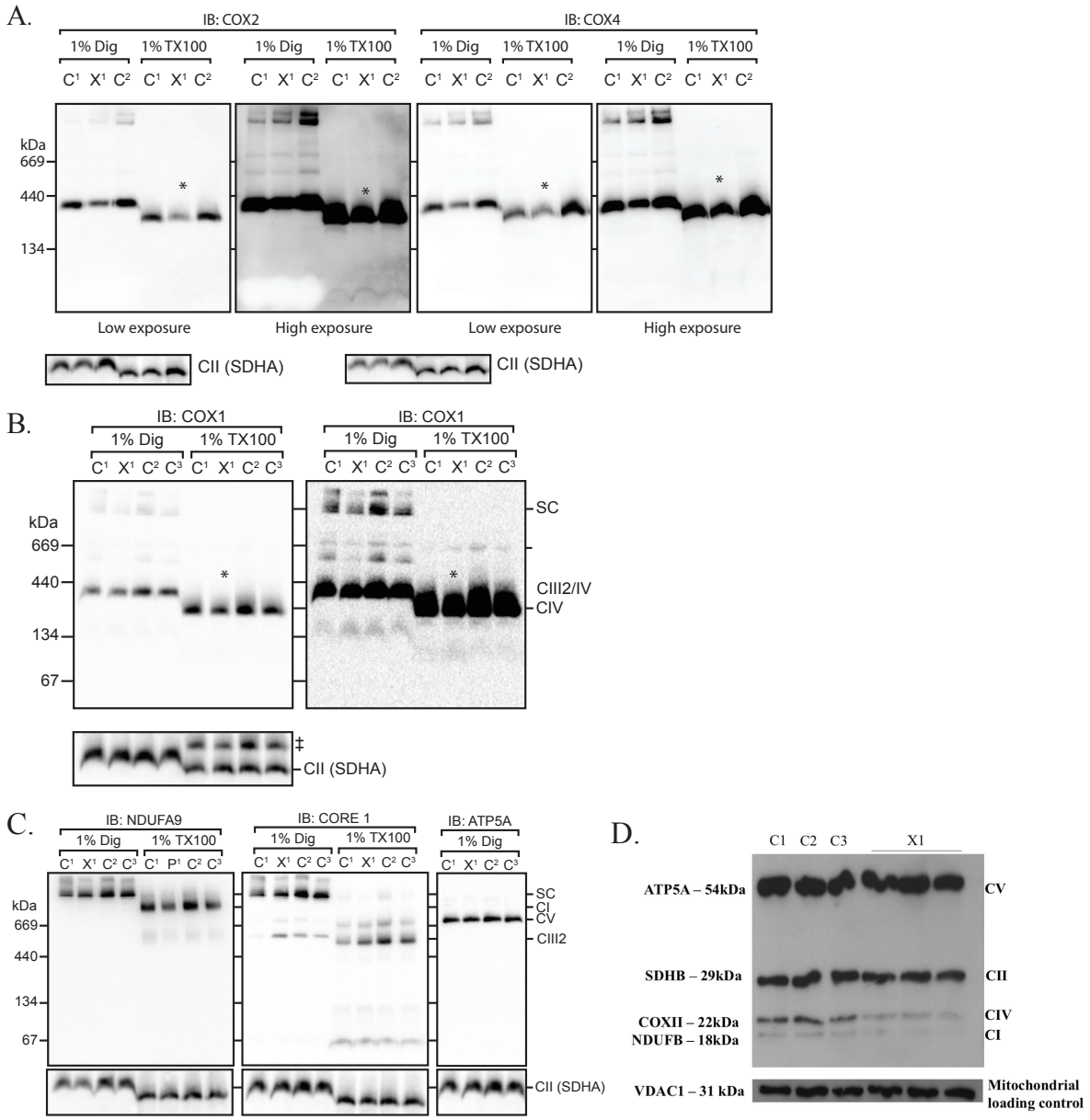

Figure 10 BN-PAGE for complex IV and OXPHOS Western blot in *COX11* fibroblasts.

Mitochondria isolated from Patient X1 and control fibroblasts analysed by BN-PAGE immunoblotting for COX2, COX4 (A) and COX2 (B) suggest a decreased level of monomeric complex IV more evident in mitochondria solubilised with Triton X-100, which separates the individual OXPHOS complexes (marked by *); ‡ indicates previous COX1 signal and complex II (SDHA) was used as a loading control. (C) The levels of other OXPHOS subunits in isolated mitochondria were similar to controls (CI- NDUFA9, CIII-CORE1, CV-ATP5A). BN-PAGE experiments and figures A-C were performed by Dr Luke Formosa (Monash University, Melbourne). (D) Representative SDS-PAGE and western blotting in whole cell lysates with an antibody cocktail targeting individual OXPHOS complex subunits showed consistently lower levels of complex IV in patient X1; porin (VDAC1) was used as a loading control.

Recently, Prof. Nakamura's group (Gladstone Institute, USA) screened the effect of individually knocking down 2,231 genes (including *COX11*) on ATP levels. They used Clustered Regularly Interspaced Short Palindromic Repeats interference (CRISPRi) and a (FRET)-based ATP sensor to study the effect of exposing these cells to different metabolic substrates. Under respiratory conditions, ATP levels were low when *COX11* was knocked down. Interestingly, when these cells were exposed to coenzyme Q_{10} (CoQ_{10}), they showed restoration of ATP levels (Mendelsohn, Bennett et al. 2018).

To assess if the NM_004375.4(*COX11*):c.730G>C p.(Ala244Pro) homozygous variant identified in patient X1 could have effects on ATP levels, our collaborators in Prof. Nakamura's group (Gladstone Institute, USA) studied patient X1 fibroblasts as described in Section 2.21.
Preliminary results suggested comparable ATP levels in patient X1 (*COX11*) and control cells under basal conditions, in the presence or absence of 50 μM CoQ_{10}. However, when glycolytic ATP generation was blocked with 2-deoxy-D-glucose, the patient X1 (COX11) fibroblasts showed markedly decreased ATP levels compared with control cells (Figure 11). In the presence of 50 μM CoQ_{10}, patient X1 (*COX11*) cells showed substantial restoration of ATP levels, while control cells showed no significant change (Figure 11).

These preliminary data suggest that the patient X1 (*COX11*) cells show a very similar functional defect to *COX11* knock-down cells in having inadequate ATP generation that can be rescued by CoQ_{10}. Replicate studies with additional controls are being conducted by our collaborators.

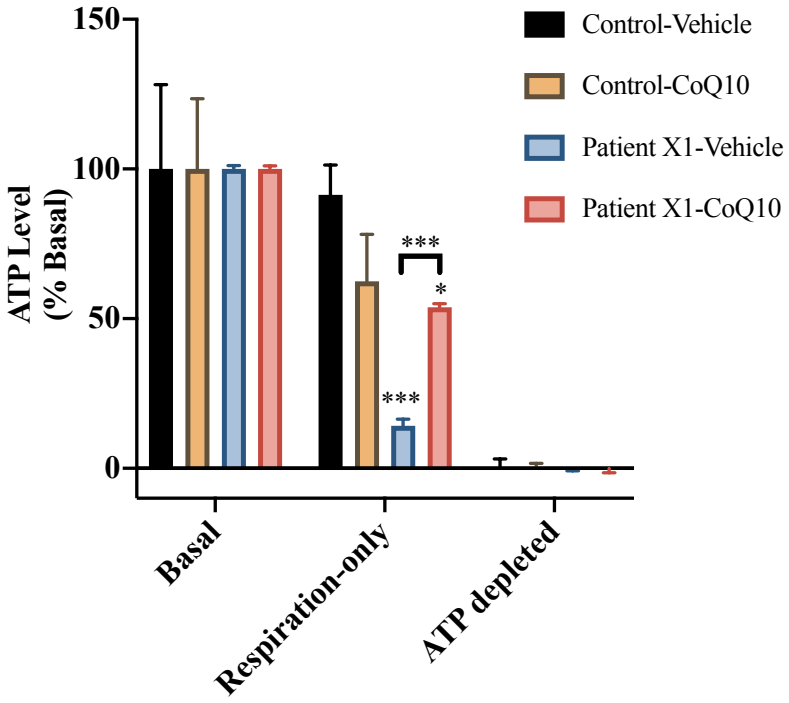

Figure 11 ATP Levels after 6 hrs of metabolic substrate in patient X1 (COX11) and control fibroblasts.

Patient XI (*COX11*) and control fibroblast ATP levels remained high with unrestricted basal metabolism. When metabolism was restricted to respiration-only by addition of 10mM 2-deoxy-D-glucose, ATP levels dropped in patient cells that were treated with the vehicle, while patient XI (*COX11*) cells maintained higher levels of ATP with CoQ_{10} treatment (***p<0.001, *< 0.05). The experiment and the figure were generated by Dr Neal Bennett (Prof. Nakamura's group, Gladstone Institute, USA).

An additional unrelated patient (X2) with feeding intolerance and high lactates was identified through Matchmaker Exchange (Philippakis, Azzariti et al. 2015). Trio exome sequencing performed at the Bioscientia Center for Human Genetics (Ingelheim, Germany) identified a homozygous variant in *COX11* NM_004375.4(*COX11*)c.35_36delinsG p.(Val12Glyfs*21) (Appendix A Figure X2).

RNA was extracted from total peripheral blood of patient X2 and reverse transcribed as described in Section 2.9 and Section 2.10. Quantitative PCR (qPCR) was performed using TaqMan gene expression assays as described in Section 2.14, showing no significant difference in *COX11* transcript expression compared to controls (Figure 12).

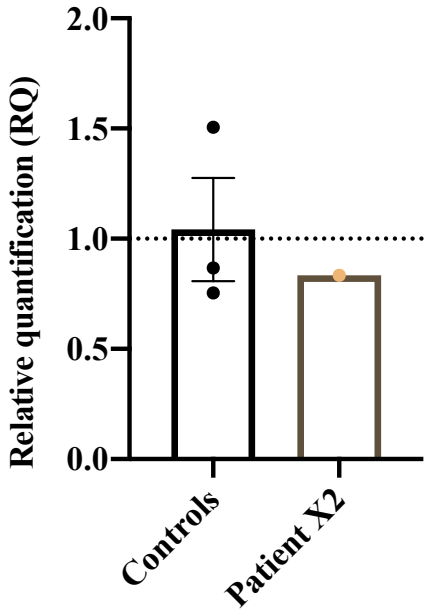

Figure 12 Relative expression of *COX11* in patient X2.

The relative expression of *COX11* mRNA extracted from blood from patient X2 was similar to controls (n=3). The expression level was normalised to *HPRT1* and *18S* and expressed relative to the average of controls. Samples were analysed from a single blood draw, with three technical replicates.

In addition, sequencing of cDNA identified the homozygous NM_004375.4(*COX11*)c.35_36delinsG variant in the transcript (Figure 13). Together, these data suggest that the mutant transcript is not degraded by nonsense mediated decay, and is therefore predicted to cause a truncated protein.

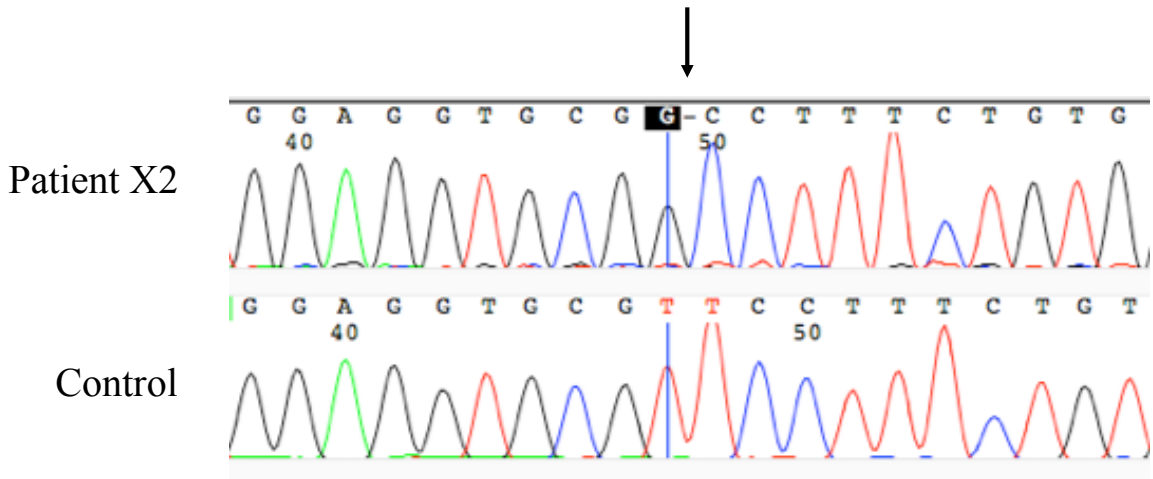

Figure 13 *COX11* cDNA sequencing in patient X2.

Partial electropherogram showing *COX11* cDNA Sanger sequencing from blood, showing a NM_004375.4(*COX11*)c.35_36delinsG p.(Val12Glyfs*21) homozygous variant in patient X2.

It seems likely that the truncating variant abolishes import of the protein into mitochondria as suggested by mitochondrial localization *in silico* prediction tools MitoFates (Fukasawa, Tsuji et al. 2015), and Tppred 3.0 (Savojardo, Martelli et al. 2015) (Table 11).

Table 11 *In silico* Predictions of the Mitochondrial Targeting Sequence

In silico tool	Predictions	COX11 WT	COX11 p.(Val12Glyfs*21)
MitoFates	Presequence predictions	Possessing mitochondrial presequence	No target sequence detected
	Probability for mitochondria localization (range 0-1)	0.996	0.112
	Predicted cleavage site	Amino acid 19	NA
Tppred 3.0	Protein localization prediction	Mitochondrion	No target sequence detected
	Predicted cleavage site	Amino acid 28	NA

NA= not applicable

While rare, variants affecting the mitochondrial targeting sequence have been described before in other genes causing mitochondrial disease such as *MLCYD*, *PDHA1* and *ISCA1,* leading to defects such as reduced mitochondrial protein import or protein mistargeting (Takakubo, Cartwright et al. 1995, Wightman, Santer et al. 2003, Torraco, Stehling et al. 2018). However, a cell line of patient X2 was not available and validation studies were not possible. Generating in vitro translated ^{35}S-labeled WT and COX11 mutants to perform import assays could be considered in the future to experimentally validate the impaired mitochondrial import predictions (Tucker, Mimaki et al. 2012, Priesnitz, Pfanner et al. 2020).

3.3 Conclusion

The identification of *TEFM* and *COX11* as two candidate novel disease genes associated with mitochondrial disorders highlights the utility of trio GS for gene discovery. Previously, Riley and colleagues identified two other novel mitochondrial disease genes in this trio GS cohort: *MECR* (Heimer, Keratar et al. 2016) and a gene that codes for a vitamin transporter that is

currently undergoing functional studies (Riley, Cowley et al. 2020). Together, this would bring the total new mitochondrial disease genes in the cohort to four.

The novel candidate genes identified in this cohort are consistent with autosomal recessive inheritance, in line with the most common pattern of paediatric onset mitochondrial disease (Gorman, Chinnery et al. 2016). Therefore, the main advantage for trio GS as opposed to a proband-only GS approach for gene discovery was the possibility to apply family filters to prioritize genes with biallelic variants.

However, even with trio GS, bioinformatic pipelines often do not prioritize variants in intronic regions. Complementing the GS analysis with RNAseq to prioritize variants in regions shown to have differential expression or splice defects was previously shown to be useful in discovering new mitochondrial disease genes such as *TIMMDC1* (Kremer, Bader et al. 2017). Using a similar approach, by combining trio GS with RNAseq we have identified *TEFM* as a novel candidate disease gene in patient T1 who had an initially inconclusive trio GS. In this case, the paired analysis allowed for detection and rapid segregation of the *TEFM* variants, including an intronic variant.

Further research is in progress to confirm *TEFM* and *COX11* as causative of mitochondrial disorders. To establish gene-disease associations in rare diseases, collaborative efforts to identify more cases and perform functional studies are needed. Tools such as Matchmaker exchange (Philippakis, Azzariti et al. 2015) have allowed us to identify more patients with variants in *TEFM* and *COX11* and to establish collaborations to perform further functional assays. Currently, studies are being conducted by our collaborators in three additional patients with candidate variants in *TEFM* that show similar clinical features to patient T1. To characterise the consequences of biallelic variants in *TEFM* our collaborators will perform an array of studies including analysis of the mitochondrial transcriptome.

Additionally, replicate studies on the effects of *COX11* variants on ATP levels are being conducted. Future research to elucidate the mechanism in which CoQ_{10} supplementation rescues the ATP reduction caused by *COX11* variants would be of interest, particularly since *COX11* does not have a known direct function in CoQ_{10} biosynthesis (Mendelsohn, Bennett et al. 2018). Further clarification of these mechanisms would be important to undertake

investigations on potential therapeutic effects of CoQ_{10} in patients with *COX11* variants (Mendelsohn, Bennett et al. 2018).

Chapter 4 Cryptic intronic *NBAS* variant reveals the genetic basis of recurrent liver failure in a child

4.1 Introduction

One of the many challenges in reaching a genetic diagnosis in mitochondrial diseases is that there are numerous disorders with phenotypic similarity to mitochondrial diseases (Parikh, Karaa et al. 2019).

This chapter contains a publication that illustrates how trio GS can be useful to diagnose those patients with phenotypic overlap with mitochondrial disease (Rius, Riley et al. 2019). A female patient from the trio GS cohort with recurrent acute liver failure was suspected to have mitochondrial disease. Liver failure can be a prominent feature in several mitochondrial diseases such as in the mtDNA depletion syndromes, while other non-mitochondrial genes such as *NBAS*, *LARS* and *IARS* can also be causative of monogenic liver failure (Parikh, Karaa et al. 2019). By combining trio GS and cDNA studies, biallelic variants in *NBAS* were identified, including a deep intronic variant resulting in the inclusion of a pseudo-exon.

Identifying the underlying genetic diagnosis in this patient had implications to guide treatment and to direct decisions regarding liver transplantation.

Molecular Genetics and Metabolism 126 (2019) 77–82

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

journal homepage: www.elsevier.com/locate/ymgme

Cryptic intronic *NBAS* variant reveals the genetic basis of recurrent liver failure in a child

Rocio Rius[a,b], Lisa G. Riley[c,d], Yiran Guo[e], Minal Menezes[c], Alison G. Compton[a,b], Nicole J. Van Bergen[a,b], Velimir Gayevskiy[f], Mark J. Cowley[f,g,h], Beryl B. Cummings[i], Louisa Adams[j], Carolyn Ellaway[d,j], David R. Thorburn[a,b], Hakon Hakonarson[e], John Christodoulou[a,b,*]

[a] Murdoch Children's Research Institute, Melbourne, Australia
[b] Department of Paediatrics, University of Melbourne, Melbourne, Australia
[c] Kids Research, The Children's Hospital at Westmead, Sydney, NSW, Australia
[d] Discipline of Child & Adolescent Health, Sydney Medical School, University of Sydney, Sydney, Australia
[e] Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, USA
[f] Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia
[g] St Vincent's Clinical School, UNSW Sydney, Sydney, Australia
[h] Children's Cancer Institute, Kensington, Australia
[i] Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, USA
[j] Western Sydney Genetics Program, Children's Hospital at Westmead, Sydney, NSW, Australia

ARTICLE INFO

Keywords:
Liver failure
Pseudo-exon
Deep-intronic
Whole Genome Sequencing

ABSTRACT

Background: In almost half of patients with acute liver failure the cause is unknown, making targeted treatment and decisions about liver transplantation a challenge. Monogenic disorders may contribute to a significant proportion of these undiagnosed patients, and so the incorporation of technologies such as next generation sequencing (NGS) in the clinic could aid in providing a definitive diagnosis. However, this technology may present a major challenge in interpretation of sequence variants, particularly those in non-coding regions.
Results: In this report we describe a case of Infantile liver failure syndrome 2 (ILFS2; MIM 616483) due to novel bi-allelic variants in the *NBAS* gene. A missense variant NM_015909.3(*NBAS*):c.2617C > T, NP_056993.2(NBAS):p.(Arg873Trp) was identified by whole genome sequencing (WGS). By combining WGS and reverse transcription-polymerase chain reaction (RT-PCR) we were able to identify a novel deep intronic variant, NM_015909.3(*NBAS*):c.2423 + 404G > C, leading to the inclusion of a pseudo-exon. This mechanism has not been described previously in this syndrome.
Conclusions: This study highlights the utility of analyzing NGS data in conjunction with investigating complementary DNA (cDNA) using techniques such as RT-PCR for detection of variants that otherwise would be likely to be missed in common NGS bioinformatic analysis pipelines. Combining these approaches, particularly when the phenotype match is strong, could lead to an increase in the diagnostic yield in acute liver failure and thus aid in targeted treatment, accurate genetic counseling and restoration of reproductive confidence.

1. Introduction

Pediatric acute liver failure is a complex clinical disorder in which rapidly progressive severe hepatic dysfunction presents in a child often without pre-existing liver disease. Patients can present with encephalopathy, ascites, seizures, coagulation abnormalities and elevated liver enzymes as a result of many different etiologies, including infections, exposure to toxins, autoimmune disorders, drug exposure, inherited metabolic and other genetic disorders [1]. Establishing a definitive diagnosis is very important as it impacts the management of these patients as well as their eligibility for possible liver transplantation [2].

The underlying cause of acute liver failure has not been determined in up to 49% of affected children [1]. It is likely that monogenic disorders could explain at least some of these undiagnosed cases [3]. Examples of monogenic disorders that present with acute liver failure in

* Corresponding author at: Brain and Mitochondrial Research Group, Murdoch Children's Research Institute, 50 Flemington Rd Parkville, 3052 Victoria, Australia.
E-mail address: john.christodoulou@mcri.edu.au (J. Christodoulou).

https://doi.org/10.1016/j.ymgme.2018.12.002
Received 5 November 2018; Received in revised form 5 December 2018; Accepted 5 December 2018
Available online 11 December 2018

children include galactosemia, tyrosinemia, Niemann-Pick C, Wilson disease, fructose intolerance, glycosylation defects, urea cycle defects, Infantile liver failure syndrome 2 (*NBAS* related) and mitochondrial diseases [1,4]. However, there is a clinical overlap between many of these disorders, making it difficult to clinically establish the diagnosis. The traditional diagnostic approach often comprises prolonged, expensive and intrusive diagnostic processes involving clinical evaluation, urine and blood biochemistry, tissue biopsies for histology and enzymology, followed by molecular genetic testing of candidate genes, which delays targeted treatment and decisions about the appropriateness of transplantation [5,6].

More recently, with the introduction of next generation sequencing (NGS) into healthcare, we have an opportunity to achieve a diagnosis in a more efficient and non-invasive manner. The utility of NGS for the diagnosis of other monogenic disorders has already been clearly established. When implemented early in the diagnostic journey in patients suspected to have a monogenic disorder, NGS can increase the diagnostic yield in a faster and cost-effective manner, as opposed to the traditional clinical and laboratory pathways, with some studies reporting a diagnostic yield of up to 60% using an NGS approach, about double that of the "standard" pathway [7,8].

As these technologies continue to become more accessible it is expected that the use of NGS will be increasingly implemented in the diagnostic approach to pediatric acute liver failure [3,6]. However, one of the main challenges of using NGS in the clinic includes the bioinformatic data analysis and interpretation. Most of the algorithms used to identify pathogenic variants do not prioritize those that are outside the protein-coding regions or exon-intron boundaries. Consequently, variants in deep intronic regions that could potentially cause splicing defects are easily overlooked [9]. To overcome this challenge, the analysis of complementary DNA (cDNA) or RNA can be conducted as an orthogonal strategy to assess the impact of candidate deep intronic variants on splicing [10]. In this paper we highlight the utility of WGS in combination with cDNA studies in a child with Infantile liver failure syndrome 2, in whom a novel disease mechanism for the *NBAS* (Neuroblastoma Amplified Sequence) gene was detected by combining both methodologies to identify a deep intronic variant leading to pseudoexon inclusion, *in trans* with a missense variant.

2. Materials and methods

2.1. Patient clinical summary

This girl was born at 38 weeks *via* emergency lower segment caesarean section for maternal pre-eclampsia to non-consanguineous parents of indigenous Australian background. Her birth weight was approximately 4 kg. She was bottle fed. There was no significant family history. She now has 5 siblings who are well.

Early growth and development were normal and she was fully immunised. From the age of thirteen months she began having episodes of fulminant liver failure, usually triggered by an intercurrent viral infection, resulting in transaminases being in the tens of thousands and liver synthetic function becoming deranged (INR up to 9.1 units; RR 1.0–1.2) as manifested by a severe coagulopathy, and associated with lactic acidosis and often hypoketotic hypoglycaemia. Blood ammonia was sometimes mildly elevated during these episodes (peaked at 340 during first admission and treated with sodium benzoate; RR 10–50 μmol/L). Other admissions had milder elevations up to 115 μmol/L.

On a number of occasions, the liver failure was so severe that she almost came to urgent liver transplant. However, with supportive dextrose-containing IV fluids she made a complete recovery over a period of 1–2 weeks each time. In between episodes, her acute hepatomegaly and liver function returned completely to normal, and blood lactate ranged from normal to being intermittently mildly elevated (1.6–3.4 mmol/L; RR 0.7–2.0 mmol/L). There was an impression that treatment with coenzyme Q at the time of acute liver crises shortened the duration of liver failure. She was consequently commenced on regular coenzyme Q and L-carnitine, with the frequency of episodes of liver failure reducing, although it was unclear whether these medications contributed to that or whether heightened awareness in the parents led to admissions earlier in the course of intercurrent illnesses, which seemed to reduce the severity and duration of liver failure (Fig. S1). Despite these episodes she continued to enjoy normal growth and development.

Formal cardiology assessment was unremarkable. Normal investigations include urine amino and organic acids, and plasma total and acylcarnitines, serum transferrin isoforms, alpha 1 antitrypsin levels (M2M2 phenotype), as well as fibroblast fatty acid oxidation studies. A liver biopsy during an acute episode showed only microvesicular steatosis.

Muscle and liver biopsies taken when well showed mild lipid accumulation but were otherwise unremarkable. Mitochondrial respiratory chain enzyme studies were performed as described previously [11]. The liver respiratory chain enzymology was normal, while the muscle results showed possible complex II + III deficiency (22% of normal relative to protein, 14% relative to citrate synthase and 15% relative to complex II). White blood cell coenzyme Q levels were normal.

Mutation testing for hereditary fructose intolerance was negative. No mtDNA deletions were detected in DNA extracted from liver.

When last reviewed at 9 years of age, her last presentation with acute liver failure was three years earlier. She had subsequently been well. The family actively avoid acetaminophen or any other potential liver toxin. They also try to avoid intercurrent illnesses, as these seemed to precipitate her events. She was otherwise growing and developing normally without any other medical issues.

3. Whole exome and genome sequencing and *in-silico* analyses

Genomic DNA was extracted from blood of the proband and parents and whole exome sequencing (WES) and primary bioinformatics analysis were performed as previously described [12]. WGS and variant calling were performed at the Kinghorn Centre for Clinical Genomics (Garvan Institute, Sydney) as described previously [13]. Variants from the MitoCarta 2.0 [14] and selected liver disease genes *POLG, TRMU, DGUOK, BCS1L, MPV17, RRM2B, TYMP, TP, LARS, CPT2, TWNK, NBAS* were prioritized using Seave [15].

In silico analyses of variants were performed using PolyPhen-2 [16] (http://genetics.bwh.harvard.edu/pph2/), SIFT [17] (http://sift.jcvi.org), CADD [18] (http://cadd.gs.washington.edu), MutationTaster [19] (http://www.mutationtaster.org/), Human Splicing Finder v3.1 [20] (http://www.umd.be/HSF3/), and SROOGLE [21] (http://sroogle.tau.ac.il). To determine allele frequencies the Genome Aggregation Database [22] (gnomAD; http://gnomad.broadinstitute.org) was used. Alignments and conservation analysis were conducted using the UCSC Genome Browser [23] (http://genome.ucsc.edu/).

4. Complementary DNA (cDNA) studies

Proband and control fibroblast cell lines were cultured at 37 °C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM). Cultured fibroblasts were grown with and without cycloheximide treatment (100 ng/μl) to inhibit nonsense-mediated decay (NMD) as previously described [24].

RNA was extracted using the miRNeasy Mini kit (QIAGEN) and cDNA was synthesized using the Superscript III kit (ThermoFisher) according to the manufacturer's protocol. cDNA was amplified by polymerase chain reaction (PCR) using 12 sets of primers designed to overlap the 52 exons of the *NBAS* sequence (Table S1) followed by Sanger sequencing.

5. Results

Homozygous and compound heterozygous variants in *NBAS* have been previously associated with ILFS2 and SOPH syndrome (Short stature, optic nerve atrophy, and Pelger-Huet anomaly, MIM 614800), the former having a strong phenotype match with the patient.

However, in the trio WES analysis, only one likely pathogenic variant was prioritized in *NBAS*. This variant, which segregated with the mother, was identified in exon 24. The missense variant NM_015909.3(*NBAS*):c.2617C > T, chr2:g.15557797G > A, NP_056993.2(NBAS):p.(Arg873Trp) is predicted to change an arginine to a tryptophan at amino acid position 873. The arginine at this position is highly conserved (Table S2), and is situated in the secretory pathway Sec39-like domain involved in vesicle transport from the endoplasmic reticulum to the Golgi apparatus [25]. *In silico* tools predict this variant to be disease-causing (Table S3). It has been observed in population databases at a frequency of 0.000008131 (2/245,964) but has not been reported in a homozygous state. It has not been previously reported to cause disease.

Also maternally inherited was a likely benign variant found in exon 38; the missense variant NM_015909.3 (*NBAS*):c.4475A > G, chr2:g.15467981 T > C, NP_056993.2(NBAS):p.(Tyr1492Cys) is predicted to change a tyrosine to a cysteine at amino acid position 1492. The tyrosine at this position is not conserved in mammals and is not situated in a known functional domain. *In silico* tools predict this variant to be tolerated. It has been observed in population databases at a frequency of 0.00000407 (1/245712) but not in a homozygous state. It has also not been previously reported to be associated with disease.

Because of the strong ILSF2 phenotype, we went on to perform trio WGS. As with the WES studies, in the initial analysis of WGS data we found only these two maternally inherited variants in *NBAS*. No candidate variants *in trans* were found in the coding region of the *NBAS* gene and a SNP array failed to detect any copy number variant (CNV) changes in the region (data not shown). Given the strong phenotype match, further analysis of the non-coding regions of *NBAS* and cDNA studies were conducted in order to determine whether a second pathogenic allele had been missed.

After re-analysis of the WGS data, a deep-intronic paternally inherited heterozygous variant was identified within intron 22. The NM_015909.3(*NBAS*):c.2423 + 404G > C, chr2:g.15567431C > G variant is located + 404 bp from the canonical splice-site of exon 22. This variant has been observed in population databases at a frequency of 0.00006460 (2/30958) but not in a homozygous state. It has not been previously reported to cause disease. *In silico* tools that combine multiple splicing regulatory element algorithms, Human Splicing Finder HSF v3.1 [20] and SROOGLE [21], suggested that the variant could disrupt a splicing silencer motif (Table S4 and Fig. S2).

Patient and control fibroblasts were grown with or without cycloheximide (which inhibits NMD) to assess the impact of this deep-intronic variant on splicing. As a consequence of the NM_015909.3(*NBAS*):c.2423 + 404G > C variant, a new 131 bp pseudo-exon containing intron 22 sequence was created and inserted into the transcript (chr2:15,567,351-15,567,481) (Fig. 1). The predicted consequence at the protein level is the insertion of 9 additional amino acids and the creation of a premature stop-codon p.(Arg809Phefs*10), with loss of over 1500C-terminal amino acids. The aberrantly spliced transcript was observed at low levels in patient cells in the absence of cycloheximide treatment, suggesting that the transcript undergoes NMD. In keeping with this, the c.2617C > T;p.(Arg873Trp) variant appeared homozygous in the cDNA from patient cells grown without cycloheximide as was the c.4475A > G;p.(Tyr1492Cys) variant. Interrogation of RNA-seq data of 151 control fibroblasts from the Genotype-Tissue Expression (GTEx) project did not show evidence of transcripts containing this pseudo-exon. No other abnormalities were detected by RT-PCR of the proband cDNA.

6. Discussion

The *NBAS* gene contains 52 exons and encodes a 2371 amino acid Neuroblastoma Amplified Sequence protein, which is involved in Golgi-to-endoplasmic reticulum (ER) retrograde transport [26]. It has been suggested that the secretory pathway Sec39-like domain is essential for this function and bi-allelic pathogenic variants in this region are associated with ILFS2 [25]. This disorder is inherited in an autosomal recessive manner and patients often present in infancy with recurrent episodes of acute liver failure associated with fever and with complete recovery between intervals [25]. This fits very closely with the clinical history of the patient described here.

Interestingly, patients with a homozygous missense p.(Arg1914His) variant in the C-terminal domain do not have acute liver failure, but have a different clinical presentation, SOPH syndrome [27]. This same variant has been reported in a compound heterozygous state in patients with all of the SOPH cardinal features including optic atrophy and a predominant skeletal phenotype, however they also presented with elevated liver transaminase levels associated with fever without developing acute liver failure [28,29]. Furthermore, it has been recently noticed that some patients with ILFS2 may have some subtle SOPH manifestations including short stature, and mild dysmorphic and skeletal features [25,27,30].

The patient reported here did not show any of the SOPH manifestations, and given that the allele harboring the deep intronic variant leads to NMD resulting in an 'apparent homozygosity' in cDNA of the p.(Arg873Trp) missense variant within the highly conserved Sec39-like domain, this is entirely in keeping with the ILFS2 presentation in this patient.

Intronic variants causing splicing defects have been reported in numerous diseases that can manifest with acute liver failure (Table 1). However, deep-intronic variants located at least 100 bp from the closest canonical splice-site have only been associated with acute liver failure in ornithine transcarbamoylase deficiency (MIM 311250) and Niemann-Pick type C (MIM 257220) (Table 1), even though deep-intronic variants have been identified in more than 76 other monogenic diseases [31,32].

Deep intronic variants in *NBAS* have not been described before, and to the best of our knowledge pseudo-exon activation by altering a regulatory splicing sequence has not been reported before in patients with monogenic acute liver failure. It remains to be seen if the lack of previous reports is due to the rarity of such events or because they have been missed by current diagnostic strategies. For instance, a recent report using a targeted amplicon-based NGS panel in children with suspected monogenic hepatopathies, identified only one pathogenic heterozygous variant in 7 patients in whom autosomal recessive inheritance was suspected, thus failing to provide a definite molecular diagnosis [33]. However, this study did not include any cDNA studies and only exon-intron boundaries were targeted for sequencing. Thus, it is cases like these that are good candidates for more in-depth cDNA studies or RNA sequencing (RNA-seq) in a bid to identify a missed second causative variant that could alter splicing.

It is therefore reasonable to speculate that the diagnostic yield in monogenic causes of acute liver failure may increase using the combination of WGS and cDNA analysis to detect deep intronic variants, as illustrated in this report. Since the first description of *NBAS* as a cause of acute liver failure in 2015 [4] more than 28 patients have been described [25,28,30,34–37], and we suggest that NGS in combination with cDNA studies should be implemented early in the diagnostic work up of children with recurrent acute liver failure, especially if it is associated with fever.

Establishing a diagnosis is important to direct treatment and most importantly decide if the patient would benefit from liver transplantation. The clinical decision making related to liver transplantation in pediatric acute liver failure is not standarised [38]. It requires a multidisciplinary team to make individualized decisions, balancing the

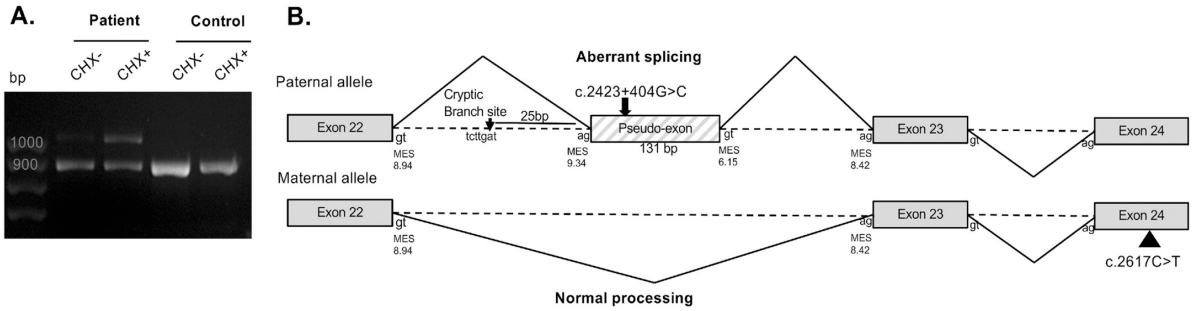

Fig. 1. A novel deep-intronic variant leads to the insertion of a pseudo-exon in *NBAS*.
A. Gel electrophoresis of RT-PCR products containing exons 18 to 25 shows a fragment of the expected size (880 bp) in both control and patient fibroblast cells as well as an additional larger fragment (~1000 bp) which is only present in the patient. This band is more prominent in the patient cells treated with cycloheximide (CHX +) to inhibit nonsense mediated decay (NMD) of the mutant transcript. B. Schematic representation of the aberrant splicing generated by the Paternally-inherited c.2423 + 404G > C deep-intronic variant, leading to inclusion of a 131 bp pseudo-exon between exons 22 and 23. The Maternally-inherited c.2617C > T;p.(Arg873Trp) variant is represented as a filled triangle and appeared homozygous in Patient cDNA generated from cells grown without cycloheximide, suggesting that the paternal transcript containing the pseudo-exon is degraded by NMD (Fig. S3). Maximum entropy scores (MES) above 3 are considered to be indicative of a splice-site (consensus values range − 20 to +20).

risks and benefits of performing a transplant, which is particularly challenging in the absence of a specific diagnosis. This is of particular concern if we take into account that acute liver failure is responsible for 11% of hepatic transplants in the USA [2], and almost half of the children with acute liver failure do not survive without this treatment [39]. In the child we described here, before the molecular diagnosis was made through WGS and cDNA studies, liver transplantation was seriously considered as a therapeutic option on several occasions. However, the patient recovered completely during episodes, making the transplant unnecessary. In addition, because of the recurrent acute liver failure, mild lactate elevation, and a low complex II + III enzyme activity in muscle there was a suspicion of mitochondrial disease as a differential diagnosis. This made the decision on liver transplantation more challenging, as patients with a multisystemic mitochondrial disease can have poor outcomes, and severe multiorgan mitochondrial disease is a contra-indication to liver transplant [2]. Fortunately,

Table 1
Selected intronic variants leading to splicing defects causing common monogenic diseases with pediatric acute liver failure.

Phenotype	Deep intronic pathogenic variants reported (ClinVar, HGMD)	Gene	Example of Variant	Consequence	References
Classic Galactosemia (MIM 230400)	No	*GALT*	c.1059 + 56C > T	54 bp insertion	[41]
Tyrosinemia type I (MIM 276700)	No	*FAH*	c.1062 + 5G > A	Skipping of exon 12	[42,43]
Hereditary fructose intolerance (MIM 229600)	No	*ALDOB*	c.-11 + 1G > C	Retention of intron 1	[44,45]
Carbamoylphosphate synthetase I deficiency (MIM 237300)	No	*CPS1*	c.3558 + 1G > C	Skipping of exon 29	[46]
Classic citrullinemia (MIM 215700)	No	*ASS1*	c.421-2A > G	Skipping of exon 7	[47]
Argininosuccinic aciduria (MIM 207900)	No	*ASL*	c.446 + 1G > A	Skipping of exon 5	[48]
Hyperornithinemia-hyperammonemia-homocitrullinuria syndrome (MIM 238970)	No	*SLC25A15*	c.56 + 1G > T	Skipping of exon 3	[49]
Wilson disease (MIM 277900)	No	*ATP7B*	c.1708-1G > C	Skipping of exon 5	[50]
Alpers-Huttenlocher syndrome (MIM 203700)	No	*POLG*	c.1251-2A > T	Skipping of exon 7	[51]
Mitochondrial DNA depletion syndrome 3 (hepatocerebral type) (MIM 251880)	No	*DGUOK*	c.444-62C > A	60 bp insertion	[52]
Mitochondrial DNA depletion syndrome 6 (hepatocerebral type) (MIM 256810)	No	*MPV17*	c.70 + 5G > A	Skipping of exon 2	[53]
Ornithine transcarbamylase deficiency (MIM 311250)	Yes	*OTC*	c.540 + 265G > A	Deep-intronic variant creates a new acceptor splice site leading to a 135 bp pseudo-exon inclusion	[54]
Niemann-Pick type C (MIM 257220)	Yes	*NPC1*	c.1554-1009G > A	Deep-intronic variant creates a new donor splice site leading to a 194 bp pseudo-exon inclusion	[55]
Infantile liver failure syndrome 2 (MIM 616483)	Yes	*NBAS*	c.2423 + 404G > C	Deep-intronic variant alters an intronic regulatory splice sequence leading to a 131 bp pseudo-exon inclusion	Present report

patients with ILFS2 tend to respond to antipyretics, high glucose and lipids during the acute episodes, and have a good overall prognosis with conservative treatment. Furthermore, when liver transplantation has been conducted during the acute episodes full recovery and no relapses have been reported [25,40].

7. Conclusion

The present study expands the spectrum of pathogenic variants causative of Infantile liver failure syndrome 2. By identifying a deep-intronic variant leading to inclusion of a pseudo-exon we describe a novel disease mechanism in the *NBAS* gene and highlight the utility of WGS and cDNA studies. We postulate that a similar approach could be conducted when only one pathogenic or likely pathogenic variant is found in a gene with a known autosomal recessive inheritance where there is a strong phenotype match. However, when no candidate variants are found, other technologies such as RNA-seq and proteomics could also aid in the diagnosis, as the current WGS bioinformatic analysis and *in silico* tools are currently inadequate for the easy identification of pathogenic deep intronic variants. This is an important area for future research as the use of NGS continues to expand and as more data are accumulated, further studies should be conducted to improve the current bioinformatic approaches to identify clinically relevant variants in non-coding regions.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymgme.2018.12.002.

Acknowledgements and funding

This research was supported by a New South Wales Office of Health and Medical Research Council Sydney Genomics Collaborative grant (JC), NHMRC project grant 1026891 (JC), NHMRC research fellowship 1102896 (DRT) and a NSW Health Early-Mid Career Fellowship (MJC). Whole exome sequencing, data analysis was done in the Center for Applied Genomics at the Children's Hospital of Philadelphia through research funding from Aevi Genomic Medicine Inc. We are grateful to the Crane and Perkins families for their generous financial support. The research conducted at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program.

References

[1] R.H. Squires Jr., B.L. Shneider, J. Bucuvalas, et al., Acute liver failure in children: the first 348 patients in the pediatric acute liver failure study group, J. Pediatr. 148 (5) (2006) 652–658.
[2] R.H. Squires, V. Ng, R. Romero, et al., Evaluation of the pediatric patient for liver transplantation: 2014 practice guideline by the American Association for the Study of Liver Diseases, American Society of Transplantation and the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition, Hepatology 60 (1) (2014) 362–398.
[3] S. Vilarinho, M. Choi, D. Jain, et al., Individual exome analysis in diagnosis and management of paediatric liver failure of indeterminate aetiology, J. Hepatol. 61 (5) (2014) 1056–1063.
[4] T.B. Haack, C. Staufner, M.G. Kopke, et al., Biallelic Mutations in NBAS Cause Recurrent Acute Liver Failure with Onset in Infancy, Am. J. Hum. Genet. 97 (1) (2015) 163–169.
[5] Y. Yang, D.M. Muzny, J.G. Reid, et al., Clinical whole-exome sequencing for the diagnosis of Mendelian disorders, N. Engl. J. Med. 369 (16) (2013) 1502–1511.
[6] E. Nicastro, L. D'Antiga, Next generation sequencing in pediatric hepatology and liver transplantation, Liver Transpl. 24 (2) (2018) 282–293.
[7] Z. Stark, D. Schofield, K. Alam, et al., Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement, Genet. Med. 19 (8) (2017) 867–874.
[8] Z. Stark, T.Y. Tan, B. Chong, et al., A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders, Genet. Med. 18 (11) (2016) 1090–1096.
[9] S. Gelfman, Q. Wang, K.M. McSweeney, et al., Annotating pathogenic non-coding variants in genic regions, Nat. Commun. 8 (1) (2017) 236.
[10] A. Anna, G. Monika, Splicing mutations in human genetic disorders: examples, detection, and confirmation, J. Appl. Genet. 59 (3) (2018) 253–268.
[11] A.E. Frazier, D.R. Thorburn, Biochemical analyses of the electron transport chain complexes by spectrophotometry, Methods Mol. Biol. 837 (2012) 49–62.
[12] Y. Guo, M.J. Menezes, M.P. Menezes, et al., Delayed diagnosis of congenital myasthenia due to associated mitochondrial enzyme defect, Neuromuscul. Disord. 25 (3) (2015) 257–261.
[13] G. Heimer, J.M. Keratar, L.G. Riley, et al., MECR mutations cause childhood-onset dystonia and optic atrophy, a mitochondrial fatty acid synthesis disorder, Am. J. Hum. Genet. 99 (6) (2016) 1229–1244.
[14] S.E. Calvo, K.R. Clauser, V.K. Mootha, MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins, Nucleic Acids Res. 44 (D1) (2016) D1251–D1257.
[15] V. Gayevskiy, T. Roscioli, M.E. Dinger, M.J. Cowley, J. Wren, Seave: a comprehensive web platform for storing and interrogating human genomic variation, Bioinformatics (2018), https://doi.org/10.1093/bioinformatics/bty540 (bty540).
[16] I.A. Adzhubei, S. Schmidt, L. Peshkin, et al., A method and server for predicting damaging missense mutations, Nat. Methods 7 (4) (2010) 248–249.
[17] Y. Choi, G.E. Sims, S. Murphy, J.R. Miller, A.P. Chan, Predicting the functional effect of amino acid substitutions and indels, PLoS One 7 (10) (2012) e46688.
[18] M. Kircher, D.M. Witten, P. Jain, B.J. O'Roak, G.M. Cooper, J. Shendure, A general framework for estimating the relative pathogenicity of human genetic variants, Nat. Genet. 46 (3) (2014) 310–315.
[19] J.M. Schwarz, C. Rodelsperger, M. Schuelke, D. Seelow, MutationTaster evaluates disease-causing potential of sequence alterations, Nat. Methods 7 (8) (2010) 575–576.
[20] F.O. Desmet, D. Hamroun, M. Lalande, G. Collod-Beroud, M. Claustres, C. Beroud, Human Splicing Finder: an online bioinformatics tool to predict splicing signals, Nucleic Acids Res. 37 (9) (2009) e67.
[21] S. Schwartz, E. Hall, G. Ast, SROOGLE: webserver for integrative, user-friendly visualization of splicing signals, Nucleic Acids Res. 37 (2009) W189–W192 Web Server issue.
[22] M. Lek, K.J. Karczewski, E.V. Minikel, et al., Analysis of protein-coding genetic variation in 60,706 humans, Nature 536 (7616) (2016) 285–291.
[23] W.J. Kent, C.W. Sugnet, T.S. Furey, et al., The human genome browser at UCSC, Genome Res. 12 (6) (2002) 996–1006.
[24] S.E. Calvo, E.J. Tucker, A.G. Compton, et al., High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency, Nat. Genet. 42 (10) (2010) 851–858.
[25] C. Staufner, T.B. Haack, M.G. Kopke, et al., Recurrent acute liver failure due to NBAS deficiency: phenotypic spectrum, disease mechanisms, and therapeutic concepts, J. Inherit. Metab. Dis. 39 (1) (2016) 3–16.
[26] T. Aoki, S. Ichimura, A. Itoh, et al., Identification of the neuroblastoma-amplified gene product as a component of the syntaxin 18 complex implicated in Golgi-to-endoplasmic reticulum retrograde transport, Mol. Biol. Cell 20 (11) (2009) 2639–2649.
[27] N. Maksimova, K. Hara, I. Nikolaeva, et al., Neuroblastoma amplified sequence gene is associated with a novel short stature syndrome characterised by optic nerve atrophy and Pelger-Huet anomaly, J. Med. Genet. 47 (8) (2010) 538–548.
[28] F. Kortum, I. Marquardt, M. Alawi, et al., Acute liver failure meets SOPH syndrome: a case report on an intermediate phenotype, Pediatrics 139 (1) (2017).
[29] M. Balasubramanian, J. Hurst, S. Brown, et al., Compound heterozygous variants in NBAS as a cause of atypical osteogenesis imperfecta, Bone 94 (2017) 65–74.
[30] F.S. Regateiro, S. Belkaya, N. Neves, et al., Recurrent elevated liver transaminases and acute liver failure in two siblings with novel bi-allelic mutations of NBAS, Eur. J. Med. Genet. 60 (8) (2017) 426–432.
[31] R. Vaz-Drago, N. Custodio, M. Carmo-Fonseca, Deep intronic mutations and human disease, Hum. Genet. 136 (9) (2017) 1093–1111.
[32] R.D. Bagnall, J. Ingles, M.E. Dinger, et al., Whole genome sequencing improves outcomes of genetic testing in patients with hypertrophic cardiomyopathy, J. Am. Coll. Cardiol. 72 (4) (2018) 419–429.
[33] A. Stalke, B. Skawran, B. Auber, et al., Diagnosis of monogenic liver diseases in childhood by next-generation sequencing, Clin. Genet. 93 (3) (2018) 665–670.
[34] J.Q. Li, Y.L. Qiu, J.Y. Gong, et al., Novel NBAS mutations and fever-related recurrent acute liver failure in Chinese children: a retrospective study, BMC Gastroenterol. 17 (1) (2017) 77.
[35] J. Wang, Z. Pu, Z. Lu, Targeted nextgeneration sequencing reveals two novel mutations of NBAS in a patient with infantile liver failure syndrome2, Mol. Med. Rep. 17 (2) (2018) 2245–2250.
[36] J.M. Capo-Chichi, C. Mehawej, V. Delague, et al., Neuroblastoma Amplified Sequence (NBAS) mutation in recurrent acute liver failure: Confirmatory report in a sibship with very early onset, osteoporosis and developmental delay, Eur. J. Med. Genet. 58 (12) (2015) 637–641.
[37] N.G. Segarra, D. Ballhausen, H. Crawford, et al., NBAS mutations cause a multisystem disorder involving bone, connective tissue, liver, immune system, and retina, Am. J. Med. Genet. A 167A (12) (2015) 2902–2912.
[38] J.E. Squires, D.A. Rudnick, R.M. Hardison, et al., Liver Transplant Listing in Pediatric Acute Liver Failure (PALF): practices and participant characteristics, Hepatology 6 (2018) 2338–2347.
[39] D.A. Rudnick, D.J. Dietzen, Y.P. Turmelle, et al., Serum alpha-NH-butyric acid may predict spontaneous survival in pediatric acute liver failure, Pediatr. Transplant. 13 (2) (2009) 223–230.
[40] P.L. Calvo, F. Tandoi, T.B. Haak, et al., NBAS mutations cause acute liver failure: when acetaminophen is not a culprit, Ital. J. Pediatr. 43 (1) (2017) 88.
[41] C. Wadelius, A. Lagerkvist, A.K. Molin, A. Larsson, U. von Dobeln, U. Pettersson, Galactosemia caused by a point mutation that activates cryptic donor splice site in the galactose-1-phosphate uridyltransferase gene, Genomics 17 (2) (1993) 525–526.
[42] R. Perez-Carro, R. Sanchez-Alcudia, B. Perez, et al., Functional analysis and in vitro

correction of splicing FAH mutations causing tyrosinemia type I, Clin. Genet. 86 (2) (2014) 167–171.
[43] M. Grompe, M. St. Louis, S.I. Demers, M. Al-Dhalimy, B. Leclerc, R.M. Tanguay, A single mutation of the fumarylacetoacetate hydrolase gene in French Canadians with hereditary tyrosinemia type I, N. Engl. J. Med. 6 (1994) 353.
[44] E.M. Coffee, D.R. Tolan, Mutations in the promoter region of the aldolase B gene that cause hereditary fructose intolerance, J. Inherit. Metab. Dis. 33 (6) (2010) 715–725.
[45] S. Bijarnia-Mahay, S. Movva, N. Gupta, et al., Molecular Diagnosis of Hereditary Fructose Intolerance: Founder Mutation in a Community from India, JIMD Rep. 19 (2015) 85–93.
[46] V. Klaus, T. Vermeulen, B. Minassian, et al., Highly variable clinical phenotype of carbamylphosphate synthetase 1 deficiency in one family: an effect of allelic variation in gene expression? Clin. Genet. 76 (3) (2009) 263–269.
[47] K. Kobayashi, H. Kakinoki, T. Fukushige, N. Shaheen, H. Terazono, T. Saheki, Nature and frequency of mutations in the argininosuccinate synthetase gene that cause classical citrullinemia, Hum. Genet. 96 (4) (1995) 454–463.
[48] M. Linnebank, E. Tschiedel, J. Haberle, et al., Argininosuccinate lyase (ASL) deficiency: mutation analysis in 27 patients and a completed structure of the human ASL gene, Hum. Genet. 111 (4–5) (2002) 350–359.
[49] A. Tessa, G. Fiermonte, C. Dionisi-Vici, et al., Identification of novel mutations in the SLC25A15 gene in hyperornithinemia-hyperammonemia-homocitrullinuria (HHH) syndrome: a clinical, molecular, and functional study, Hum. Mutat. 30 (5) (2009) 741–748.
[50] G.R. Thomas, J.R. Forbes, E.A. Roberts, J.M. Walshe, D.W. Cox, The Wilson disease gene: spectrum of mutations and their consequences, Nat. Genet. 9 (2) (1995) 210–217.
[51] A. Schaller, D. Hahn, C.B. Jackson, et al., Molecular and biochemical characterisation of a novel mutation in POLG associated with Alpers syndrome, BMC Neurol. 11 (4) (2011).
[52] N. Brahimi, M. Jambou, E. Sarzi, et al., The first founder DGUOK mutation associated with hepatocerebral mitochondrial DNA depletion syndrome, Mol. Genet. Metab. 97 (3) (2009) 221–226.
[53] A. Navarro-Sastre, E. Martin-Hernandez, Y. Campos, et al., Lethal hepatopathy and leukodystrophy caused by a novel mutation in MPV17 gene: description of an alternative MPV17 spliced form, Mol. Genet. Metab. 94 (2) (2008) 234–239.
[54] W. Ogino, Y. Takeshima, A. Nishiyama, et al., Mutation analysis of the ornithine transcarbamylase (OTC) gene in five Japanese OTC deficiency patients revealed two known and three novel mutations including a deep intronic mutation, Kobe J. Med. Sci. 53 (5) (2007) 229–240.
[55] L. Rodriguez-Pascau, M.J. Coll, L. Vilageliu, D. Grinberg, Antisense oligonucleotide treatment for a pseudoexon-generating mutation in the NPC1 gene causing Niemann-Pick type C disease, Hum. Mutat. 30 (11) (2009) E993–E1001.

The Supplementary data related to this article is included in Appendix B

Chapter 5 Clinical Spectrum and Functional Consequences associated with Biallelic *PNPT1* variants

5.1 Introduction

This chapter contains a publication that illustrates the utility of trio GS to identify novel variants in known mitochondrial disease genes, in this case, *PNPT1* (Rius, Van Bergen et al. 2019).

In the article, patient (P2) is from the trio GS cohort. He presented with multisystemic disease suggesting a mitochondrial disorder, but the initial analysis of trio GS data identified only a single missense variant in *PNPT1*. However, when the filters were relaxed to allow synonymous variants, an apparently synonymous variant was identified *in trans*. *In silico* and cDNA studies later confirmed that this variant leads to a splicing defect. Through our collaborators, we identified a further six individuals with candidate variants in *PNPT1* and conducted a clinical review of previously published patients.

The article also illustrates how functional studies are needed to confirm pathogenicity of novel variants. It is important to consider that often single *in vitro* assays may not fully explain the molecular effect leading to disease pathogenesis. For instance, detection of hypomorphic variants in functional assays can lead to false positives as the effects of the variant may not cause sufficient loss of function to be a penetrant pathogenic allele. This has been previously described by Lake and colleagues who identified homozygous nonsense variants in two genes, *PDHX* and *TIMMDC1*. The authors found that loss of the C-terminal of TIMMDC1 impacted modestly on complex I assembly, however it did not seem to be the cause of the patient's phenotype, and the pathogenic *PDHX* variant was consistent with this patient's clinical features (Lake, Formosa et al. 2019).

The interpretation of functional studies in context is therefore important as they can have different degrees of specificity. For instance, as it will be discussed in this chapter, normal OXPHOS enzymology does not always exclude pathogenicity of a variant in a mitochondrial disease gene. Conversely, abnormal OXPHOS enzymology can be caused by numerous disease genes and therefore as a single test may not be specific to validate pathogenicity of a variant in a single gene. Therefore, multiple lines of evidence for pathogenicity are often needed, as will be highlighted in this chapter, where the choice and interpretation of functional studies focused on the primary role of *PNPT1* in RNA processing (Matilainen, Carroll et al. 2017, Dhir, Dhir et al. 2018).

Journal of
Clinical Medicine

MDPI

Article

Clinical Spectrum and Functional Consequences Associated with Bi-Allelic Pathogenic *PNPT1* Variants

Rocio Rius [1,2], Nicole J. Van Bergen [1,2], Alison G. Compton [1,2], Lisa G. Riley [3,4], Maina P. Kava [5,6], Shanti Balasubramaniam [6,7,8], David J. Amor [1,2,9], Miriam Fanjul-Fernandez [2,9], Mark J. Cowley [10,11,12], Michael C. Fahey [13], Mary K. Koenig [14], Gregory M. Enns [15], Simon Sadedin [1,9], Meredith J. Wilson [16,17], Tiong Y. Tan [1,2,9], David R. Thorburn [1,2,9] and John Christodoulou [1,2,4,9,*]

1 Murdoch Children's Research Institute, Melbourne, VIC 3052, Australia; rocio.rius@mcri.edu.au (R.R.); nicole.vanbergen@mcri.edu.au (N.J.V.B.); alison.compton@mcri.edu.au (A.G.C.); david.amor@mcri.edu.au (D.J.A.); simon.sadedin@vcgs.org.au (S.S.); tiong.tan@vcgs.org.au (T.Y.T.); david.thorburn@mcri.edu.au (D.R.T.)
2 Department of Paediatrics, University of Melbourne, Melbourne, VIC 3052, Australia; miriam.fanjul@vcgs.org.au
3 Kids Research, The Children's Hospital at Westmead, Sydney, NSW 2145, Australia; lisa.riley@health.nsw.gov.au
4 Discipline of Child & Adolescent Health, Sydney Medical School, University of Sydney, Sydney, NSW 2050, Australia
5 Department of Neurology, Perth Children's Hospital, Perth, WA 6009, Australia; maina.kava@health.wa.gov.au
6 Department of Metabolic Medicine and Rheumatology, Perth Children's Hospital, Perth, WA 6009, Australia; shanti.balasubramaniam@health.nsw.gov.au
7 Genetic Metabolic Disorders Service, Western Sydney Genetics Program, The Children's Hospital at Westmead, Sydney, NSW 2145, Australia
8 Discipline of Genetic Medicine, Sydney Medical School, University of Sydney, Sydney, NSW 2145, Australia
9 Victorian Clinical Genetic Services, Melbourne, VIC 3052, Australia
10 Precision Medicine Theme, Children's Cancer Institute, Kensington, NSW 2750, Australia; MCowley@ccia.org.au
11 Kinghorn Centre for Clinical Genomics, Garvan Institute, University of New South Wales, Randwick, NSW 2010, Australia
12 School of Women's and Children's Health, University of New South Wales, Randwick, NSW 2031, Australia
13 Department of Paediatrics, Monash University, Melbourne, VIC 3168, Australia; michael.fahey@monash.edu
14 The University of Texas McGovern Medical School, Houston, TX 77030, USA; Mary.K.Koenig@uth.tmc.edu
15 Department of Pediatrics, Division of Medical Genetics, Stanford University, Stanford, CA 94305, USA; genns@stanford.edu
16 Department of Clinical Genetics, The Children's Hospital at Westmead, Sydney, NSW 2145, Australia; meredith.wilson@health.nsw.gov.au
17 Discipline of Genomic Medicine, Faculty of Medicine and Health, University of Sydney, Sydney, NSW 2006, Australia
* Correspondence: john.christodoulou@mcri.edu.au; Tel.: +613-9936-6353

Received: 9 October 2019; Accepted: 14 November 2019; Published: 19 November 2019

Abstract: *PNPT1* (PNPase—polynucleotide phosphorylase) is involved in multiple RNA processing functions in the mitochondria. Bi-allelic pathogenic *PNPT1* variants cause heterogeneous clinical phenotypes affecting multiple organs without any established genotype–phenotype correlations. Defects in PNPase can cause variable combined respiratory chain complex defects. Recently, it has been suggested that PNPase can lead to activation of an innate immune response. To better understand the clinical and molecular spectrum of patients with bi-allelic *PNPT1* variants, we captured detailed clinical and molecular phenotypes of all 17 patients reported in the literature, plus seven new patients, including a 78-year-old male with the longest reported survival. A functional follow-up of genomic

sequencing by cDNA studies confirmed a splicing defect in a novel, apparently synonymous, variant. Patient fibroblasts showed an accumulation of mitochondrial unprocessed *PNPT1* transcripts, while blood showed an increased interferon response. Our findings suggest that functional analyses of the RNA processing function of PNPase are more sensitive than testing downstream defects in oxidative phosphorylation (OXPHPOS) enzyme activities. This research extends our knowledge of the clinical and functional consequences of bi-allelic pathogenic *PNPT1* variants that may guide management and further efforts into understanding the pathophysiological mechanisms for therapeutic development.

Keywords: mitochondrial; *PNPT1*; PNPase; interferon; OXPHOS; respiratory chain; mutation; splice defect

1. Introduction

PNPT1 encodes for polynucleotide phosphorylase (PNPase), a conserved homotrimeric 3′-to-5′ exoribonuclease predominantly localized in the mitochondrial matrix and intermembrane space [1]. It is primarily involved in mitochondrial RNA (mtRNA) processing and degradation [2].

PNPase has been suggested to play a role in RNA import into mitochondria [3,4]; however, experimental data have been contradictory and, to date, there is no general agreement about an RNA import mechanism [5]. Recent reports suggest that disrupted PNPase RNA processing could lead to the accumulation of double-stranded mtRNAs, with the possibility of triggering an altered immune response [6,7].

Patients with bi-allelic *PNPT1* pathogenic variants have shown wide clinical heterogeneity ranging from non-syndromic hearing loss to multisystemic Leigh syndrome [8,9]. To date, no clear phenotype–genotype correlations have been drawn.

In a bid to better understand the clinical phenotype and functional consequences of patients with *PNPT1*-related diseases, we reported seven new patients with bi-allelic *PNPT1* variants and expanded upon the mutational spectrum. We also conducted functional studies and performed a thorough clinical review of previously published patients with *PNPT1* variants [6,8–13].

2. Experimental Section

2.1. Patients

We report seven new patients (P1, 2, 3, 3.2, 4, 7, 8) with bi-allelic variants in *PNPT1*, and provided updates for three previously published cases (P5, 6, 9). We also conducted a PubMed search using the terms PNPase (All Fields) OR (PNPT1 (All Fields) AND (“persons” (MeSH Terms) OR “persons” (All Fields) OR “individual” (All Fields))) up to September 2019 for publications limited to human subjects to describe the clinical features of all the patients with bi-allelic *PNPT1* pathogenic variants reported in the literature to date [6,8–13].

Functional studies were conducted in samples collected from four patients (P1, 2, 3, 4) in which the *PNPT1* variants were identified by whole genome sequencing (WGS; Garvan Institute, Sydney) or whole exome sequencing (WES; Victorian Clinical Genetics Services (VCGS), Melbourne; Broad Institute, Cambridge, MA, USA; and Baylor College of Medicine, Houston, TX, USA).

This study was performed in accordance with the Helsinki Declaration and ethical standards of the responsible ethics committees. The project was approved by the Human Research Ethics Committees of the Sydney Children’s Hospitals Network (ID number HREC/10/CHW/114), Melbourne Health (ID number HREC/16/MH/251), and the Royal Children’s Hospital (ID number HREC/16/RCHM/150).

2.2. Next Generation Sequencing (NGS) and in Silico Tools

The variants in P1 were identified through whole exome sequencing (WES) performed by the Genomics Platform at the Broad Institute of Harvard and MIT (Broad Institute, Cambridge, MA, USA). The variants in P2 were identified through trio whole genome sequencing (WGS) performed at the Kinghorn Centre for Clinical Genomics (Garvan Institute, Sydney) as previously described [14]. The variants in P3 and P4 were identified through WES performed at Victorian Clinical Genetics Services (VCGS), Melbourne. P7 and P8 variants were identified through WES performed at Baylor College of Medicine, Medical Genetics Laboratories, Whole Genome Laboratory (Houston, TX, USA).

In silico prediction analyses were performed using PolyPhen-2 [15], SIFT [16], Combined Annotation Dependent Depletion CADD [17], MutationTaster [18], and Human Splicing Finder v3.1 [19]. Visualization of variants in the Pfam [20] protein domains was conducted with MutationMapper [21]. Allele frequencies were determined using the Genome Aggregation Database [22].

2.3. Western Blotting

Fibroblast protein extraction and Western blotting were performed using total Abcam OXPHOS human WB antibody cocktail (ab11041) and PNPase (ab96176) as previously published [10].

2.4. Mitochondrial Oxidative Phosphorylation (OXPHOS) Enzyme Activities

Spectrophotometric analysis of OXPHOS enzyme activities in muscle and fibroblasts was performed as previously described [23]. The mitochondrial respiratory chain complex I (CI) and complex IV (CIV) dipstick activity assays (Abcam, Melbourne, VIC, Australia) were performed using 15 µg of whole-cell lysates fibroblasts as previously published [10].

2.5. Fibroblast Culture, RNA Extraction, and Complementary DNA (cDNA) Studies

Cycloheximide treatment of cultured fibroblasts from P2 and controls was performed as published [24]. Cultured fibroblasts from patients P1, P2, P3, and P4, and three control lines were incubated in the presence or absence of 100 U/mL interferon α-2a Roferon-A (Roche, Sydney, Australia) in HyClone Dulbecco's Modified Eagle Medium (GE Healthcare, Rydalmere, NSW, Australia) at 37 °C and 5% CO_2 for 24 h.

RNA was isolated from fibroblasts using the RNeasy Plus kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Reverse transcription was performed using SuperScript III First-Strand synthesis kit (Thermo Fisher Scientific, Carlsbad, CA, USA) following the manufacturer's instructions.

2.6. RNA Extraction from Blood

PAXgene blood RNA tubes (PreAnalytix by Qiagen, Hombrechtikon, Switzerland) were used to collect peripheral blood samples. After collection, the tubes were left at room temperature between 2 h and 72 h before extracting RNA using the PAXgene blood RNA kit (PreAnalytix by Qiagen) following the manufacturer's instructions.

2.7. PCR Quantification of Unprocessed Mitochondrial Transcripts

qPCR for the quantification of unprocessed transcripts was performed with AccuPower 2X Greenstar qPCR Master Mix (Bioneer, Daejeon, Korea) using primers previously published by Matilainen and collaborators [9]. The relative accumulation of unprocessed mitochondrial transcripts was calculated using the 2(-ΔΔCt) method [25].

2.8. Interferon Signature Analysis

The relative expression of six interferon-stimulated genes was analyzed by qPCR using cDNA from 40 ng RNA, TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific), and Taq Man probes for *IFI27* (Hs01086370_m1), *IFI44L* (Hs00199115_m1), *IFIT1* (Hs00356631_g1), *ISG15* (Hs00192713_m1),

RSAD2 (Hs01057264_m1), and *SIGLEC1* (Hs00988063_m1). The relative abundance was normalized to *HPRT1* (Hs03929096_g1) and *18S* (Hs999999001_s1). Median expression was used to calculate an interferon score as described by Dhir, Rice, and collaborators [6,26].

3. Results

A total of 24 patients with bi-allelic variants in *PNPT1* were identified in 15 families, including seven new patients (P1, 2, 3, 3.1, 4, 7, 8), three previously published individuals with additional clinical information since the previous publication (P5, 6, 9), and fourteen other patients reported in the literature (Tables S1 and S2). A total of 22 different pathogenic variants were found, of which only three were reported in more than one family. The p.(Ala507Ser) variant was reported in 27% of the families (4/15), followed by the p.(Thr531Arg) variant at 13% (2/15), and p.(Arg136His) at 13% (2/15) (Figure 1).

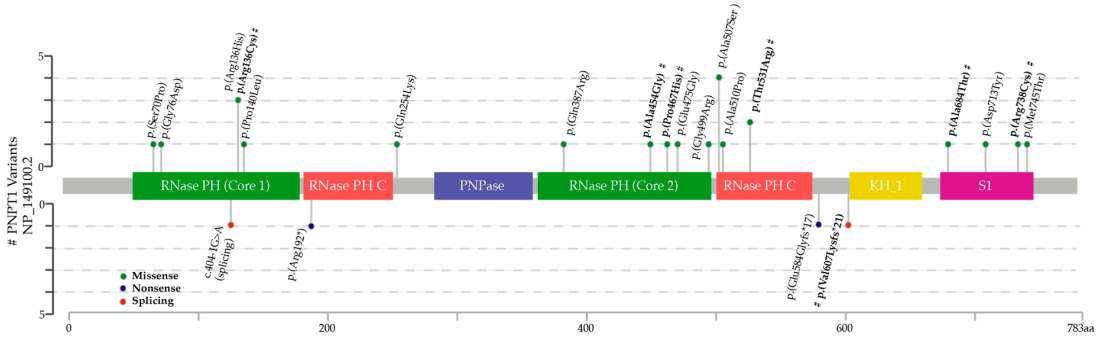

Figure 1. Lollipop plot depicting the polynucleotide phosphorylase (PNPase) protein and location of pathogenic variants reported in 30 alleles (15 families). Missense variants are shown above the protein. Nonsense and splicing variants are shown below. Novel variants reported in this article are marked with #.

Most of the patients (75% $n = 18$) presented within the first year of life. The most common initial manifestations were tone abnormalities (42% $n = 10$), feeding difficulties (17% $n = 4$), and mild to severe sensorineural hearing loss (38% $n = 9$). Other reported initial features were regression, choreoathetosis, visual loss, and cataracts (Table S1).

The median age at the last follow up was 8.9 years (range 1–78). Three of the patients died due to acute encephalopathy (P15 at 4 years), infection (P16 at 2.4 years), and unstated reasons (P21 at 2 years) (Table S1).

The clinical symptoms of all the reported patients to date are summarized in Figure 2 and Table S1.

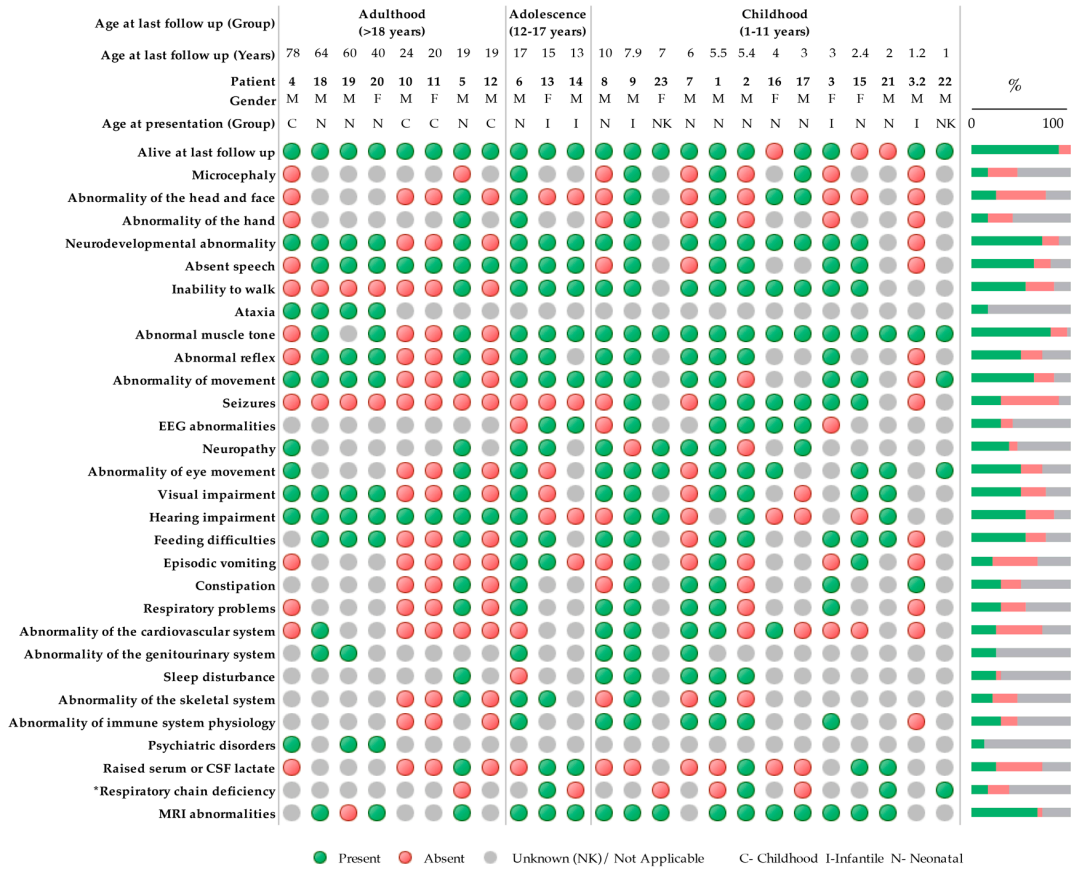

Figure 2. Clinical symptoms using Human Phenotype Ontology (HPO) terminology of patients with bi-allelic *PNPT1* variants (n = 24). Green indicates a present phenotype, red indicates an absent phenotype, and gray is unknown or not applicable. * Respiratory chain deficiencies refer to reported reductions in oxidative phosphorylation (OXPHOS) enzyme activity measured by spectrophotometric analysis in any tissue. Neonatal (0–28 days), infant (28 days–1 year), and childhood (>1 year).

3.1. cDNA Studies Identify a Splicing Defect in P2

The NM_033109.4:c.1818T>G heterozygous variant identified in P2 was originally believed to be synonymous p.(Val606=), however it was located 5 bp from the donor splice site and in silico studies suggested the possibility of a splicing defect (Table S2). cDNA studies performed from cultured fibroblasts grown with and without cycloheximide treatment to inhibit nonsense-mediated decay (NMD) confirmed a splicing defect resulting in a frameshift deletion and premature termination codon p.(Val607Lysfs*21), leading the transcript to be subject to degradation by NMD (Figure 3).

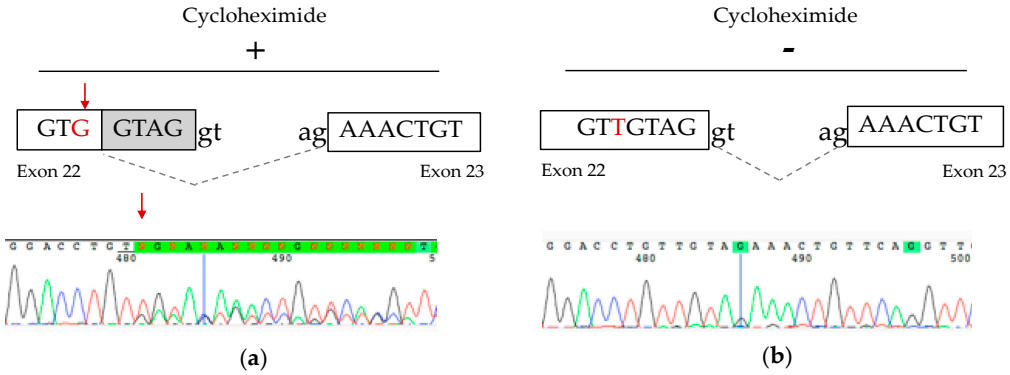

Figure 3. cDNA studies from P2 fibroblast cells. (**a**) Schematic representation and partial electropherogram showing the creation of a new splice site generated by the NM_033109.4:c.1818T>G variant in cells treated with cycloheximide. (**b**) Without cycloheximide treatment, the heterozygous NM_033109.4:c.1818T>G variant was not detectable, suggesting that the mutant RNA is subject to nonsense-mediated decay.

3.2. Mitochondrial OXPHOS Protein Expression and Activity

In the fibroblasts, there was a reduction of the mitochondrial complex IV COX II subunit and the complex I NDUFB8 subunit in patients P1–4, although the levels of other OXPHOS subunits were normal (Figure 4a). The activities of complex I and complex IV were determined using a dipstick assay. Compared to the control means, there was a relatively mild reduction in complex I by 20% (P1), 45% (P2), 50% (P4), and a modest reduction in complex IV by 40% (P1), 57% (P2), 20% (P3), and 80% (P4) (Figure 4b). P5 was a previously reported patient with *PNPT1* variants [10] and was used as a positive control.

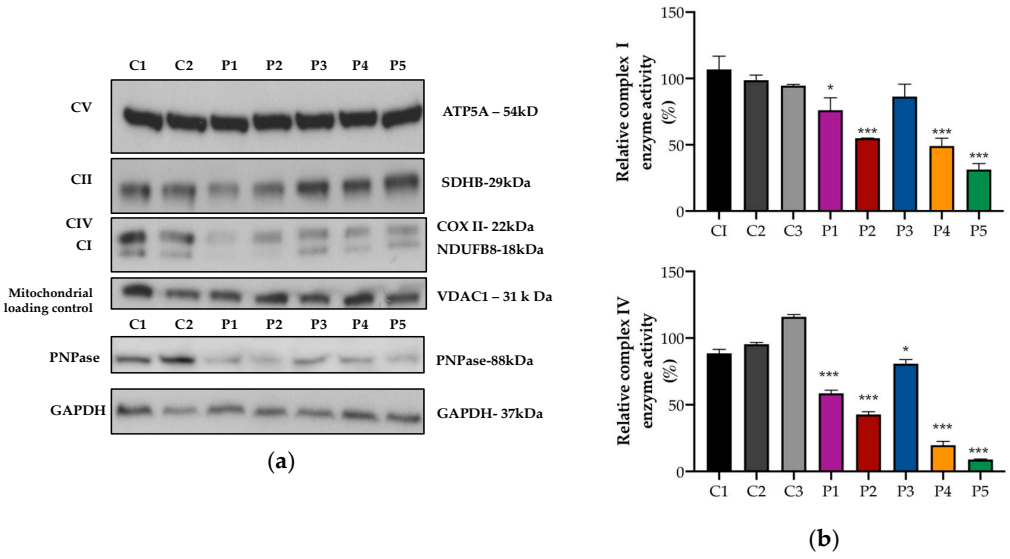

Figure 4. Protein expression and OXPHOS activity in fibroblasts. (**a**) Representative images of Western blot analysis performed in fibroblast proteins from controls (C) and patients (P). In the patients, there was a reduction in protein expression of PNPase (lower panel) and complex I and complex IV OXPHOS mitochondrial subunits (upper panel). (**b**) Complex I (upper panel) and complex IV (lower panel) enzyme activity was analyzed using dipstick activity assays. In the patients, there was a mild/moderate reduction in complex I and complex IV enzyme activities. The mean and variation (SEM) between three independent experiments are shown. (*** $p < 0.001$, * $p < 0.033$).

However, spectrophotometric analysis of OXPHOS enzymes was normal in fibroblasts from P1, and was only borderline low for complex IV activity relative to protein and citrate synthase in muscle from P2 (Table 1).

Table 1. Spectrophotometric analysis of oxidative phosphorylation (OXPHOS) enzyme activities in fibroblasts from P1 and muscle from P2.

	P1 (Fibroblasts)			P2 (Muscle)		
Enzyme	**Residual Activity (%)**	**CS Ratio (%)**	**CII Ratio (%)**	**Residual Activity (%)**	**CS Ratio (%)**	**CII Ratio (%)**
I	150	107	75	75	73	126
II	194	140	-	59	58	-
III	266	187	133	69	65	113
IV	91	67	47	34	34	59
CS	137			102		

Activities of OXPHOS enzyme complexes I, II, III, IV, and citrate synthase (CS) are expressed as % relative to protein (residual activity), citrate synthase (CS ratio), and CII (CII ratio) of control samples.

3.3. Accumulation of Mitochondrial Unprocessed Transcripts

To assess the PNPase mitochondrial RNA processing activity, we measured the accumulation of unprocessed polycistronic mitochondrial transcripts in fibroblasts.

Compared to controls, fibroblasts from patients P1–4 showed an accumulation of aberrantly processed mitochondrial RNA transcripts (Figure 5) in line with the PNPase function in mtRNA processing. The accumulation of unprocessed transcripts in myoblasts from a patient with bi-allelic pathogenic variants in *PNPT1* was previously described [9]. Notably, our patients presented different patterns of accumulation, which were not restricted to transcripts around *MT-ND6*. Further studies may help us to understand if the differences in the accumulation of unprocessed mitochondrial transcripts are related to specific *PNPT1* variants.

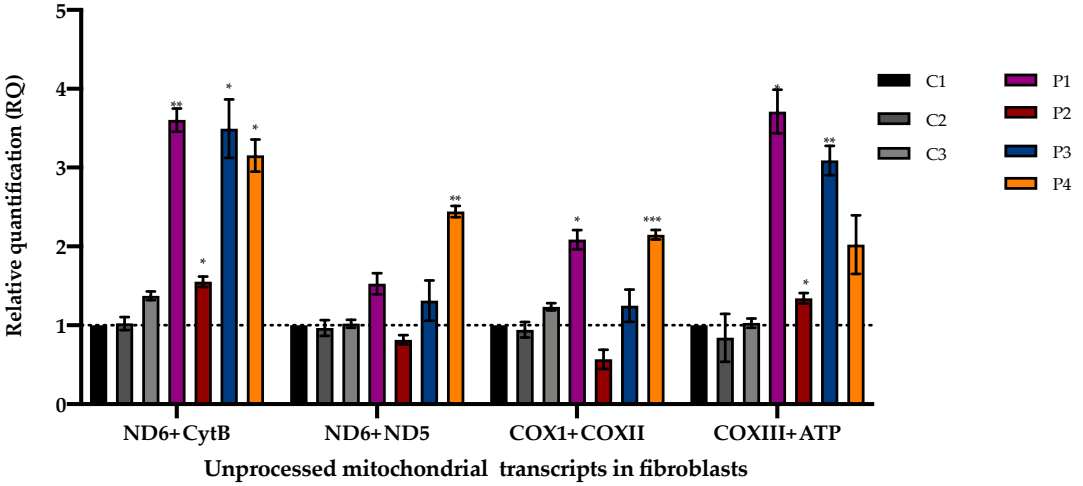

Figure 5. Unprocessed polycistronic mitochondrial transcripts in fibroblasts measured by qPCR. Mean and variation (SEM) between three independent experiments are shown (*** $p < 0.001$, ** $p < 0.002$, * $p < 0.033$).

3.4. Interferon Signature in Patients with PNPT1 Variants

Increased expression of interferon-stimulated genes has been previously described in blood from P14, P22, and P23 with bi-allelic *PNPT1* variants [6]. To assess if the PNPase defects could elicit an

increased interferon response in different tissues, we calculated the interferon score in fibroblasts (P1–4) and in blood (only available from P2 and P4). In fibroblasts from P1–4 stimulated with IFN-α2a (100 U/mL), the interferon response was not different to controls (interferon score below the 1.3 threshold in fibroblasts). On the other hand, in blood from P1 and P4, a positive interferon score was identified with a median of 2.53 and 4.65, respectively (above the 1.8 threshold in blood) (Figure 6).

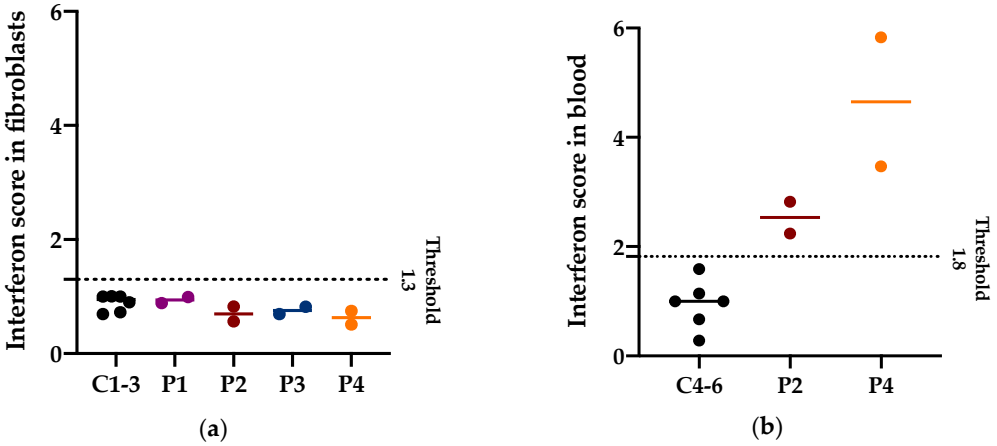

Figure 6. Interferon (IFN) score in fibroblasts and in blood. (**a**) Median and individual values of the IFN score in fibroblasts of three controls and patients with *PNPT1* variants (P1–4) from two independent experiments. The IFN score threshold of 1.3 corresponds to the mean IFN score of controls + 2SD. Higher values are considered positive. (**b**) Median and individual values of the IFN score in blood of three controls and patients with *PNPT1* variants (P2 and P4) from two independent experiments. The IFN score threshold of 1.8 corresponds to the mean IFN score of controls + 2SD. Higher values are considered positive.

4. Discussion

This cohort illustrated the marked phenotypic heterogeneity associated with bi-allelic *PNPT1* variants. Severe developmental delay and regression were the most common neurodevelopmental abnormalities. Sensorineural hearing loss, optic atrophy, tone, and movement abnormalities—including choreoathetosis, dystonia, myoclonus, and ataxia—were also amongst the most prevalent clinical features. Most patients had a history of hypotonia (n = 19) and four patients presented a pattern of central hypotonia with peripheral hypertonia. Brain imaging was available in 17 patients, of which 16 patients had MRI abnormalities. The most common were white matter abnormalities (n = 8), followed by thin corpus callosum (n = 5), and basal ganglia abnormalities (n = 4) (Figure S1).

Interestingly, most of the patients in the cohort had normal to mildly deficient OXPHOS when measured by spectrophotometric enzyme assays. This finding has several implications. First, a normal OXPHOS enzyme analysis does not rule out the diagnosis of *PNPT1*-related disease. Second, reduced OXPHOS protein expression or activity measured by dipstick assays does not always correlate with spectrophotometric assays, as shown in P1 and P2. This discrepancy could be attributed to sample preparation and differences in the methodology, which has been discussed previously [10]. Finally, in order to confirm pathogenicity, functional analyses focusing on the primary role of *PNPT1* (mtRNA processing) and interferon signaling in blood rather than fibroblasts appear to be more sensitive, and potentially more specific than evaluating the downstream effects on the OXPHOS enzyme activities.

Most of the patients with *PNPT1*-related disease have been identified through NGS approaches. The identification of an apparently synonymous variant in P2 that was later demonstrated to result in abnormal mRNA splicing highlights the importance of careful interpretation of apparently "silent" variants in NGS data, the utility of in silico tools, and the need for cDNA studies when following up potential splicing variants.

The identification of an interferon signature in blood from patients with *PNPT1* variants raises the possibility that an immune response could contribute to disease pathogenesis [6]. Further longitudinal studies could elucidate if there is a correlation between the interferon signature and disease severity, or the possibility of being a useful clinical biomarker and/or therapeutic target.

So far, little is known about possible genotype–phenotype correlations, and it is not clear what factors could be contributing to the clinical spectrum of *PNPT1*-related diseases. Studies suggest that the rate of disease progression could be, in part, related to having two deleterious variants affecting the first core domain [9]. In line with this, three patients with rapidly progressive disease and early death (median age 2.4 years) were described (Table S1), in which the variants were located in the first core domain. P21 had the p.(Arg136His) variant in a homozygous state. P15 also had the p.(Arg136His) variant but in compound heterozygosity with the p.(Pro140Leu) variant, which was also located in the first core domain. P16, who had a p.(Gly76Asp) located in the first core domain and a p.(Arg192*), was predicted to result in nonsense-mediated decay [6,9,12]. The novel variant p.(Arg136Cys) identified in P3 was an alternate amino acid change located in the same active site as described in P21 and P15, but her second pathogenic variant p.(Pro467His) was located in the second core domain. Interestingly, she is currently three years old and has a relatively slower progression. Her brother (P3.2) also seemed to be more mildly affected. Intrafamilial variability was also described in other families with *PNPT1* variants (P5, 6, 7, 8; Table S1), but bona fide genetic modifiers of disease severity have yet to be identified.

P4 had the longest survival reported to date (78 years), with a slowly progressive clinical course. He presented at 11 years old with hearing and visual loss. Subsequently, he developed blepharospasm and ptosis. Over the decades, he developed progressive ataxia, poor night vision, and external ophthalmoplegia. His hearing worsened during his 60s and, therefore, he received a cochlear implant. Eaton and collaborators [13] also described one family that presented with a slowly progressive pattern. Both affected individuals had congenital sensorineural hearing loss and developed neurological symptoms in adulthood. It is interesting to note that the pathogenic *PNPT1* variants in both these patients were distal variants located within the S1 domain. Further work and more patients are required to formally establish genotype–phenotype correlations.

Careful phenotypic analysis in rare monogenic diseases is valuable for identifying common clinical features [27,28]. In our cohort, this approach led us to describe the most common phenotypic findings associated with *PNPT1*-related disease that should increase the diagnostic suspicion of this genetic etiology, and could guide screening and management in patients with a molecular diagnosis. On the other hand, no single finding was present in 100% of the patients and, thus, the lack of any of these features does not rule out the diagnosis. Some of the clinical features have only been described in adult patients, such as obsessive–compulsive disorder (P19), depression (P4, P19), and paranoia (P20), or in single cases, such as renal artery stenosis in P9. The causal relationship of some of these findings is unknown and it remains to be seen whether some of these features are coincidental rather than linked to the pathophysiology of *PNPT1* dysfunction.

The diverse clinical heterogeneity of patients with bi-allelic *PNPT1* variants presents a diagnostic and management challenge to the clinician. As more patients are diagnosed, it is important to share detailed clinical information to understand the natural history, elucidate new insights into the mechanisms underlying the pathology, and aid with the identification of potential treatment strategies.

Supplementary Materials: The following are available online at http://www.mdpi.com/2077-0383/8/11/2020/s1, Table S1: Clinical phenotype of patients with bi-allelic PNPT1 variants; Table S2: In silico studies and population frequency of PNPT1 variants identified in new families; Figure S1: Brain MRI findings in patients with bi-allelic *PNPT1* variants.

Author Contributions: Conceptualization, J.C.; data curation, R.R., M.F.-F., M.K.K., and S.S.; formal analysis, R.R. and L.G.R.; funding acquisition, D.R.T., and J.C.; clinical evaluation and investigation, M.P.K., S.B., D.J.A., G.M.E., M.C.F., M.K.K., M.J.W., and T.Y.T.; methodology, R.R. and N.J.V.B.; software, M.J.C.; supervision, N.J.V.B., A.G.C., D.R.T., and J.C.; writing—Original draft, R.R.; writing—Review and editing, N.J.V.B., A.G.C., L.G.R., M.P.K., S.B., D.J.A., M.F.-F., M.J.C., M.C.F., M.K.K., G.M.E., S.S., M.J.W., T.Y.T., D.R.T., and J.C.

Funding: This research was supported by: A New South Wales Office of Health and Medical Research (OHMR) Council Sydney Genomics Collaborative grant (JC); the Australian Genomics Health Alliance (Australian Genomics) project, funded by an NHMRC Targeted Call for Research grant (GNT1113531; JC, DRT); NHMRC research fellowship 1155244 (DRT); OHMR Fellowship (MJC); and a CONACYT Postgraduate Research Scholarship (RR). Sequencing and analysis were provided by the Broad Institute of MIT and Harvard Center for Mendelian Genomics (Broad CMG) and were funded by the National Human Genome Research Institute, the National Eye Institute, the National Heart, Lung and Blood Institute grant UM1 HG008900, and, in part, by a National Human Genome Research Institute grant R01 HG009141.

Acknowledgments: We are grateful to the Crane and Perkins families for their generous financial support. The research conducted at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, H.W.; Rainey, R.N.; Balatoni, C.E.; Dawson, D.W.; Troke, J.J.; Wasiak, S.; Hong, J.S.; McBride, H.M.; Koehler, C.M.; Teitell, M.A.; et al. Mammalian polynucleotide phosphorylase is an intermembrane space RNase that maintains mitochondrial homeostasis. *Mol. Cell. Biol.* **2006**, *26*, 8475–8487. [CrossRef]
2. Borowski, L.S.; Dziembowski, A.; Hejnowicz, M.S.; Stepien, P.P.; Szczesny, R.J. Human mitochondrial RNA decay mediated by PNPase-hSuv3 complex takes place in distinct foci. *Nucleic Acids Res.* **2013**, *41*, 1223–1240. [CrossRef] [PubMed]
3. Wang, G.; Chen, H.W.; Oktay, Y.; Zhang, J.; Allen, E.L.; Smith, G.M.; Fan, K.C.; Hong, J.S.; French, S.W.; McCaffery, J.M.; et al. PNPASE regulates RNA import into mitochondria. *Cell* **2010**, *142*, 456–467. [CrossRef] [PubMed]
4. Golzarroshan, B.; Lin, C.L.; Li, C.L.; Yang, W.Z.; Chu, L.Y.; Agrawal, S.; Yuan, H.S. Crystal structure of dimeric human PNPase reveals why disease-linked mutants suffer from low RNA import and degradation activities. *Nucleic Acids Res.* **2018**, *46*, 8630–8640. [CrossRef] [PubMed]
5. Gammage, P.A.; Moraes, C.T.; Minczuk, M. Mitochondrial Genome Engineering: The Revolution May Not Be CRISPR-Ized. *Trends Genet.* **2018**, *34*, 101–110. [CrossRef]
6. Dhir, A.; Dhir, S.; Borowski, L.S.; Jimenez, L.; Teitell, M.; Rotig, A.; Crow, Y.J.; Rice, G.I.; Duffy, D.; Tamby, C.; et al. Mitochondrial double-stranded RNA triggers antiviral signalling in humans. *Nature* **2018**, *560*, 238–242. [CrossRef]
7. Pajak, A.; Laine, I.; Clemente, P.; El-Fissi, N.; Schober, F.A.; Maffezzini, C.; Calvo-Garrido, J.; Wibom, R.; Filograna, R.; Dhir, A.; et al. Defects of mitochondrial RNA turnover lead to the accumulation of double-stranded RNA in vivo. *PLoS Genet.* **2019**, *15*, e1008240. [CrossRef]
8. Von Ameln, S.; Wang, G.; Boulouiz, R.; Rutherford, M.A.; Smith, G.M.; Li, Y.; Pogoda, H.M.; Nürnberg, G.; Stiller, B.; Volk, A.E.; et al. A mutation in PNPT1, encoding mitochondrial-RNA-import protein PNPase, causes hereditary hearing loss. *Am. J. Hum. Genet.* **2012**, *91*, 919–927. [CrossRef]
9. Matilainen, S.; Carroll, C.J.; Richter, U.; Euro, L.; Pohjanpelto, M.; Paetau, A.; Isohanni, P.; Suomalainen, A. Defective mitochondrial RNA processing due to PNPT1 variants causes Leigh syndrome. *Hum. Mol. Genet.* **2017**, *26*, 3352–3361. [CrossRef]
10. Alodaib, A.; Sobreira, N.; Gold, W.A.; Riley, L.G.; Van Bergen, N.J.; Wilson, M.J.; Bennetts, B.; Thorburn, D.R.; Boehm, C.; Christodoulou, J. Whole-exome sequencing identifies novel variants in PNPT1 causing oxidative phosphorylation defects and severe multisystem disease. *Eur. J. Hum. Genet.* **2016**, *25*, 79. [CrossRef]
11. Vedrenne, V.; Gowher, A.; De Lonlay, P.; Nitschke, P.; Serre, V.; Boddaert, N.; Altuzarra, C.; Mager-Heckel, A.M.; Chretien, F.; Entelis, N.; et al. Mutation in PNPT1, which encodes a polyribonucleotide nucleotidyltransferase, impairs RNA import into mitochondria and causes respiratory-chain deficiency. *Am. J. Hum. Genet.* **2012**, *91*, 912–918. [CrossRef] [PubMed]
12. Sato, R.; Arai-Ichinoi, N.; Kikuchi, A.; Matsuhashi, T.; Numata-Uematsu, Y.; Uematsu, M.; Fujii, Y.; Murayama, K.; Ohtake, A.; Abe, T.; et al. Novel biallelic mutations in the PNPT1 gene encoding a mitochondrial-RNA-import protein PNPase cause delayed myelination. *Clin. Genet.* **2018**, *93*, 242–247. [CrossRef] [PubMed]
13. Eaton, A.; Bernier, F.P.; Goedhart, C.; Caluseriu, O.; Lamont, R.E.; Boycott, K.M.; Parboosingh, J.S.; Innes, A.M. Care4Rare Canada Consortium. Is PNPT1-related hearing loss ever non-syndromic? Whole exome sequencing of adult siblings expands the natural history of PNPT1-related disorders. *Am. J. Med. Genet. A* **2018**, *176*, 2487–2493. [CrossRef] [PubMed]

14. Riley, L.G.; Cowley, M.J.; Gayevskiy, V.; Roscioli, T.; Thorburn, D.R.; Prelog, K.; Bahlo, M.; Sue, C.M.; Balasubramaniam, S.; Christodoulou, J. A SLC39A8 variant causes manganese deficiency, and glycosylation and mitochondrial disorders. *J. Inherit. Metab. Dis.* **2017**, *40*, 261–269. [CrossRef]
15. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249. [CrossRef]
16. Choi, Y.; Sims, G.E.; Murphy, S.; Miller, J.R.; Chan, A.P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **2012**, *7*, e46688. [CrossRef]
17. Kircher, M.; Witten, D.M.; Jain, P.; O'Roak, B.J.; Cooper, G.M.; Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **2014**, *46*, 310–315. [CrossRef]
18. Schwarz, J.M.; Rodelsperger, C.; Schuelke, M.; Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **2010**, *7*, 575–576. [CrossRef]
19. Desmet, F.O.; Hamroun, D.; Lalande, M.; Collod-Beroud, G.; Claustres, M.; Beroud, C. Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **2009**, *37*, e67. [CrossRef]
20. Punta, M.; Coggill, P.C.; Eberhardt, R.Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2012**, *40*, D290–D301. [CrossRef]
21. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404. [CrossRef] [PubMed]
22. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**, *536*, 285–291. [CrossRef] [PubMed]
23. Frazier, A.E.; Thorburn, D.R. Biochemical analyses of the electron transport chain complexes by spectrophotometry. *Methods Mol. Biol.* **2012**, *837*, 49–62. [PubMed]
24. Calvo, S.E.; Tucker, E.J.; Compton, A.G.; Kirby, D.M.; Crawford, G.; Burtt, N.P.; Rivas, M.; Guiducci, C.; Bruno, D.L.; Goldberger, O.A.; et al. High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat. Genet.* **2010**, *42*, 851–858. [CrossRef]
25. Schmittgen, T.D.; Livak, K.J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **2008**, *3*, 1101–1108. [CrossRef]
26. Rice, G.I.; Forte, G.M.A.; Szynkiewicz, M.; Chase, D.S.; Aeby, A.; Abdel-Hamid, M.S.; Ackroyd, S.; Allcock, R.; Bailey, K.M.; Balottin, U.; et al. Assessment of interferon-related biomarkers in Aicardi-Goutières syndrome associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, and ADAR: A case-control study. *Lancet Neurol.* **2013**, *12*, 1159–1169. [CrossRef]
27. Fuchs, S.A.; Schene, I.F.; Kok, G.; Jansen, J.M.; Nikkels, P.G.J.; van Gassen, K.L.I.; Terheggen-Lagro, S.W.J.; van der Crabben, S.N.; Hoeks, S.E.; Niers, L.E.M.; et al. Aminoacyl-tRNA synthetase deficiencies in search of common themes. *Genet. Med.* **2019**, *21*, 319–330. [CrossRef]
28. Robinson, P.N. Deep phenotyping for precision medicine. *Hum. Mutat.* **2012**, *33*, 777–780. [CrossRef]

The Supplementary data related to this article is included in Appendix C

Chapter 6 Biparental inheritance of mitochondrial DNA in humans is not a common phenomenon

6.1 Introduction

The analyses in Chapter 3-5 have highlighted the diagnostic utility of sequencing family trios by identifying variants in known and novel disease genes. This chapter contains a publication that illustrates an additional bonus of trio GS by evaluating the utility of interrogating the data to investigate mtDNA inheritance (Rius, Cowley et al. 2019).

In response to a controversial article from Luo and colleagues (Luo, Valencia et al. 2018), which raised the possibility of biparental mitochondrial DNA transmission in several three-generation families, we used data from the trio GS cohort to interrogate mtDNA inheritance and examine the frequency of this phenomenon. Our results were consistent with strictly maternal mtDNA inheritance. This was the first cohort of paediatric patients with suspected mitochondrial disease where mtDNA inheritance was systematically studied by trio GS.

Since the manuscript was published, other groups have used trio GS data to look for evidence of mtDNA transmission. Using data from the Genomics England 100,000 Genomes Rare Disease project, Wei and colleagues analysed 11,035 trio GS and found 7 families with apparently similar heteroplasmy patterns with the families described by Luo and colleagues that suggested biparental mtDNA inheritance. However, by analysing nuclear genome sequences the group was able to explain that in these families, the pattern was likely due to paternally transmitted nuclear-mitochondrial DNA segments (NUMTs) and not due to paternal mtDNA transmission (Wei, Pagnamenta et al. 2020).

BRIEF COMMUNICATION | Genetics in Medicine

Biparental inheritance of mitochondrial DNA in humans is not a common phenomenon

Rocio Rius, MD MPH[1,2], Mark J. Cowley, PhD[3,4,5], Lisa Riley, PhD[6,7], Clare Puttick, BSc MStats[3,8], David R. Thorburn, PhD FSc(RCPA)[1,2,9] and John Christodoulou, PhD FRACP[1,2,6,9]

Purpose: A recent report has raised the possibility of biparental mitochondrial DNA (mtDNA) inheritance, which could lead to concerns by health-care professionals and patients regarding investigations and genetic counseling of families with pathogenic mitochondrial DNA variants. Our aim was to examine the frequency of this phenomenon by investigating a cohort of patients with suspected mitochondrial disease.

Methods: We studied genome sequencing (GS) data of DNA extracted from blood samples of 41 pediatric patients with suspected mitochondrial disease and their parents.

Results: All of the mtDNA variants in the probands segregated with their mother or were apparently de novo. There were no variants that segregated only with the father and none of these families showed evidence of biparental inheritance of their mtDNA.

Conclusion: Paternal mitochondrial transmission is unlikely to be a common occurrence and therefore at this point we would not recommend changes in clinical practice.

Genetics in Medicine (2019) https://doi.org/10.1038/s41436-019-0568-0

Keywords: mitochondrial genome; genetic testing; genetic counseling; mitochondrial diseases; paternal inheritance

INTRODUCTION

The underlying molecular diagnosis of patients with mitochondrial diseases can be due to nuclear or mitochondrial DNA (mtDNA) pathogenic variants. In humans, mtDNA is generally inherited from the mother. Thus, the cascade diagnostic investigations, genetic counseling, and reproductive choices for patients with pathogenic mtDNA variants are currently based on that premise.[1]

In a recent study, Luo et al.[2] reported evidence of paternal transmission of mtDNA to offspring in three families. Some of the issues emerging from these findings relate specifically to mitochondrial disease diagnosis and genetic counseling and could prompt concerns about current practice.

The possibility of paternal inheritance of mtDNA has been reported previously in mice, including across multiple generations.[3,4] Tissue-specific inheritance of paternal mtDNA in humans has also been reported in a single individual with mitochondrial myopathy whose skeletal muscle had a de novo homoplasmic 2-bp deletion in the *MT-ND2* gene on the paternal mtDNA haplotype, while only maternal mtDNA was detected in other tissues tested.[5] The recent finding by Luo et al.[2] prompts reconsideration of whether paternal inheritance may be more common in humans than previously considered. To our knowledge, there have not been any other systematic investigations in human trios.

MATERIALS AND METHODS

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all patients included in the study. This project was approved by the Human Research Ethics Committee of the Sydney Children's Hospitals Network (ID number 10/CHW/114).

We studied genome sequencing (GS) data of DNA extracted from blood samples of 41 pediatric patients recruited in Australia with suspected mitochondrial disease and their parents, as previously reported.[6] Thirty-seven of the patients had trio GS, while in four patients DNA from the father was not available.

[1]Murdoch Children's Research Institute, Melbourne, Australia; [2]Department of Paediatrics, University of Melbourne, Melbourne, Australia; [3]Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia; [4]St Vincent's Clinical School, UNSW Sydney, Sydney, Australia; [5]Children's Cancer Institute, Kensington, Australia; [6]Kids Research, The Children's Hospital at Westmead, Sydney, NSW, Australia; [7]Discipline of Child & Adolescent Health, Sydney Medical School, University of Sydney, Sydney, Australia; [8]The Francis Crick Institute, London, UK; [9]Victorian Clinical Genetics Services, Royal Children's Hospital, Melbourne, VIC, Australia. Correspondence: John Christodoulou (john.christodoulou@mcri.edu.au)

Submitted 2 March 2019; accepted: 24 May 2019
Published online: 07 June 2019

BRIEF COMMUNICATION

Table 1 Mitochondrial DNA (mtDNA) variant segregation

				Proband			Mother			Father		
Patient	All inherited variants shared with the mother	Paternal specific variants	Number of de novo variants	Number of variants	Heteroplasmic variants	Haplotype[14]	Number of variants	Heteroplasmic variants	Haplotype[14]	Number of variants	Heteroplasmic variants	Haplotype[14]
1	Yes	No	0	31	0	M35b	31	0	M35b	16	0	HV
2	Yes	No	0	9	0	H1aq	9	0	H1aq	9	0	H1aq
3	Yes	NA	1	38	1 m.7925G>A (70%)	T2a1b1a	37	0	T2a1b1a	NA	NA	NA
4	Yes	No	0	9	0	H1	9	0	H1	32	0	U5b2b4a
5	Yes	No	0	10	0	H1	10	0	H1	33	0	T2b30
6	Yes	No	0	43	0	M1a1i	43	0	M1a1i	13	0	H1m1
7	Yes	No	0	17	0	V10a	17	0	V10a	10	0	H1c
8	Yes	No	0	33	1 m.8489A>C (3%)	K1a2a	33	1 m.8489A>C (28%)	K1a2a	36	0	T2b5a
9	Yes	No	0	29	0	A1a	29	0	A1a	35	0	T2b
10	Yes	No	0	12	0	H1a3a	12	0	H1a3a	35	0	J1c2c1
11	Yes	No	0	0	0	H2a2a1	0	0	H2a2a1	29	0	J1c2
12	Yes	No	0	39	0	J2a1a1	39	0	J2a1a1	12	0	H45
13	Yes	No	0	16	0	H4a1a1a	16	0	H4a1a1a	2	1 m.16183A>C (20%)	H2a2a1
14	Yes	No	0	33	0	J1c2b1	33	0	J1c2b1	35	0	J1c2b1
15	Yes	No	0	11	0	H1	12	1 m.16150C>T (25%)	H1	11	0	H5
16	Yes	No	1	34	1 m.3243A>G (40%)	U5b2b4a	33	0	U5b2b4a	31	0	U3a1a
17	Yes	No	0	36	0	U2d2	36	0	U2d2	32	0	X4
18	Yes	No	0	24	0	HV2a1	24	0	HV2a1	39	0	G2a2
19	Yes	No	0	31	0	J1c2	31	0	J1c2	NA	NA	NA
20	Yes	No	0	31	0	M3c2	31	0	M3c2	31	0	U7a2
21	Yes	No	0	35	0	T2g2	35	0	T2g2	14	0	H1bb
22	Yes	No	1	36	0	A12a	35	0	A12a	15	0	H13a1a1a
23	Yes	No	0	37	0	K1a4f1	37	0	K1a4f1	NA	NA	NA
24	Yes	No	0	31	0	P4b1	31	0	P4b1	42	0	M42a
25	Yes	No	0	28	0	J1c3g	28	0	J1c3g	16	0	V10a
26	Yes	No	0	31	0	J1c1b	31	0	J1c1b	11	0	H1c
27	Yes	No	0	41	0	I1a1	41	0	I1a1	15	0	H1e1a3
28	Yes	No	0	27	0	U5a2 + 16294	27	0	U5a2 + 16294	34	0	I2c
29	Yes	No	0	8	0	H2a3b	8	0	H2a3b	25	0	U5a2a2
30	Yes	No	0	14	1 m.16399A>G (8%)	H1a	14	1 m.16399A>G (83%)	H1as	38	1 m.16188CT>C (20%)	T2
31	Yes	No	0	12	0	V + @72	12	0	V + @72	24	1 m.16188CT>C (70%)	B4c2
32	Yes	No	2	35	2 m.15617G>A (12%) m.153A>G (96%)	T2b4a1	33	0	T2b4a1	39	0	I1a1b
33	Yes	No	0	12	0	H3 + 152	12	0	H3 + 152	11	0	H3z
34	Yes	No	0	29	0	B4b1a2	29	0	B4b1a2	32	0	W4a1
35	Yes	No	0	12	0	H11a	12	0	H11a	9	0	H1k
36	Yes	No	0	35	0	K1a3a1b	35	0	K1a3a1b	6	0	H2a2a2
37	Yes	No	0	13	0	H45b	13	0	H45b	13	0	H3a1a
38	Yes	No	0	40	0	U1a1a	40	0	U1a1a	24	0	U5a1g
39	Yes	No	0	15	0	H11a2a	15	0	H11a2a	34	0	U4c1
40	Yes	No	0	37	0	F1f	37	0	F1f	NA	NA	NA
41	Yes	No	0	41	0	U2e2a1a	41	0	U2e2a1a	14	0	H1c3

Variants are reported relative to the Revised Cambridge Reference Sequence NC_012920.1.

Variants were filtered using Seave[7] and reported relative to the Revised Cambridge Reference Sequence NC_012920.1. Variants with heteroplasmy levels >5% were included in the analysis and if discordant with the mother, lower levels of heteroplasmy were further investigated using the bioinformatic pipeline *mity* (Puttick et al., unpublished). Variants in positions where common sequencing errors occur due to a homopolymer stretch (MT:302–315 and MT:3105–3109) were excluded.

RESULTS

Using GS we observed a median of medians coverage of 4586× (range 1179–22,965×) in the mitochondrial genome. We identified a median of 31 mtDNA variants per proband (Table **1**). In all studied families, homoplasmic variants detected in the mother were also found in the offspring.

Only five probands had any heteroplasmic variants. In two families, heteroplasmic variants detected in the mother presented lower heteroplasmy levels in the probands (families 8 and 30); in one family (family 15) the m.16150C>T heteroplasmic variant (mutant load 25%) detected in the mother was not present in the proband. These different heteroplasmy levels can be explained by the mitochondrial bottleneck during oogenesis.

In one family (family 32), two de novo heteroplasmic variants (m.15617G>A; mutant load 12% and m.153A>G; mutant load 96%) were detected in the proband that were not detected in the mother or father. In family 3, a de novo heteroplasmic variant (m.7925G>A; mutant load 70%) was present in the proband that was not identified in the mother's blood; no DNA was available from the father, however, all 37 variants present in the mother were also identified in the proband and they both shared the same haplotype (T2a1b1a).

We did not find any homoplasmic or heteroplasmic variants in the cohort that segregated only with the father. In addition, none of these families presented with high numbers of heteroplasmic variants or other suggestive patterns of biparental inheritance of their mtDNA. However, because only blood samples were analyzed, tissue-specific heteroplasmy cannot be ruled out.

DISCUSSION

Previous studies reporting paternal mtDNA transmission have been controversial, with most cases later regarded as sample cross-contamination.[8] The recent study by Luo et al.[2] has reactivated the concern about the possibility of paternal mtDNA transmission; they described three multigeneration families with a large number of mtDNA variants with high levels of heteroplasmy, corresponding to mixed haplogroups in a pattern apparently corresponding to autosomal dominant inheritance. They attempted to exclude sample contamination by conducting the mtDNA sequencing in different laboratories, however, their results have also been challenged. Lutz-Bonengel[9] and Vissing[10] speculated that the apparent biparental mtDNA inheritance pattern in the families could be attributed to amplification of nuclear mtDNA segments (NuMTs).[9,10] Luo et al.[11] contested that claim by arguing that it is unlikely that the NuMT would amplify with the different sequencing methods used. In addition, they argued that there would be a low probability of all offspring studied inheriting the NuMT variants. Nonetheless, they plan to conduct further single-cell sequencing studies with additional families to rule out that possibility. Even if paternal mtDNA transmission is confirmed, the mechanism of this event is still not clear, as well as the potential consequences of this phenomenon, given that no actual causal relationship with mitochondrial disease was established in the reported families.

Our results further support the prevailing view that paternal inheritance of mtDNA appears to be a rare event in humans, and suggest that the scenario described by Luo et al.[2] is uncommon. We suggest that genetic counseling and genetic investigations for families with suspected mitochondrial disease do not need to change. Nonetheless, it may be worth considering sequencing paternal mtDNA where a patient has an apparent de novo mtDNA variant.

We believe that this is the largest cohort of patients with suspected mitochondrial disease undergoing trio GS reported to date. As access to GS technology and genomic data sharing become more widespread in clinical practice,[12] the analysis of nuclear and mtDNA sequencing data in larger cohorts could aid in better understanding the frequency and consequences of biparental mtDNA inheritance events, and could lead to the identification of variants in nuclear genes involved in paternal mtDNA elimination that may be contributing to this mechanism.[2,13]

ACKNOWLEDGEMENTS

This research was supported by a New South Wales (NSW) Office of Health and Medical Research Council Sydney Genomics Collaborative grant (J.C.), National Health and Medical Research Council (NHMRC) project grant 1026891 (J.C.), NHMRC research fellowship 11022896 (D.R.T.), and a NSW Health Early-Mid Career Fellowship (M.J.C.). We are grateful to the Crane and Perkins families for their generous financial support. The research conducted at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program.

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

1. Gorman GS, Chinnery PF, DiMauro S, et al. Mitochondrial diseases. Nat Rev Dis Primers. 2016;2:16080.
2. Luo S, Valencia CA, Zhang J, et al. Biparental inheritance of mitochondrial DNA in humans. Proc Natl Acad Sci USA. 2018;115:13039–13044.
3. Gyllensten U, Wharton D, Josefsson A, Wilson AC. Paternal inheritance of mitochondrial DNA in mice. Nature. 1991;352:255–257.

4. Kidgotko OV, Kustova MY, Sokolova VA, Bass MG, Vasilyev VB. Transmission of human mitochondrial DNA along the paternal lineage in transmitochondrial mice. Mitochondrion. 2013;13:330–336.
5. Schwartz M, Vissing J. Paternal inheritance of mitochondrial DNA. N Engl J Med. 2002;347:576–580.
6. Heimer G, Keratar JM, Riley LG, et al. MECR mutations cause childhood-onset dystonia and optic atrophy, a mitochondrial fatty acid synthesis disorder. Am J Hum Genet. 2016;99:1229–1244.
7. Gayevskiy V, Roscioli T, Dinger ME, Cowley MJ, Wren J. Seave: a comprehensive web platform for storing and interrogating human genomic variation. Bioinformatics. 2018;35:122–125.
8. Bandelt HJ, Parson W. Consistent treatment of length variants in the human mtDNA control region: a reappraisal. Int J Legal Med. 2008;122: 11–21.
9. Lutz-Bonengel S, Parson W. No further evidence for paternal leakage of mitochondrial DNA in humans yet. Proc Natl Acad Sci USA. 2019;116:1821–1822.
10. Vissing J. Paternal comeback in mitochondrial DNA inheritance. Proc Natl Acad Sci USA. 2019;116:1475–1476.
11. Luo S, Valencia CA, Zhang J, et al. Reply to Lutz-Bonengel et al.: biparental mtDNA transmission is unlikely to be the result of nuclear mitochondrial DNA segments. Proc Natl Acad Sci USA. 2019;116:1823–1824.
12. Stark Z, Dolman L, Manolio TA, et al. Integrating genomics into healthcare: a global responsibility. Am J Hum Genet. 2019;104:13–20.
13. McWilliams TG, Suomalainen A. Mitochondrial DNA can be inherited from fathers, not just mothers. Nature. 2019;565:296–297.
14. Weissensteiner H, Pacher D, Kloss-Brandstätter A, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 2016;44:W58–W63.

Chapter 7 Singleton GS for diagnosis of suspected mitochondrial disease

7.1 Introduction

The Australian Genomics Health Alliance is a project aiming to evaluate the incorporation of genomics to the clinic (Australian Genomics Health Alliance 2020). Several clinical flagship projects have been designed to study the application of genomics in specific diseases, including mitochondrial diseases through the mitochondrial disease flagship. This flagship brings together national expertise to examine the clinical utility of genomic sequencing to diagnose mitochondrial diseases (Stark, Boughtwood et al. 2019). Prospectively recruited patients with a probable or definite MD diagnosis, based on the modified Nijmegen criteria described in Table 3, were randomly assigned to genome sequencing or exome sequencing with mitochondrial DNA sequencing using DNA extracted from blood. Cases in which the modified Nijmegen score was below 5 but where there was a strong clinical suspicion of mitochondrial disease were reviewed by an expert committee in order to qualify for inclusion.

Singleton genome sequencing was conducted by the Garvan Institute as described in Section 2.3 and variant curation was performed as part of this PhD project as described in Section 2.4. Exome sequencing was performed and analysed by VCGS as described in Section 2.2. At this stage the analysis of both groups (ES+mtDNAseq and GS) was limited to coding regions and known disease genes.

7.2 Prospective AGHA mitochondrial flagship

A total of 132 cases with suspected mitochondrial disease have been analysed through genome (n=60) or exome + mtDNA sequencing (n=72). These patients were recruited from New South Wales (NSW), Queensland (QLD), South Australia (SA), Victoria (VIC), Tasmania (TAS) and Western Australia (WA). 59% of patients (n=78) had onset of symptoms by 16 years (paediatric group) and 41% of patients (n=54) presented in adulthood. The patient characteristics and molecular diagnosis by sequencing arm is summarised in Table 12.

Table 12 Mitochondrial disease flagship patient characteristics

Sequencing arm	Exome +mtDNA		Genome		Total	
	n= 72		n=60		n=132	
Gender						
Female	41	57%	39	65%	80	61%
Male	31	43%	21	35%	52	39%
Age of onset						
Paediatric	42	58%	36	60%	78	59%
Adult	30	42%	24	40%	54	41%
State of residence						
NSW	13	18%	14	23%	27	20%
QLD	20	28%	17	28%	37	28%
SA	8	11%	4	7%	12	9%
VIC	21	29%	17	28%	38	29%
WA	8	11%	8	13%	16	12%
TAS	2	3%	0	0	2	2%
Median modified Nijmegen score	6	IQR 2	6	IQR 2	6	IQR 2
Genomic diagnosis	25	35%	22	37%	47	36%

A genomic diagnosis was reached in 36% (n=47) of the total cohort; 77% (n=36) of the diagnoses were in known mitochondrial disease genes, of which 67% (n=24) were encoded in the nuclear genome, and 33% (n=12) in the mitochondrial DNA. In 23% (n=11) of the diagnoses, the causative genes were not known to have a mitochondrial function (Figure 14).

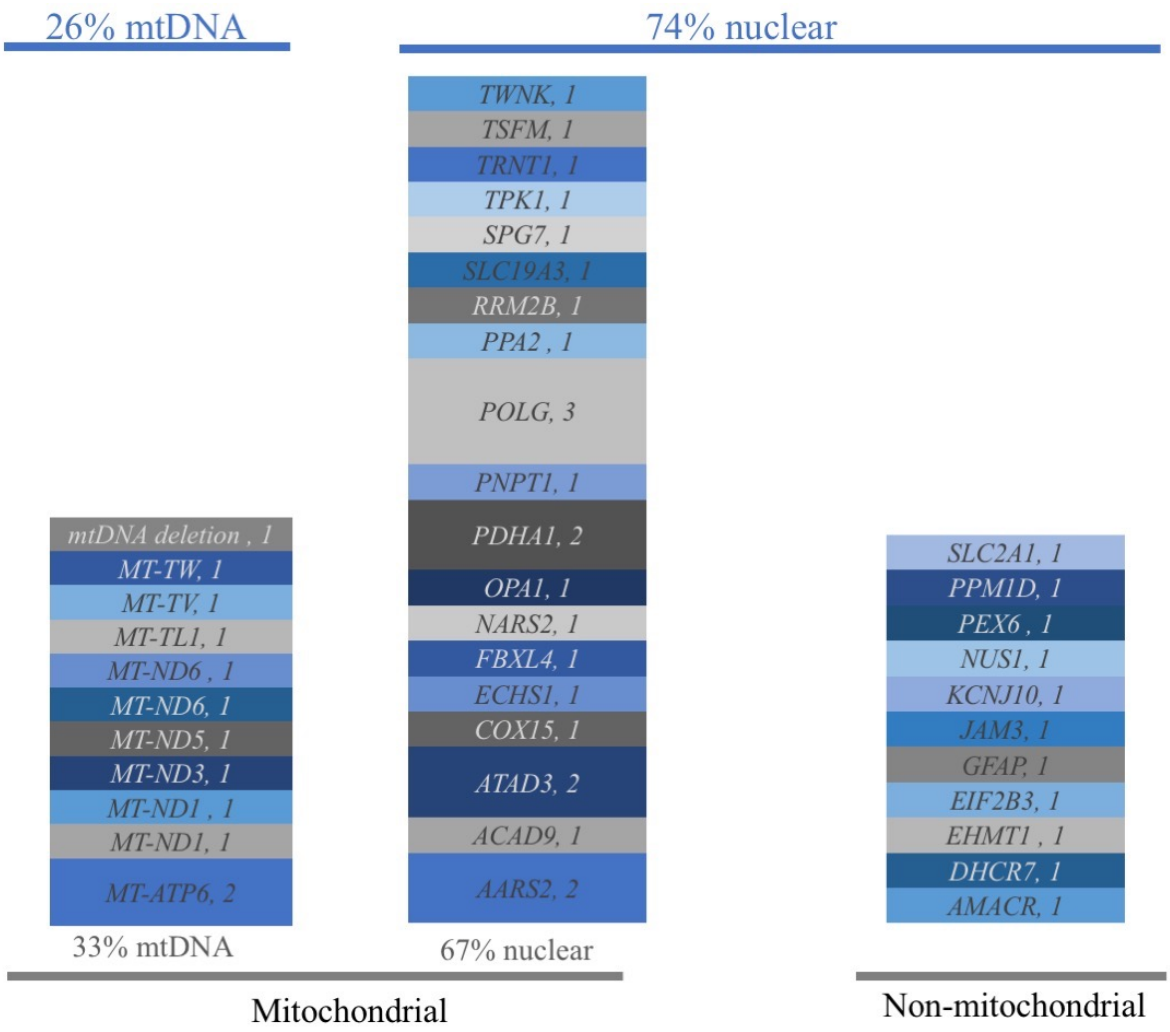

Figure 14 Molecular diagnosis in patients from the mitochondrial flagship.

In 47 patients a molecular diagnosis was identified in 40 different genes and one single large mtDNA deletion encompassing 23 genes.

Interestingly, the diagnostic yield in the cases with paediatric onset was 45% (n=35) while in the adult onset cases it was 22% (n=12).

As shown in Figure 15, the paediatric group had a higher median modified Nijmegen score (6 IQR 3) than the adult group (5 IQR 2) (p= 0.0009). However, a more severe and multisystemic presentation that could result in a higher modified Nijmegen score does not fully explain the increased diagnostic yield in children.

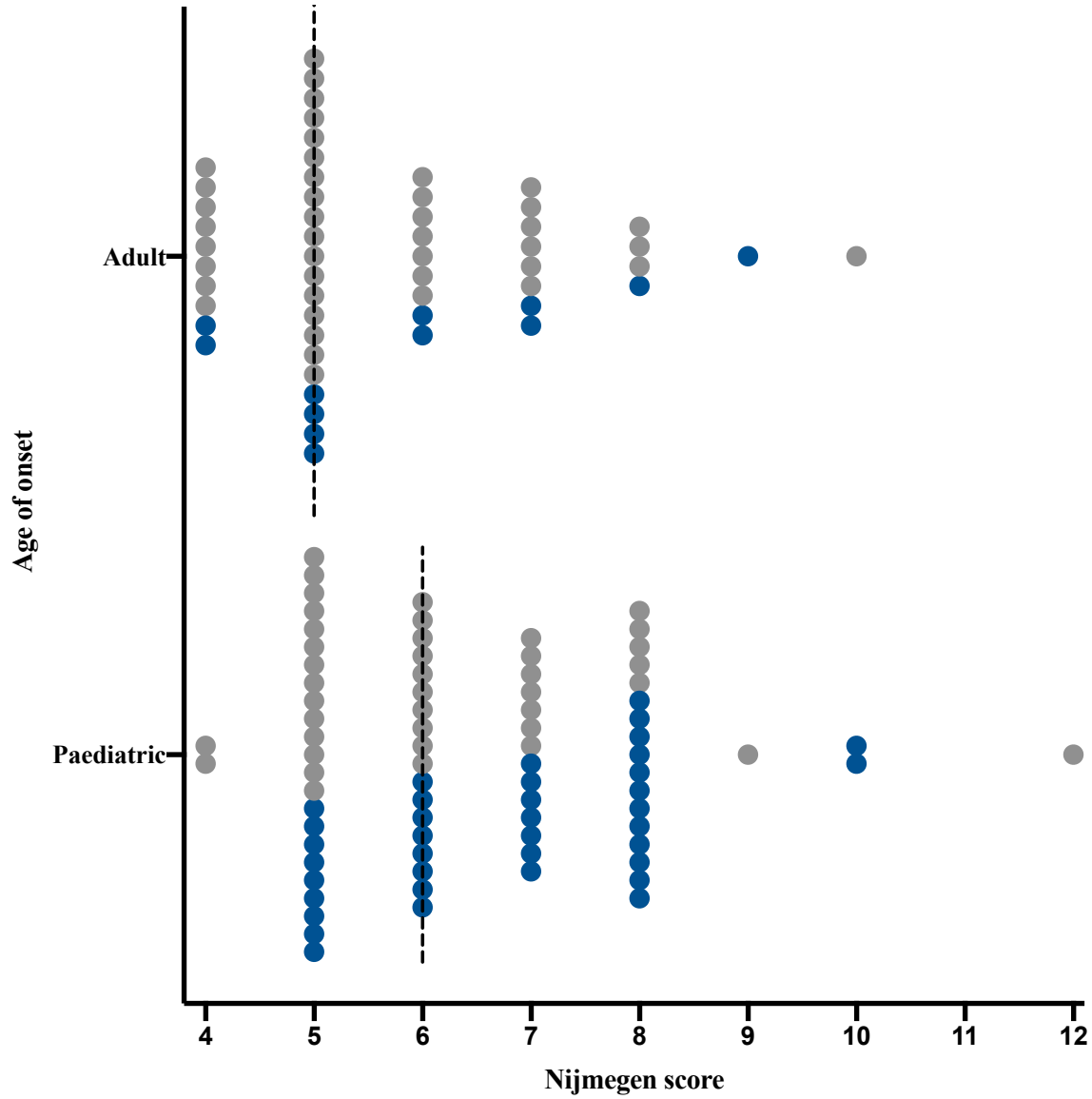

Figure 15 Modified Nijmegen score by age of onset.

The dotted lines represent the median modified Nijmegen scores in the paediatric (6 IQR 3) and adult onset groups (5 IQR 2) (p= 0.0009), the blue circles represent the cases with a molecular diagnosis, and the grey circles represent cases with no molecular diagnosis.

For instance, a "definite" modified Nijmegen score (8-12) was only associated with a molecular diagnosis in the paediatric group (odds ratio (OR) 3.43, 95% confidence interval (CI)In adults, no difference was found between "definite" or "probable" modified Nijmegen scores and the proportion of molecular diagnosis achieved (OR 1.900 95% CI:0.3209 to 9.46, p=0.6048) (Table 13).

Table 13 Molecular diagnosis by modified Nijmegen score and onset groups

	Adult				Paediatric			
Modified Nijmegen score	**No Diagnosis**	**Diagnosis**	**Total**	**p**	**No Diagnosis**	**Diagnosis**	**Total**	**p**
Definite (n=27)	4	2	6	0.604	7	14	21	0.023
Probable (n=105)	38	10	48		36	21	57	
Total	**42**	**12**	**54**		**43**	**35**	**78**	

An alternative explanation might be that in a proportion of the adult cases, variants in mtDNA have not been detected due to the use of blood as the source of DNA. In blood, the heteroplasmy levels of some mtDNA variants can decline and mtDNA deletions may be undetectable (Rahman, Poulton et al. 2001, Grady, Pickett et al. 2018). This may also explain why in the adult onset group a molecular diagnosis was identified in mtDNA in only 25% (n=3), which contrasts with prior literature that suggests that up to 75% of adult onset mitochondrial disease can result from mtDNA variants (Stenton and Prokisch 2020). Testing skeletal muscle in patients with a phenotype compatible with a mtDNA deletion or muscle-specific phenotypes could help to clarify this discrepancy.

In addition to testing relevant tissue, the apparent lower diagnosis in mtDNA in the adult group, may also be explained by the likelihood of prior targeted molecular testing in patients with well-defined phenotypes. To illustrate, Leber Hereditary Optic Neuropathy (LHON) is one of the most common adult onset mitochondrial diseases. Three variants are found in 95% of patients diagnosed with LHON: m.3460G>A in *MT-ND1*, m.11778G>A in *MT-ND4*, and m.14484T>C in *MT-ND6* (Yu-Wai-Man, Griffiths et al. 2009). However, we did not identify any of these variants in our cohort. Patients presenting with bilateral subacute visual impairment (the well-defined phenotype of LHON) were likely to have had prior targeted molecular testing rather than being recruited to our cohort.

7.3 Utility of Genome Sequencing: Case studies

While the current diagnostic yield in the mitochondrial disease flagship was comparable between the ES (35%) and GS (37%) arms, several cases from this cohort highlight the utility of GS.

7.3.1 Genome sequencing detects low levels of mtDNA heteroplasmy in blood

In an adult patient (M1) with bilateral cataracts, hearing loss, myoclonic epilepsy, peripheral neuropathy and exercise intolerance, GS identified a pathogenic heteroplasmic mtDNA variant NC_012920.1(*MT-TL1*):m.3243A>T with a 2% mutant load in blood (Figure 16). *MT-TL1* encodes a mitochondrial tRNA (Leu (UUR)). The NC_012920.1(*MT-TL1*):m.3243A>T variant is located in the D-loop of the tRNA. The A at this position has high (97.78%) evolutionary conservation.

In silico software predicts this variant to be disease-causing and it has been previously described as pathogenic in six unrelated families with different mitochondrial phenotypes including MELAS, CPEO, hearing loss with cataracts and rhabdomyolysis. The variant was heteroplasmic in blood (5-46%) and muscle (30-81%) from the affected individuals (Alston, Bender et al. 2010, Czell, Abicht et al. 2012, Ikeda, Osaka et al. 2018). Using *in vitro* transcribed tRNA molecules, Sohm et al. (2003) showed that the m.3243A>T variant had a tRNALeu(UUR) aminoacylation efficiency only 1/300th that of the wild-type tRNA (Sohm, Frugier et al. 2003). Alston et al. (2010) used single fibre studies to show that patient muscle had higher amounts of mutant mtDNA in COX-deficient fibres compared with COX-positive fibres (Alston, Bender et al. 2010). An alternative change at this nucleotide, m.3243A>G, is the most common pathogenic mtDNA variant known to cause MELAS, and is also associated with maternally inherited diabetes with or without deafness, myopathy, CPEO, and other mitochondrial cytopathies (Schaefer, McFarland et al. 2008).

.

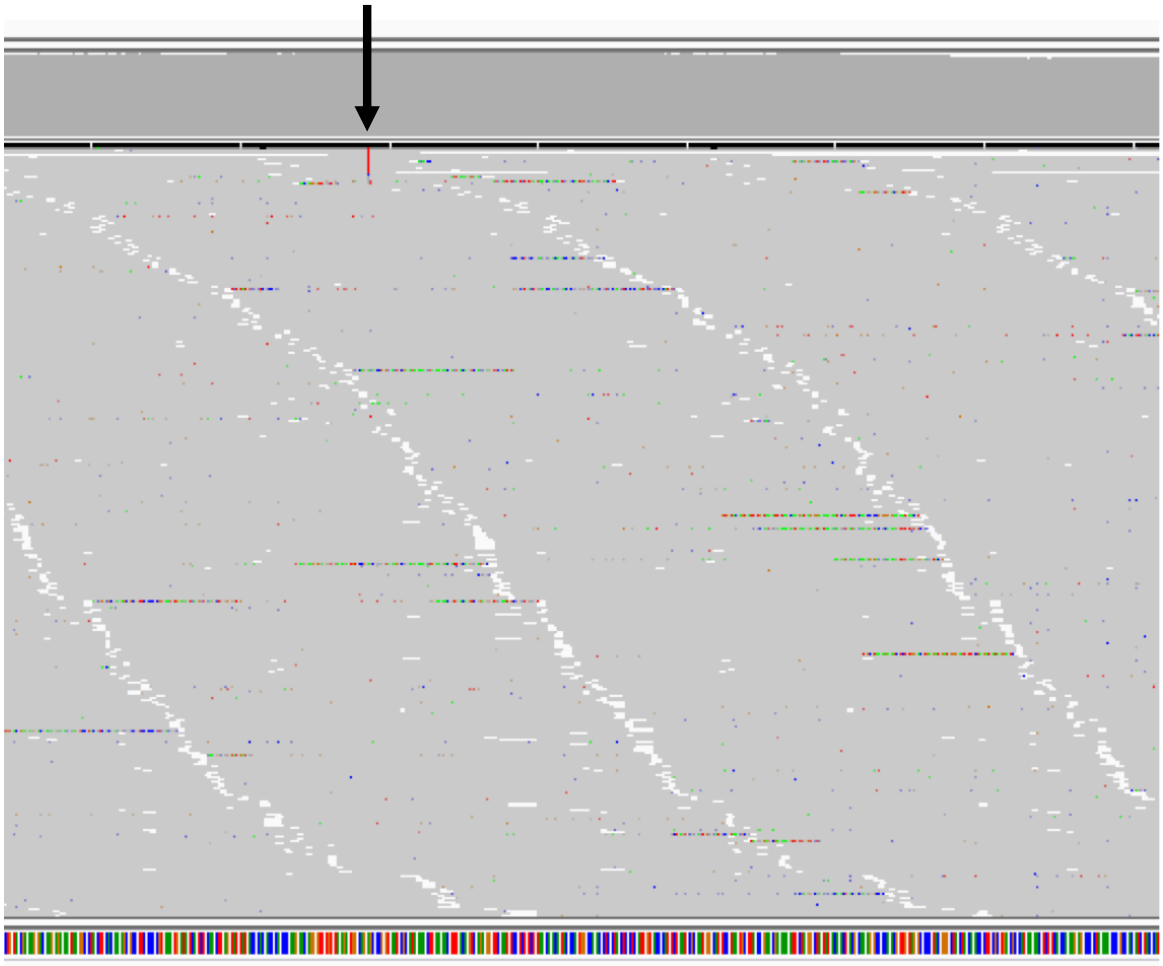

Figure 16 GS detects low levels of heteroplasmy in blood in patient M1.

IGV browser screenshot shows the identification of NC_012920.1(*MT-TL1*):m.3243A>T. The black arrow points to the variant (in red) present in 2% of the reads aligned to the m.3243 position. The total sample read depth at that position was 4235X.

This case illustrates how GS can achieve a very high mtDNA read depth that allows the detection of low-level heteroplasmy variants (Puttick, Kumar et al. 2019). This is particularly relevant when using blood as the source of DNA. For instance, it is known that the mutant load in blood of the common MELAS mutation (the alternative change of the variant identified in patient M1) decreases with age (Grady, Pickett et al. 2018). While multiple lines of evidence show that the NC_012920.1(MT-TL1):m.3243A>T variant identified in patient M1 is pathogenic, the mutant load of 2% in blood is too low to predict whether the mutation may be present at higher levels in other tissues to cause disease. Currently, this patient is being followed up by VCGS through mtDNAseq of DNA extracted from urine to determine the mutation load.

7.3.2 Genome sequencing detects an intragenic nuclear DNA deletion in *AARS2*

Patient A1, presented with respiratory distress in the newborn period, persistent metabolic acidosis, and failure to thrive. GS identified heterozygous variants in the *AARS2* gene, a missense variant NM_020745.3(*AARS2*):c.2870C>T NP_065796.1(AARS2):p.(Ser957Leu) and a 4.1kb intragenic deletion spanning exons 5 through 7. The deletion was identified through ClinSV, a bioinformatic tool that integrates depth of coverage, plus split and spanning reads to detect SV in GS data (Minoche et al., in preparation), and by visual inspection of BAM files in IGV browser (Figure 17). The deletion has been previously identified in two patients with *AARS2*-related mitochondrial disease (Sommerville, Zhou et al. 2019, Srivastava, Butala et al. 2019). However, the missense variant NM_020745.3(*AARS2*):c.2870C>T p.(Ser957Leu) has not been previously reported in other clinical cases. The serine at position 957 has very high evolutionary conservation, and the change to leucine represents a major amino acid change. *In silico* software predicts this variant to be disease-causing (SIFT, Mutation Taster, PROVEAN, CADD). It has been observed in the gnomAD population database at a frequency of 0.00002031 (5/246130) and has not been reported in a homozygous state.

Allele-specific PCR using primers designed to only amplify the allele without the deletion confirmed compound heterozygosity (Figure 18). Segregation was also confirmed by parental testing through VCGS.

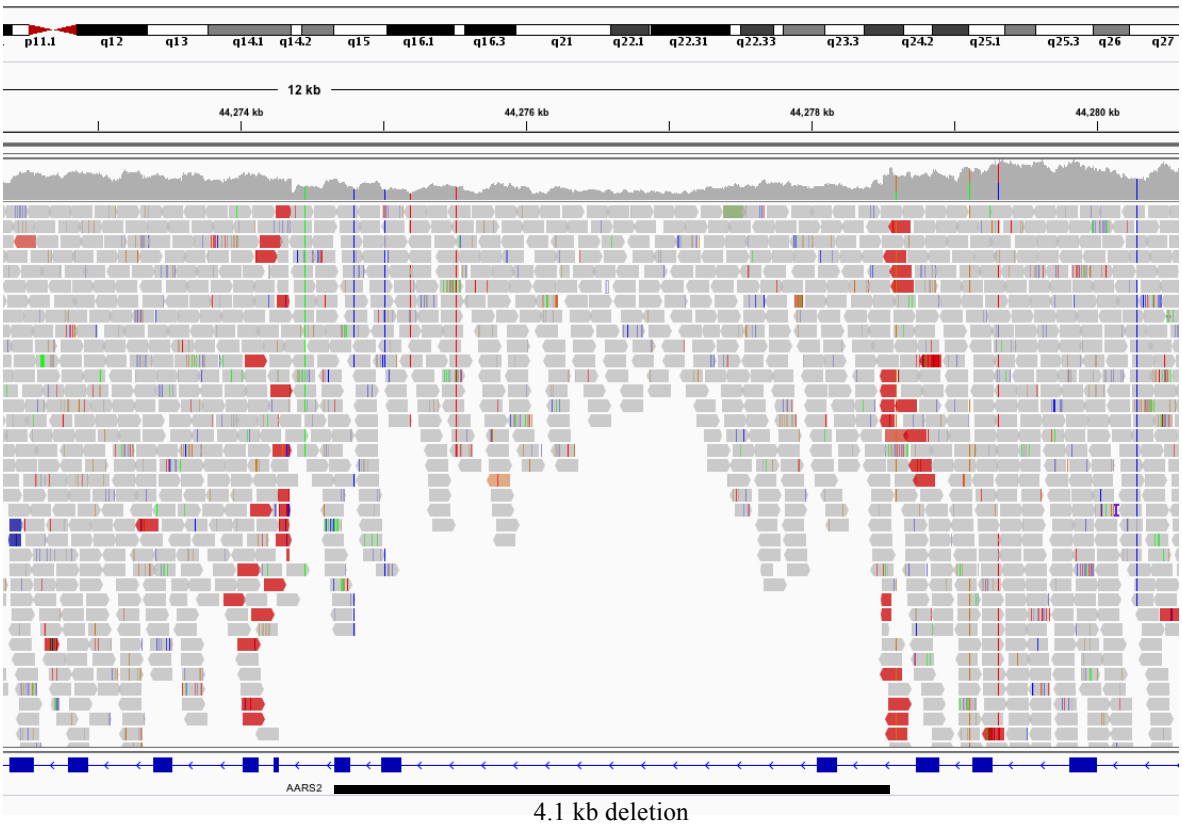

Figure 17 GS detects an intragenic nuclear DNA deletion in *AARS2* in patient A1.

IGV coverage plot shows ~ 50% less reads in the region spanning the deletion. In the alignment track the reads that are coloured red have an inferred insert size greater than expected and are therefore flanking the deletion.

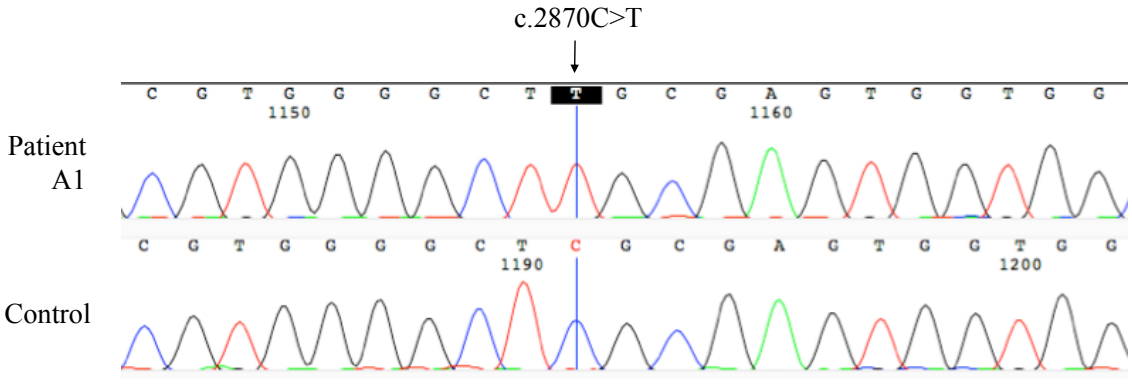

Figure 18 Allele-specific PCR confirms compound heterozygosity in *AARS2*.

Sanger sequencing shows the presence of the NM_020745.3(AARS2):c.2870C>T variant in the allele without the deletion, consistent with the variants being *in trans*.

cDNA studies performed in fibroblasts identified three different splicing transcripts resulting from the allele with the deletion, which are predicted to induce a frameshift and a premature truncation NP_065796.2:p.(Glu251Serfs*27), NP_065796.2:p.(Gly98Profs*6), or a shorter

protein lacking 257 amino acids from the aminoacyl-tRNA synthetase IIc domain NP_065796.2:p.(Glu146_Glu403del) (Figure 19).

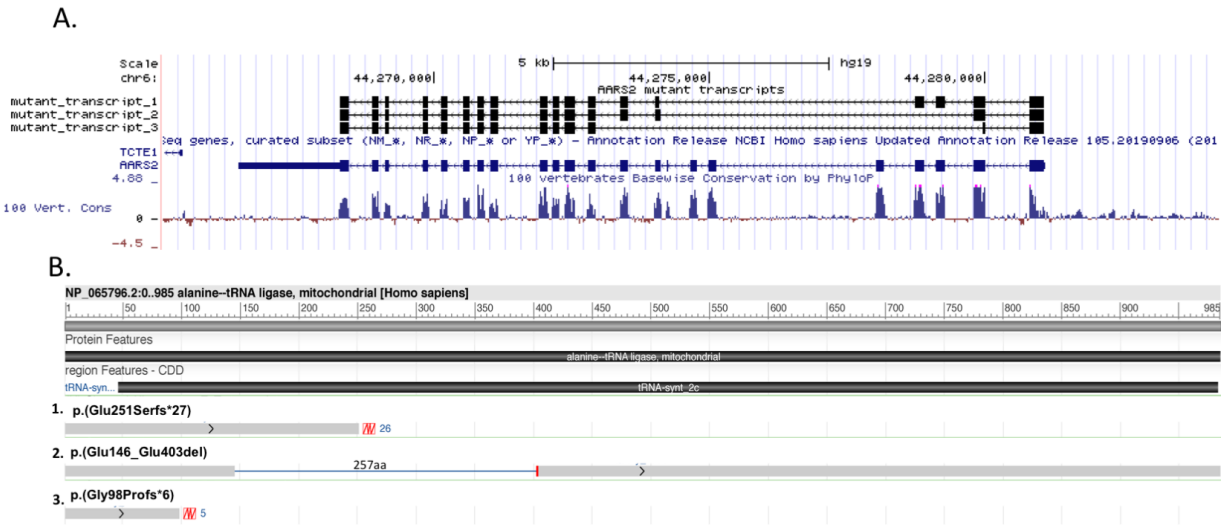

Figure 19 *AARS2* deletion leads to three mutant transcripts.

University of California, Santa Cruz (UCSC) Genome Browser screenshot showing the aligned cDNA sequences resulting from the allele with the deletion. The mutant transcript 1 resulted from a splicing event from exon 4 to exon 9, mutant transcript 2 from exon 2 to exon 9 and mutant transcript 3 from a cryptic splice site in exon 2 to exon 11 (A). At a protein level, the National Center for Biotechnology Information (NCBI) Sequence Viewer graphical display shows the mutant transcripts 1 and 3 are predicted to result in a frameshift and a truncation, and mutant transcript 2 to result in a shorter protein lacking 257 amino acids from the functional domain (B).

Western blotting of fibroblasts using total Abcam OXPHOS human WB antibody cocktail in fibroblasts showed a decreased protein expression of OXPHOS CI and CIV subunits in the patient A1 compared to control cells (Figure 20). Patient A2 (AARS2 control) with likely pathogenic *AARS2* variants was recently published by Hock and colleagues (Hock, Reljic et al. 2020) and was used as a positive control.

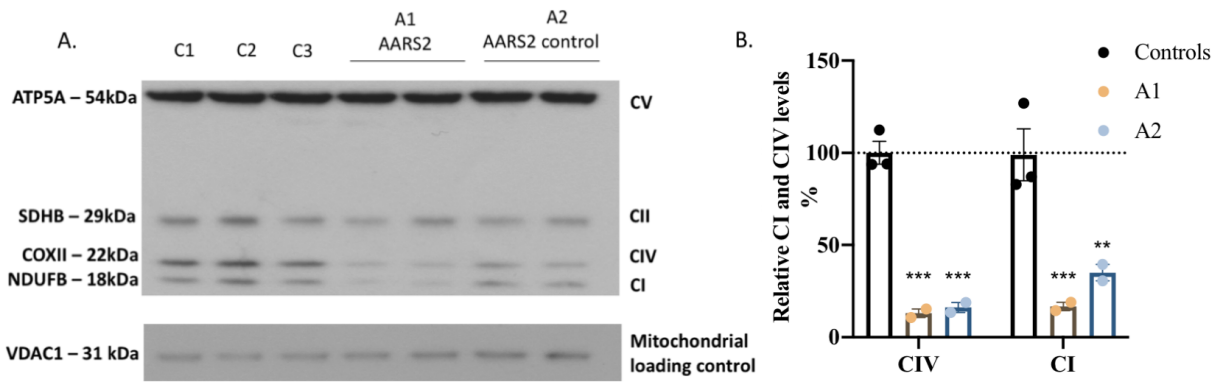

Figure 20 OXPHOS protein expression in patient A1 *AARS2* and control fibroblasts.

(A) Representative Western blot showing decreased protein levels of NDUFB8 (CI) and COXII (CIV) in patient A1 similar to patient A2 (positive control) who harbors likely pathogenic *AARS2* variants c.1874G>A p.(Arg625His) and c.665C>T p.(Thr222Ile) in *trans*, VDAC1 (porin) was used as a mitochondrial loading control. (B) Densitometry analysis suggests 85% lower levels of COXII (CIV) in both A1 and A2, and suggested 85% lower NDUFB8 (CI) in A1 and 65% lower in A2 relative to controls (***p<0.001, **< 0.01).

This case has shown the utility of GS and the related bioinformatic algorithms such as ClinSV to detect nuclear SVs. This approach allowed the prioritization and subsequently diagnostic confirmation of *AARS2* variants as causative of the clinical features in patient A1.

7.3.3 GS detects a mtDNA deletion in blood

Patient D1 is a 13-year-old female with short stature, progressive external ophthalmoplegia, ptosis, retinitis pigmentosa, and insulin dependent diabetes mellitus. By GS, a single large-scale pathogenic deletion in the mtDNA at a level of heteroplasmy in blood of 50% was detected (Figure 21 A), this deletion is 9031bp in size and spans 23 mitochondrial genes (Figure 21 B) (*MT-CO1* to *MT-CYB*). Several deletions that overlap the patient deletion have been previously reported in the MitoBreak database associated with Mitochondrial DNA Deletion Syndromes (Damas, Carneiro et al. 2014).

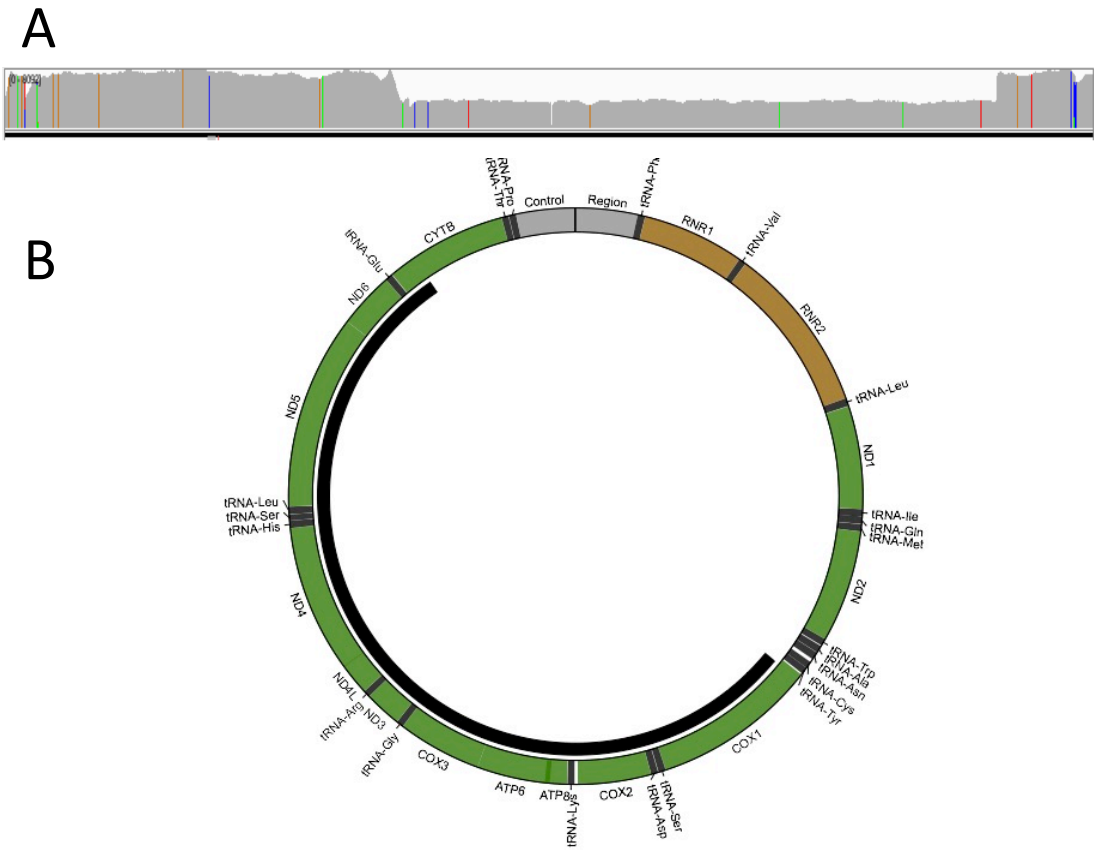

Figure 21 Deletion in the mtDNA identified by GS.

(A) IGV coverage plot shows 50% fewer reads in the area spanning 9031bp of the mtDNA. (B) MitoBreak figure depicts the single large-scale heteroplasmic deletion spaning 23 mitochondrial genes (black bar).

One advantage of PCR-free GS, in the identification of mtDNA deletions, is that it can provide an estimated level of heteroplasmy by looking at read depth (Figure 21 A). Alternative MPS approaches for whole mtDNA genome sequencing often require a single long-range PCR step. This can likely introduce amplification bias due to preferential replication of shorter amplicons, potentially overestimating the heteroplasmy levels of a mtDNA deletion (Zambelli, Vancampenhout et al. 2017). In order to accurately determine heteroplasmy levels, additional quantitative methods can be used (e.g. Taqman real-time PCR assays) (Grady, Murphy et al. 2014). Another challenge with PCR based approaches is that the amplification efficiency can be influenced if SNVs are located at primer binding sites (Cui, Li et al. 2013).

The case reported here illustrates the feasibility of GS in identifying and determining heteroplasmy levels of a mtDNA deletion in blood without requiring additional testing. It is important to note that mtDNA deletions can be detectable in the blood in the paediatric

population (Broomfield, Sweeney et al. 2015), such as in patient D1 (13-year-old). However, in adults, mtDNA deletions are often undetectable and are preferentially found in post-mitotic tissues. Consequently, other tissues such as skeletal muscle would be a better source of DNA for testing (Sadikovic, Wang et al. 2010).

The natural history of some patients with Pearson syndrome is a good illustration of the shift in heteroplasmy levels of single large mtDNA deletions in mitotic tissues. Pearson syndrome presents in infants with a severe multisystemic disease with predominantly haematological manifestations due to a single large mtDNA deletion detectable in blood (Lee, Lee et al. 2007). Interestingly, patients that survive childhood can spontaneously recover from the hematologic dysfunction and the mtDNA deletion is lost from the blood. However, it can be identified in post-mitotic skeletal muscle, and the phenotype can evolve over time to KSS (Broomfield, Sweeney et al. 2015, Farruggia, Di Cataldo et al. 2016)

7.3.1 GS aids in the diagnosis of diseases that show clinical overlap with mitochondrial disease and in variant phasing

A female patient (P1) presenting at 3 years of age with sensorineural hearing loss, who subsequently developed primary ovarian insufficiency (POI) at 13 years of age, was clinically diagnosed with Perrault syndrome. Additional clinical features developed, including peripheral neuropathy, tremor and ataxia, with brain MRI showing leukodystrophy. From the age of 17 years she started showing signs of regression and also developed pigmentary retinopathy that led to a rapid loss of vision.

GS identified two variants in *PEX6*, NM_000287.3:c.2356C>T NP_000278.3: p.(Arg786Trp) and NM_000287.3:c.371T>C NP_000278.3(PEX6):p.(Leu124Pro) a likely pathogenic and a variant of uncertain significance, respectively (Figure 22). The *PEX6* gene is associated with an autosomal recessive peroxisomal disease and has not been previously linked to Perrault syndrome (Ebberink, Kofster et al. 2010). This prompted analysis of very-long-chain fatty acids (VLCFA) in plasma, which were consistent with a defect in VLCFA oxidation (C26:C22 of 0.049; normal range 0–0.030, and C24:C22 of 0.993; normal range 0.550–1.050) (De Biase, Tortorelli et al. 2020).

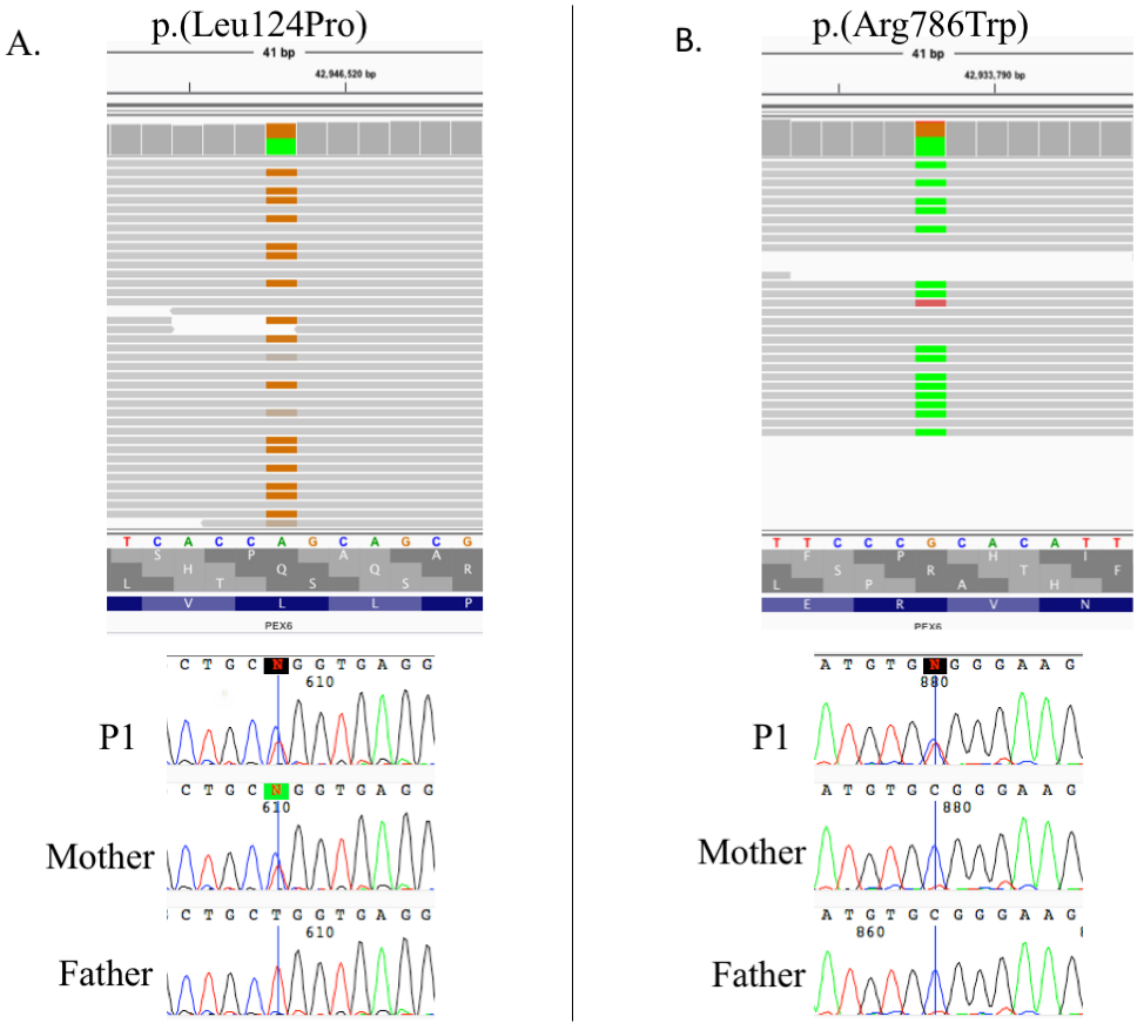

Figure 22 Identification of *PEX6* variants.

(A) IGV browser and sanger confirmation shows the identification of NM_000287.3:c.371T>C NP_000278.3(PEX6):p.(Leu124Pro), segregating with the mother. (B) IGV browser and sanger confirmation shows the identification of NM_000287.3:c.2356C>T NP_000278.3: p.(Arg786Trp) only present in proband P1.

Sequencing of parental DNA indicated that the NP_000278.3(PEX6):p.(Leu124Pro) variant was inherited from the mother (Figure 22 A), but the NP_000278.3(PEX6):p.(Arg786Trp) variant was not identified in either parent (Figure 22 B). To confirm paternity, VCGS performed a microarray in the father and compared it with 61040 SNVs in the proband using the GS data. This yielded a kinship coefficient of 0.24, which would be expected for a first-degree relationship (Manichaikul, Mychaleckyj et al. 2010), thus confirming the parent-proband relationship, suggesting that the NP_000278.3(PEX6):p.(Arg786Trp) variant arose from a *de novo* event.

Confirmation that the *de novo* variant arose on the paternal allele was needed to confirm compound heterozygosity in the proband, so a population and read-base haplotype phasing approach (Choi, Chan et al. 2018), followed by allele-specific PCR, was used to phase the variants. In brief, all SNVs identified in the *PEX6* gene were interrogated for linkage disequilibrium (LD) and haplotype frequencies with LDlink tool (Machiela and Chanock 2015). Genotyping parental tag SNVs in *PEX6* revealed that the parents and the proband were heterozygous carriers of a common haplotype— "Haplotype A", with a population frequency of 29% (1000 Genomes Project V5) (Figure 23 A).

The *PEX6* SNVs were also inspected using IGV browser to confirm haplotype blocks when the SNVs were covered by a single read. Inspection of the proband's GS reads showed the maternal NP_000278.3(PEX6):p.(Leu124Pro) variant was *in trans* with haplotype A, indicating that haplotype A was inherited from the father (Figure 23 B). Compound heterozygosity was confirmed by using allele-specific PCR and amplifying only the DNA containing the *de novo* or the WT variant, showing that the *de novo* NP_000278.3(PEX6):p.(Arg786Trp) was *in cis* with the haplotype inherited from the father (Figure 24).

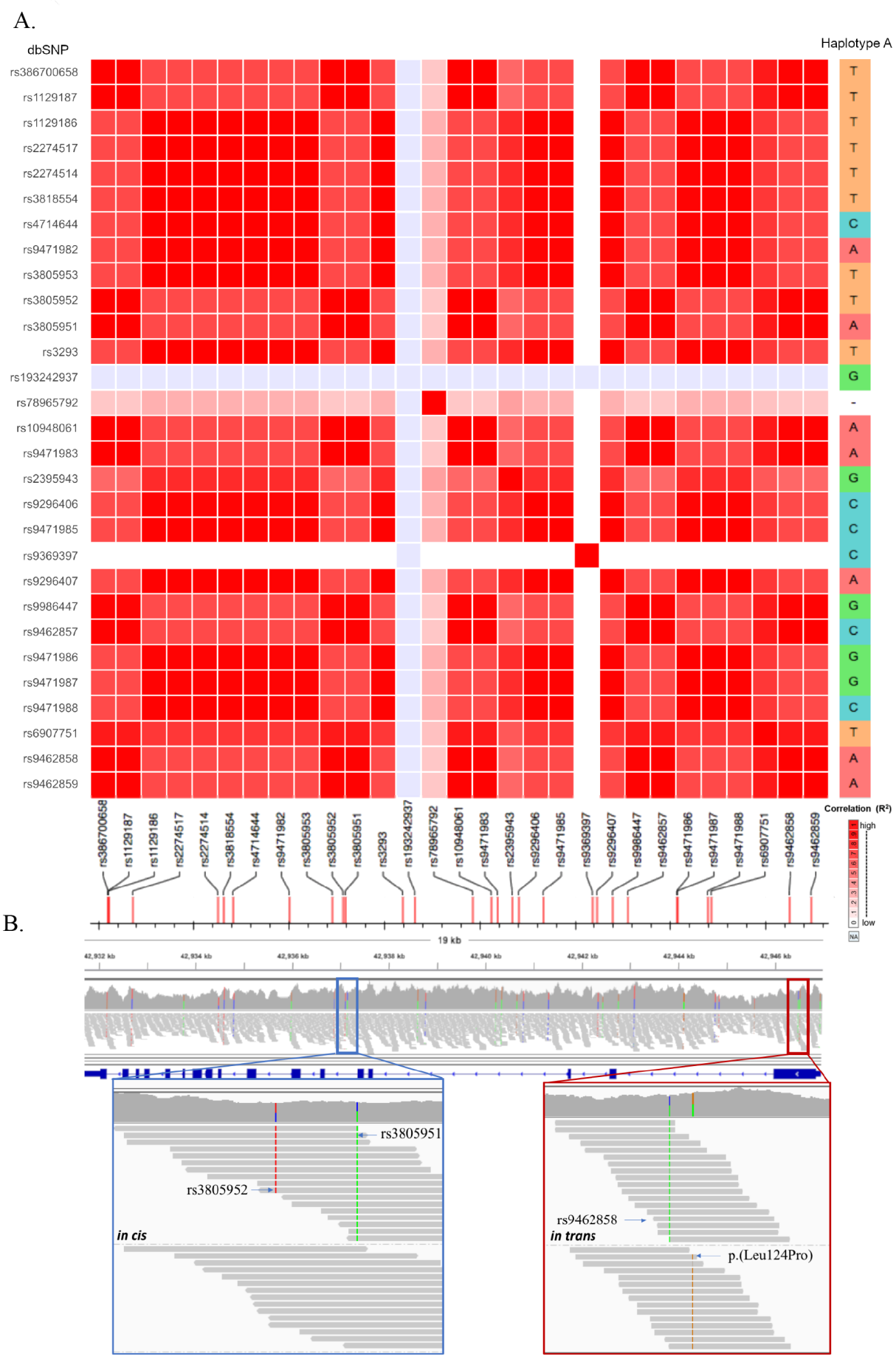
A.
dbSNP
Haplotype A
rs386700658
rs1129187
rs1129186
rs2274517
rs2274514
rs3818554
rs4714644
rs9471982
rs3805953
rs3805952
rs3805951
rs3293
rs193242937
rs78965792
rs10948061
rs9471983
rs9471985
rs9369397
rs9296407
rs9986447
rs9462857
rs9471986
rs9471987
rs9471988
rs6907751
rs9462858
rs9462859
Correlation (R²)
high
low
B.
19 kb
rs3805951
rs3805952
in cis
rs9462858
in trans
p.(Leu124Pro)

Figure 23 Population and read-based haplotype phasing of *PEX6* variants.

(A) Linkage disequilibrium (LD) and correlation matrix using variants called in *PEX6* by GATK HaplotypeCaller, which correlates with a known common haplotype ("haplotype A"). (B) The IGV browser view shows the GS reads in *PEX6*, the zoom panel highlighted in blue is representative of inspected regions where heterozygous SNVs covered by single reads confirmed haplotype blocks (rs3805952 and rs3805951 from haplotype A are in *cis*). The zoom panel highlighted in red shows the p.(Leu124Pro) variant inherited from the mother to be in *trans* with haplotype A rs9462858. This figure was adapted from FS1 (Tucker, Rius et al. 2020).

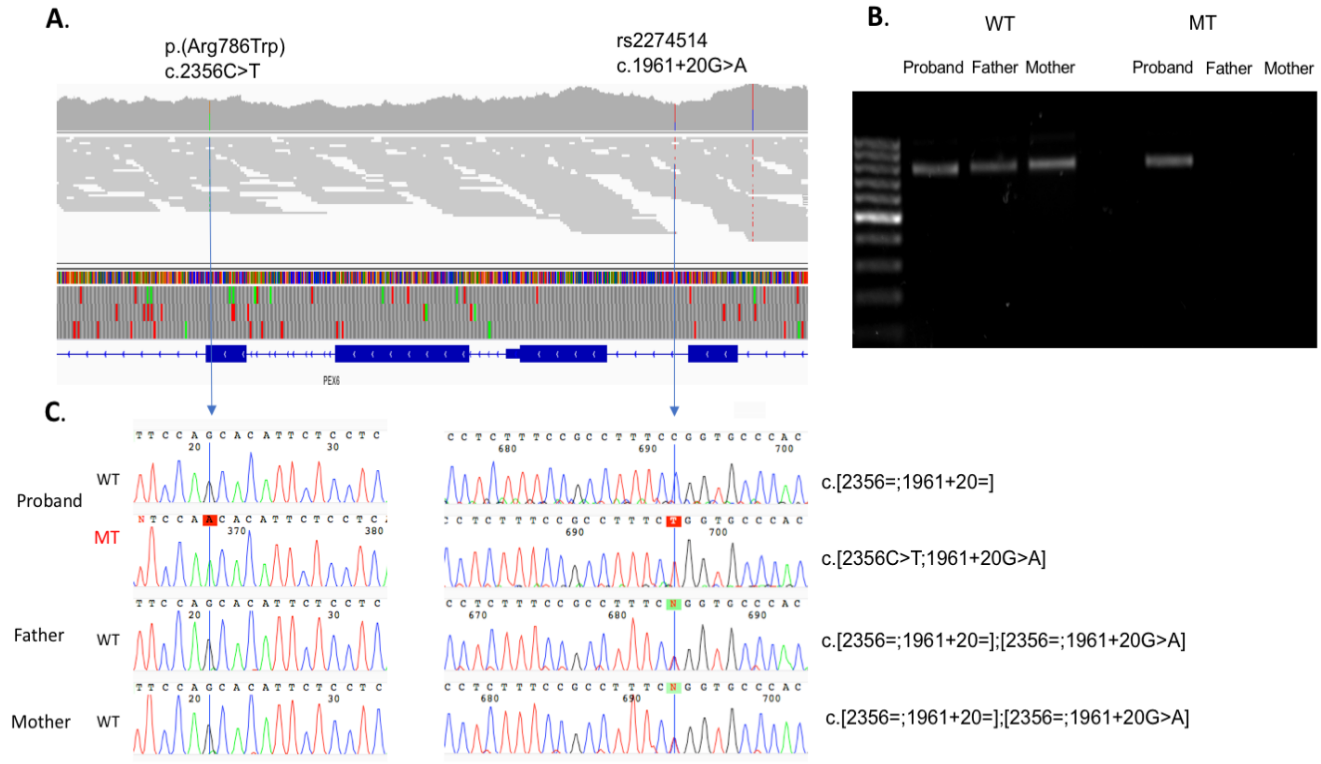

Figure 24 Allele-specific PCR confirms compound heterozygosity in *PEX6*.

(A) IGV browser screenshot from the proband's GS reads showing the *de novo* heterozygous c.2356C>T p.(Arg786Trp) variant and the heterozygous c.1961+20G>A (rs2274514) variant from haplotype A. (B) Gel electrophoresis of allele-specific PCR products shows amplification of the c.2356= wild type (WT) allele in the proband and parents, and the c.2356C>T mutant (MT) allele only in the proband. (C) Sanger sequencing of the allele specific products confirms the *de novo* c.2356C>T ; p.(Arg786Trp) variant to be in *cis* with haplotype A rs2274514 SNP NC_000006.11:g.42934500C>T c.1961+20G>A. Figure was adapted from FS2 (Tucker, Rius et al. 2020).

The confirmation of compound heterozygosity and the abnormal VLCFA profile in plasma confirmed the molecular diagnosis of patient P1 in *PEX6,* highlighting the clinical overlap between mitochondrial diseases and peroxisomal disorders.

7.4 Conclusion

Using an MPS approach, the mitochondrial flagship has identified the molecular diagnosis in 36% of patients in the cohort. At this stage of analysis there is no significant difference between the diagnostic yield in the ES+mtDNAseq (35%) and GS (37%) arms (p=0.8565). It is important to take in to account that at this stage of analysis, the GS data were interrogated only for variants in coding regions of known disease genes, therefore, most of the variants in the GS arm could potentially also be identified by ES+ mtDNAseq. Further research is ongoing to study non-coding regions in the GS data that could potentially increase the diagnostic yield in this group.

In paediatric, but not adult, patients a "definite" modified Nijmegen score was associated with a higher rate of molecular diagnosis. The diagnostic yield was also higher than the adults independently of their modified Nijmegen score classification. Testing other tissues, such as skeletal muscle, could clarify if the use of blood as the source of DNA in adults in this study could be partially contributing to the lower diagnostic yield. Another contributing factor could be ascertainment bias, whereby patients with well-defined mitochondrial phenotypes would be more likely to have had prior molecular testing rather than being recruited as adults to our cohort.

Interestingly 23% of the molecular diagnoses were in non-mitochondrial disease genes, causing predominantly neurological phenotypes. In this category there were even patients with biochemical evidence of mitochondrial disease. For instance, a patient with persistent 3-methylglutaconic aciduria (3MGA), which is often associated with phospholipid remodelling or mitochondrial membrane-associated disorders (Wortmann, Duran et al. 2013), was instead diagnosed with Kleefstra syndrome due to a variant in *EHMT1,* a gene not known to cause mitochondrial disease (Willemsen, Vulto-van Silfhout et al. 2012). Upon re-evaluation the patient had a strong phenotype match with Kleefstra syndrome and had no variants in mitochondrial genes typically associated with 3MGA-uria that could suggest a blended phenotype. Testing for 3MGA in patients with Kleefstra syndrome could clarify if there is an underlying secondary mechanism associated with the persistent 3MGA-uria. This highlights

the utility of non-targeted MPS approaches that can expand the genomic analysis to genes that represent a differential diagnosis to mitochondrial diseases.

Chapter 8 Discussion and conclusion

Mitochondrial diseases can be caused by nuclear or mitochondrially-encoded genes that often lead to multisystemic and progressive symptoms, making establishment of a molecular diagnosis difficult. This PhD has shown the feasibility of GS to diagnose patients with suspected mitochondrial disease by studying patients from a trio GS and a singleton mitochondrial flagship cohort.

The modified Nijmegen score was used in both cohorts as a key inclusion criterion for recruitment and to categorise patients in terms of the prior likelihood of them having mitochondrial disease at recruitment. Interestingly, the results from the mitochondrial flagship singleton cohort showed that in paediatric, but not adult, cases a positive molecular diagnosis was associated with a higher modified Nijmegen score (67% in "definite" group vs 37% in "possible/probable" (OR 3.43, 95% CI:1.2 to 9.0, p=0.0230) (Figure 25 A). In the paediatric trio GS cohort, no statistically significant difference in achieving a genetic diagnosis was seen in the "definite" modified Nijmegen score group with 70% diagnosis, compared to 48% in the probable/possible group (OR 2.62, 95% CI:0.7284 to 9.631, p=0.2024) (Figure 25 B).

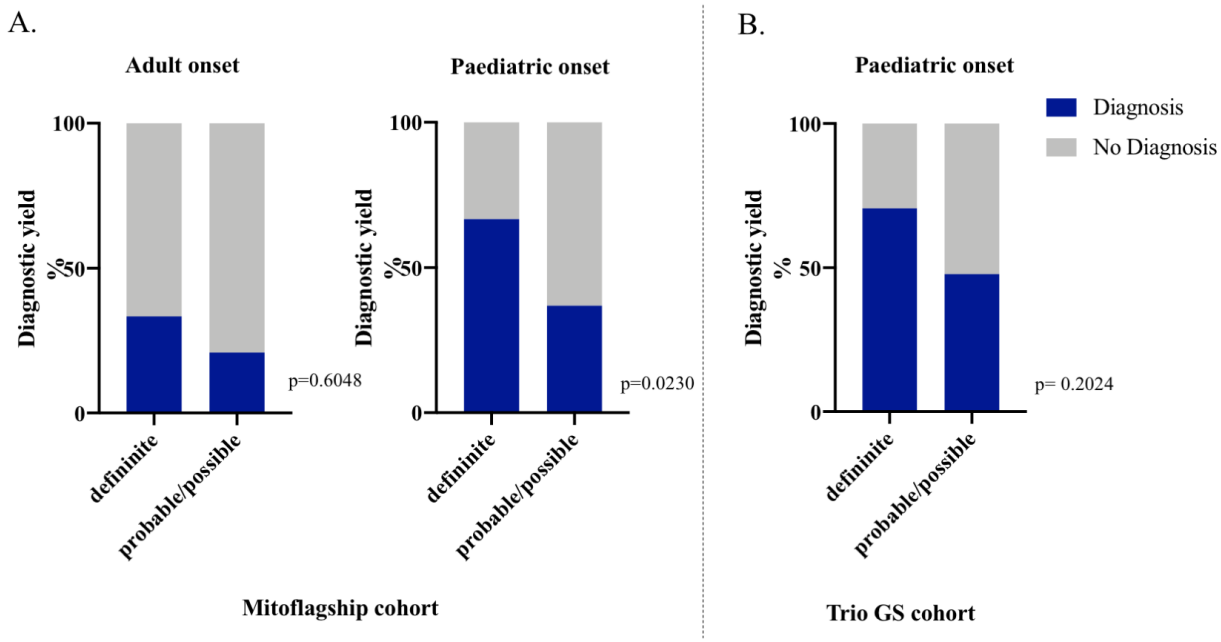

Figure 25 Modified Nijmegen score and diagnostic yield in patients from the singleton mitochondrial flagship and trio GS cohorts.

(A) Modified Nijmegen score was not associated with a molecular diagnosis in the patients with onset in adulthood. In the paediatric onset group, a significant increase in the rate of genetic diagnoses was seen in the "definite" modified Nijmegen score group with 67 % diagnosis, compared to 37 % in the probable/possible group. (B) In the

paediatric trio GS cohort, no significant difference in genetic diagnosis was seen in the "definite" modified Nijmegen score group with 70 % diagnosis, compared to 48 % in the probable/possible group.

Interestingly, all the diagnoses to date in patients from the mitochondrial flagship with a "definite" modified Nijmegen score have been in mitochondrial disease genes (n=16). Conversely, all the diagnoses in non-mitochondrial disease genes (n=11) have been in patients with a "probable/possible" modified Nijmegen score. In the trio GS, only one non-mitochondrial diagnosis was made in a patient with a "definite" modified Nijmegen score (*EPG5*, patient 22 from (Riley, Cowley et al. 2020). This suggests that a higher modified Nijmegen score is more likely to select patients with primary mitochondrial disease. These findings are consistent with those of Pronicka and colleagues, who studied a cohort of paediatric patients with suspected mitochondrial disease analysed through ES. They found a higher modified Nijmegen score was associated with a higher diagnostic rate in mitochondrial disease genes (Pronicka, Piekutowska-Abramczuk et al. 2016). In their study, the Nijmegen criteria were modified by not including the results of muscle biopsy. Consequently, the patients with the highest scores were from 5–8 and classified as "high probability" of mitochondrial disease. In this group, the diagnostic yield was 90% (n=36), with only one diagnosis made in a non-mitochondrial disease gene (Pronicka, Piekutowska-Abramczuk et al. 2016).

Overall, there seems to be some evidence to indicate that the modified Nijmegen score is a valuable tool in predicting patients more likely to reach a molecular diagnosis in a mitochondrial disease gene. To date, ~300 of the ~1500 genes known to have a mitochondrial function have been associated with mitochondrial disease (Calvo, Clauser et al. 2016, Stenton and Prokisch 2018). Therefore, patients with high modified Nijmegen scores who are still molecularly undiagnosed could be prioritised for gene discovery efforts targeting genes that encode the mitochondrial proteome.

Patients from the mitochondrial flagship cohort will be further analysed with an expanded "MitoExome" gene list that includes known and candidate genes encoding mitochondrial proteins. Therefore, it would be reasonable to expect an increase in the diagnostic yield by uncovering novel mitochondrial disease genes, particularly in patients with high Nijmegen scores. In the expanded analysis, the GS arm has the advantage of the possibility to interrogate non-coding regions and a more accurate SV detection.

In both cohorts, paediatric patients with lower modified Nijmegen scores were more likely to receive a diagnosis in a non-mitochondrial gene than the patients with high Nijmegen scores. The possibility of a non-mitochondrial disorder is important to consider when providing counselling to patients and their families. For instance, when patients are suspected to have a mitochondrial disease but are still molecularly undiagnosed, Parikh and colleagues suggest avoiding use of the term "possible mitochondrial disease" and rather use the term "diagnosis uncertain" to avoid anxiety or inappropriate management (Parikh, Karaa et al. 2019).

The molecular diagnosis of adult-onset cases in the mitochondrial flagship has been particularly challenging, even in patients with high modified Nijmegen scores. This may be due to the use of blood as source of DNA, and how patients are selected, as discussed in Chapter 7. It could also reflect the complicated analysis of mtDNA variants due to heteroplasmy, tissue-specificity, and the lack of mtDNA variant data in large population references such as gnomAD (Shen, Attimonelli et al. 2018).

The results outlined in this thesis provide important insights into evaluating different testing strategies between paediatric and adult-onset patients. For instance, in patients with a CPEO phenotype it would be reasonable to start the analysis with deletion testing in muscle, and if negative, then continue the analysis through GS. A MPS-based testing in muscle could also be considered as a first step in such patients with the advantage of potentially detecting both deletions and SNVs that could be missed in blood. Testing muscle in the patients with an initial negative result in blood from the mitochondrial flagship will help clarify if such a strategy could be more effective.

On the other hand, invasive studies in paediatric-onset cases could be initially avoided by starting the GS analysis in DNA extracted from blood. However, fibroblasts or other relevant tissues might still be needed to confirm pathogenicity of novel variants or novel disease genes as highlighted in Section 3.2 and in Chapter 5.

The current diagnostic yield of 36% in the mitochondrial flagship and 58% in the trio GS cohort, are higher than with a traditional (pre-MPS) diagnostic pathway (Neveling, Feenstra et al. 2013) and similar to the 30-60% diagnostic yield achieved in other cohorts studied using MPS technologies (Stenton and Prokisch 2018). However, 42-64% of the patients currently remain molecularly undiagnosed. A major challenge of GS analysis is the prioritization and

interpretation of the ~4 million variants per genome; it can be especially complicated to interpret variants in non-coding regions where the current understanding of variation is limited (Smedley, Schubach et al. 2016).

The use of GS in combination with other technologies and "omic" approaches could help to overcome this limitation. For instance, RNAseq transcriptome studies can detect aberrant gene expression, mono-allelic expression or splicing defects (Kremer, Wortmann et al. 2018). Quantitative proteomics can detect protein loss and characteristic protein signatures (Lake, Webb et al. 2017). Metabolomics, lipidomics and glycomics can detect signature metabolite profiles and biomarkers (Coene, Kluijtmans et al. 2018, Wanders, Vaz et al. 2019). However, tissue-specificity, inter-individual variability, and data analysis can be challenging when analysing data from such "omic" approaches (Stenton and Prokisch 2020).

It is important to consider that these complementary techniques rely on the availability of high-quality patient biological samples which highlights the importance of optimal storage of patient tissue samples and cell lines following biobanking best practices (Coppola, Cianflone et al. 2019).

Nevertheless, the use of GS in combination with these "omic" technologies can be useful not only for confirming pathogenicity of variants but can also be used to identify and prioritize candidate genes and variants as illustrated in Section 3.2.1. Large systematic studies in the context of mitochondrial diseases are needed to better determine the utility of these complementary "omic" approaches as diagnostic tools (Stenton and Prokisch 2020).

Other emerging tools that could potentially increase the current diagnostic yield are long-read sequencing technologies. Several different platforms are available (e.g., Oxford Nanopore Technologies, Pacific Biosciences) and can generate reads of 10-1000 kb (Logsdon, Vollger et al. 2020). Two major advantages of this technology include genome phasing and increased sensitivity of SV detection. For instance, it is estimated that long-read sequencing can detect up to 2.8 times more SVs than short-read GS (Eichler 2019). Base calling accuracy with long-read sequencing is improving, but is yet not comparable to short-read sequencing, and the cost is still significantly higher (Logsdon, Vollger et al. 2020). Therefore, for selected patients, it is reasonable to consider this technology as a complementary strategy to short-read GS or to resolve SV's.

Taken together, the results of this PhD support the diagnostic utility of genome sequencing for the diagnosis of patients with suspected mitochondrial disease. They also provide additional evidence to guide future strategies in the implementation of GS in the clinic. More research is required to determine the efficacy of GS in shortening the diagnostic odyssey, changing management, improving quality of life, and reducing costs to accurately measure the clinical impact and optimal testing strategies in patients with suspected mitochondrial disease. The AGHA mitochondrial flagship has the capacity and is currently in the process of collecting data to answer those research questions. In particular, in collaboration with Dr. Ilias Goranitis and Eunice Wu (Health Economics Unit, University of Melbourne), we are undertaking studies to capture health costs associated with mitochondrial diseases and cost-effectiveness analysis of GS.

References

Abyzov, A., A. E. Urban, M. Snyder and M. Gerstein (2011). "CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing." Genome Res **21**(6): 974-984.

Alston, C. L., A. Bender, I. P. Hargreaves, H. Mundy, C. Deshpande, T. Klopstock, R. McFarland, R. Horvath and R. W. Taylor (2010). "The pathogenic m.3243A>T mitochondrial DNA mutation is associated with a variable neurological phenotype." Neuromuscul Disord **20**(6): 403-406.

Alston, C. L., M. C. Rocha, N. Z. Lax, D. M. Turnbull and R. W. Taylor (2017). "The genetics and pathology of mitochondrial disease." J Pathol **241**(2): 236-250.

Australian Genomics Health Alliance, A. (2020). "Australian Genomics." Retrieved 05/08/2020, 2020, from https://www.australiangenomics.org.au.

Balsa, E., R. Marco, E. Perales-Clemente, R. Szklarczyk, E. Calvo, M. O. Landazuri and J. A. Enriquez (2012). "NDUFA4 is a subunit of complex IV of the mammalian electron transport chain." Cell Metab **16**(3): 378-386.

Bentley, D. R., 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, A. Spiridou, 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, E. C. M. Chiara, 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 and A. J. Smith (2008). "Accurate whole human genome sequencing using reversible terminator chemistry." Nature **456**(7218): 53-59.

Bernier, F. P., A. Boneh, X. Dennett, C. W. Chow, M. A. Cleary and D. R. Thorburn (2002). "Diagnostic criteria for respiratory chain disorders in adults and children." Neurology **59**(9): 1406-1411.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Bourens, M. and A. Barrientos (2017). "A CMC1-knockout reveals translation-independent control of human mitochondrial complex IV biogenesis." EMBO Rep **18**(3): 477-494.

Bris, C., D. Goudenege, V. Desquiret-Dumas, M. Charif, E. Colin, D. Bonneau, P. Amati-Bonneau, G. Lenaers, P. Reynier and V. Procaccio (2018). "Bioinformatics Tools and Databases to Assess the Pathogenicity of Mitochondrial DNA Variants in the Field of Next Generation Sequencing." Front Genet **9**: 632.

Broomfield, A., M. G. Sweeney, C. E. Woodward, C. Fratter, A. M. Morris, J. V. Leonard, L. Abulhoul, S. Grunewald, P. T. Clayton, M. G. Hanna, J. Poulton and S. Rahman (2015). "Paediatric single mitochondrial DNA deletion disorders: an overlapping spectrum of disease." J Inherit Metab Dis **38**(3): 445-457.

Calvo, S. E., K. R. Clauser and V. K. Mootha (2016). "MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins." Nucleic Acids Res **44**(D1): D1251-1257.

Choi, Y., A. P. Chan, E. Kirkness, A. Telenti and N. J. Schork (2018). "Comparison of phasing strategies for whole human genomes." PLoS Genet **14**(4): e1007308.

Coene, K. L. M., L. A. J. Kluijtmans, E. van der Heeft, U. F. H. Engelke, S. de Boer, B. Hoegen, H. J. T. Kwast, M. van de Vorst, M. Huigen, I. Keularts, M. F. Schreuder, C. D. M. van Karnebeek, S. B. Wortmann, M. C. de Vries, M. C. H. Janssen, C. Gilissen, J. Engel and R. A. Wevers (2018). "Next-generation metabolic screening: targeted and untargeted metabolomics for the diagnosis of inborn errors of metabolism in individual patients." J Inherit Metab Dis **41**(3): 337-353.

Collins, R. L., H. Brand, K. J. Karczewski, X. Zhao, J. Alfoldi, L. C. Francioli, A. V. Khera, C. Lowther, L. D. Gauthier, H. Wang, N. A. Watts, M. Solomonson, A. O'Donnell-Luria, A. Baumann, R. Munshi, M. Walker, C. W. Whelan, Y. Huang, T. Brookings, T. Sharpe, M. R. Stone, E. Valkanas, J. Fu, G. Tiao, K. M. Laricchia, V. Ruano-Rubio, C. Stevens, N. Gupta, C. Cusick, L. Margolin, T. Genome Aggregation Database Production, C. Genome Aggregation Database, K. D. Taylor, H. J. Lin, S. S. Rich, W. S. Post, Y. I. Chen, J. I. Rotter, C. Nusbaum, A. Philippakis, E. Lander, S. Gabriel, B. M. Neale, S. Kathiresan, M. J. Daly, E. Banks, D. G. MacArthur and M. E. Talkowski (2020). "A structural variation reference for medical and population genetics." Nature **581**(7809): 444-451.

Coppola, L., A. Cianflone, A. M. Grimaldi, M. Incoronato, P. Bevilacqua, F. Messina, S. Baselice, A. Soricelli, P. Mirabelli and M. Salvatore (2019). "Biobanking in health care: evolution and future directions." J Transl Med **17**(1): 172.

Cui, H., F. Li, D. Chen, G. Wang, C. K. Truong, G. M. Enns, B. Graham, M. Milone, M. L. Landsverk, J. Wang, W. Zhang and L. J. Wong (2013). "Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders." Genet Med **15**(5): 388-394.

Cummings, B. B., J. L. Marshall, T. Tukiainen, M. Lek, S. Donkervoort, A. R. Foley, V. Bolduc, L. B. Waddell, S. A. Sandaradura, G. L. O'Grady, E. Estrella, H. M. Reddy, F. Zhao, B. Weisburd, K. J. Karczewski, A. H. O'Donnell-Luria, D. Birnbaum, A. Sarkozy, Y. Hu, H. Gonorazky, K. Claeys, H. Joshi, A. Bournazos, E. C. Oates, R. Ghaoui, M. R. Davis, N. G.

Laing, A. Topf, C. Genotype-Tissue Expression, P. B. Kang, A. H. Beggs, K. N. North, V. Straub, J. J. Dowling, F. Muntoni, N. F. Clarke, S. T. Cooper, C. G. Bonnemann and D. G. MacArthur (2017). "Improving genetic diagnosis in Mendelian disease with transcriptome sequencing." Sci Transl Med **9**(386).

Czell, D., A. Abicht, J. Hench and M. Weber (2012). "Exercise-induced myalgia and rhabdomyolysis in a patient with the rare m.3243A>T mtDNA mutation." BMJ Case Rep **2012**.

Damas, J., J. Carneiro, A. Amorim and F. Pereira (2014). "MitoBreak: the mitochondrial DNA breakpoints database." Nucleic Acids Res **42**(Database issue): D1261-1268.

De Biase, I., S. Tortorelli, L. Kratz, J. S. S, K. Cusmano-Ozog, N. Braverman and A. L. Q. A. Committee (2020). "Laboratory diagnosis of disorders of peroxisomal biogenesis and function: a technical standard of the American College of Medical Genetics and Genomics (ACMG)." Genet Med **22**(4): 686-697.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler and M. J. Daly (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." Nat Genet **43**(5): 491-498.

Desmet, F. O., D. Hamroun, M. Lalande, G. Collod-Beroud, M. Claustres and C. Beroud (2009). "Human Splicing Finder: an online bioinformatics tool to predict splicing signals." Nucleic Acids Res **37**(9): e67.

Dhir, A., S. Dhir, L. S. Borowski, L. Jimenez, M. Teitell, A. Rotig, Y. J. Crow, G. I. Rice, D. Duffy, C. Tamby, T. Nojima, A. Munnich, M. Schiff, C. R. de Almeida, J. Rehwinkel, A. Dziembowski, R. J. Szczesny and N. J. Proudfoot (2018). "Mitochondrial double-stranded RNA triggers antiviral signalling in humans." Nature **560**(7717): 238-242.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner." Bioinformatics **29**(1): 15-21.

Ebberink, M. S., J. Kofster, R. J. A. Wanders and H. R. Waterham (2010). "Spectrum ofPEX6mutations in Zellweger syndrome spectrum patients." Human Mutation **31**(1): E1058-E1070.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Res **32**(5): 1792-1797.

Eichler, E. E. (2019). "Genetic Variation, Comparative Genomics, and the Diagnosis of Disease." N Engl J Med **381**(1): 64-74.

El-Hattab, A. W., A. M. Adesina, J. Jones and F. Scaglia (2015). "MELAS syndrome: Clinical manifestations, pathogenesis, and treatment options." Mol Genet Metab **116**(1-2): 4-12.

Farruggia, P., A. Di Cataldo, R. M. Pinto, E. Palmisani, A. Macaluso, L. L. Valvo, M. E. Cantarini, A. Tornesello, P. Corti, F. Fioredda, S. Varotto, B. Martire, I. Moroni, G. Puccio, G. Russo, C. Dufour and M. Pillon (2016). "Pearson Syndrome: A Retrospective Cohort Study from the Marrow Failure Study Group of A.I.E.O.P. (Associazione Italiana Emato-Oncologia Pediatrica)." JIMD Rep **26**: 37-43.

Formosa, L. E., L. Muellner-Wong, B. Reljic, A. J. Sharpe, T. D. Jackson, T. H. Beilharz, D. Stojanovski, M. Lazarou, D. A. Stroud and M. T. Ryan (2020). "Dissecting the Roles of Mitochondrial Complex I Intermediate Assembly Complex Factors in the Biogenesis of Complex I." Cell Rep **31**(3): 107541.

Frazier, A. E., D. R. Thorburn and A. G. Compton (2019). "Mitochondrial energy generation disorders: genes, mechanisms, and clues to pathology." J Biol Chem **294**(14): 5386-5395.

Frazier, A. E., A. E. Vincent, D. M. Turnbull, D. R. Thorburn and R. W. Taylor (2020). "Assessment of mitochondrial respiratory chain enzymes in cells and tissues." Methods Cell Biol **155**: 121-156.

Fukasawa, Y., J. Tsuji, S. C. Fu, K. Tomii, P. Horton and K. Imai (2015). "MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites." Mol Cell Proteomics **14**(4): 1113-1126.

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

Gayevskiy, V., T. Roscioli, M. E. Dinger and M. J. Cowley (2019). "Seave: a comprehensive web platform for storing and interrogating human genomic variation." Bioinformatics **35**(1): 122-125.

Gorman, G. S., P. F. Chinnery, S. DiMauro, M. Hirano, Y. Koga, R. McFarland, A. Suomalainen, D. R. Thorburn, M. Zeviani and D. M. Turnbull (2016). "Mitochondrial diseases." Nat Rev Dis Primers **2**: 16080.

Grady, J. P., J. L. Murphy, E. L. Blakely, R. G. Haller, R. W. Taylor, D. M. Turnbull and H. A. Tuppen (2014). "Accurate measurement of mitochondrial DNA deletion level and copy number differences in human skeletal muscle." PLoS One **9**(12): e114462.

Grady, J. P., S. J. Pickett, Y. S. Ng, C. L. Alston, E. L. Blakely, S. A. Hardy, C. L. Feeney, A. A. Bright, A. M. Schaefer, G. S. Gorman, R. J. McNally, R. W. Taylor, D. M. Turnbull and R. McFarland (2018). "mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease." EMBO Mol Med **10**(6).

Gray, M. W. (2014). "The pre-endosymbiont hypothesis: a new perspective on the origin and evolution of mitochondria." Cold Spring Harb Perspect Biol **6**(3).

Grier, J., M. Hirano, A. Karaa, E. Shepard and J. L. P. Thompson (2018). "Diagnostic odyssey of patients with mitochondrial disease: Results of a survey." Neurol Genet **4**(2): e230.

Haack, T. B., B. Haberberger, E. M. Frisch, T. Wieland, A. Iuso, M. Gorza, V. Strecker, E. Graf, J. A. Mayr, U. Herberg, J. B. Hennermann, T. Klopstock, K. A. Kuhn, U. Ahting, W. Sperl, E. Wilichowski, G. F. Hoffmann, M. Tesarova, H. Hansikova, J. Zeman, B. Plecko, M. Zeviani, I. Wittig, T. M. Strom, M. Schuelke, P. Freisinger, T. Meitinger and H. Prokisch (2012). "Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing." J Med Genet **49**(4): 277-283.

Hecht, N. B., H. Liem, K. C. Kleene, R. J. Distel and S.-m. Ho (1984). "Maternal inheritance of the mouse mitochondrial genome is not mediated by a loss or gross alteration of the paternal

mitochondrial DNA or by methylation of the oocyte mitochondrial DNA." Developmental Biology **102**(2): 452-461.

Heimer, G., J. M. Keratar, L. G. Riley, S. Balasubramaniam, E. Eyal, L. P. Pietikainen, J. K. Hiltunen, D. Marek-Yagel, J. Hamada, A. Gregory, C. Rogers, P. Hogarth, M. A. Nance, N. Shalva, A. Veber, M. Tzadok, A. Nissenkorn, D. Tonduti, F. Renaldo, G. University of Washington Center for Mendelian, I. Kraoua, C. Panteghini, L. Valletta, B. Garavaglia, M. J. Cowley, V. Gayevskiy, T. Roscioli, J. M. Silberstein, C. Hoffmann, A. Raas-Rothschild, V. Tiranti, Y. Anikster, J. Christodoulou, A. J. Kastaniotis, B. Ben-Zeev and S. J. Hayflick (2016). "MECR Mutations Cause Childhood-Onset Dystonia and Optic Atrophy, a Mitochondrial Fatty Acid Synthesis Disorder." Am J Hum Genet **99**(6): 1229-1244.

Hikmat, O., C. Tzoulis, W. K. Chong, L. Chentouf, C. Klingenberg, C. Fratter, L. J. Carr, P. Prabhakar, N. Kumaraguru, P. Gissen, J. H. Cross, T. S. Jacques, J. W. Taanman, L. A. Bindoff and S. Rahman (2017). "The clinical spectrum and natural history of early-onset diseases due to DNA polymerase gamma mutations." Genet Med **19**(11): 1217-1225.

Hillen, H. S., D. Temiakov and P. Cramer (2018). "Structural basis of mitochondrial transcription." Nat Struct Mol Biol **25**(9): 754-765.

Hillman, R. T., R. E. Green and S. E. Brenner (2004). "An unappreciated role for RNA surveillance." Genome Biol **5**(2): R8.

Hock, D. H., B. Reljic, C. S. Ang, L. Muellner-Wong, H. S. Mountford, A. G. Compton, M. T. Ryan, D. R. Thorburn and D. A. Stroud (2020). "HIGD2A is Required for Assembly of the COX3 Module of Human Mitochondrial Complex IV." Mol Cell Proteomics **19**(7): 1145-1160.

Huttenlocher, P. R., G. B. Solitare and G. Adams (1976). "Infantile diffuse cerebral degeneration with hepatic cirrhosis." Arch Neurol **33**(3): 186-192.

Ikeda, T., H. Osaka, H. Shimbo, M. Tajika, M. Yamazaki, A. Ueda, K. Murayama and T. Yamagata (2018). "Mitochondrial DNA 3243A>T mutation in a patient with MELAS syndrome." Hum Genome Var **5**: 25.

Jaganathan, K., S. Kyriazopoulou Panagiotopoulou, J. F. McRae, S. F. Darbandi, D. Knowles, Y. I. Li, J. A. Kosmicki, J. Arbelaez, W. Cui, G. B. Schwartz, E. D. Chow, E. Kanterakis, H.

Gao, A. Kia, S. Batzoglou, S. J. Sanders and K. K. Farh (2019). "Predicting Splicing from Primary Sequence with Deep Learning." Cell **176**(3): 535-548 e524.

Jarvik, G. P. and J. P. Evans (2017). "Mastering genomic terminology." Genet Med **19**(5): 491-492.

Jiang, S., C. Koolmeister, J. Misic, S. Siira, I. Kuhl, E. Silva Ramos, M. Miranda, M. Jiang, V. Posse, O. Lytovchenko, I. Atanassov, F. A. Schober, R. Wibom, K. Hultenby, D. Milenkovic, C. M. Gustafsson, A. Filipovska and N. G. Larsson (2019). "TEFM regulates both transcription elongation and RNA processing in mitochondria." EMBO Rep **20**(6): e48101.

Kearns, T. P. and G. P. Sayre (1958). "Retinitis Pigmentosa, External Ophthalmoplegia, and Complete Heart Block." A.M.A. Archives of Ophthalmology **60**(2): 280-289.

Kohda, M., Y. Tokuzawa, Y. Kishita, H. Nyuzuki, Y. Moriyama, Y. Mizuno, T. Hirata, Y. Yatsuka, Y. Yamashita-Sugahara, Y. Nakachi, H. Kato, A. Okuda, S. Tamaru, N. N. Borna, K. Banshoya, T. Aigaki, Y. Sato-Miyata, K. Ohnuma, T. Suzuki, A. Nagao, H. Maehata, F. Matsuda, K. Higasa, M. Nagasaki, J. Yasuda, M. Yamamoto, T. Fushimi, M. Shimura, K. Kaiho-Ichimoto, H. Harashima, T. Yamazaki, M. Mori, K. Murayama, A. Ohtake and Y. Okazaki (2016). "A Comprehensive Genomic Analysis Reveals the Genetic Landscape of Mitochondrial Respiratory Chain Complex Deficiencies." PLoS Genet **12**(1): e1005679.

Kremer, L. S., D. M. Bader, C. Mertes, R. Kopajtich, G. Pichler, A. Iuso, T. B. Haack, E. Graf, T. Schwarzmayr, C. Terrile, E. Konarikova, B. Repp, G. Kastenmuller, J. Adamski, P. Lichtner, C. Leonhardt, B. Funalot, A. Donati, V. Tiranti, A. Lombes, C. Jardel, D. Glaser, R. W. Taylor, D. Ghezzi, J. A. Mayr, A. Rotig, P. Freisinger, F. Distelmaier, T. M. Strom, T. Meitinger, J. Gagneur and H. Prokisch (2017). "Genetic diagnosis of Mendelian disorders via RNA sequencing." Nat Commun **8**: 15824.

Kremer, L. S., S. B. Wortmann and H. Prokisch (2018). ""Transcriptomics": molecular diagnosis of inborn errors of metabolism via RNA-sequencing." J Inherit Metab Dis **41**(3): 525-532.

Kuhlbrandt, W. (2015). "Structure and function of mitochondrial membrane protein complexes." BMC Biol **13**: 89.

Kunze, M. and J. Berger (2015). "The similarity between N-terminal targeting signals for protein import into different organelles and its evolutionary relevance." Front Physiol **6**: 259.

Lake, N. J., A. G. Compton, S. Rahman and D. R. Thorburn (2016). "Leigh syndrome: One disorder, more than 75 monogenic causes." Ann Neurol **79**(2): 190-203.

Lake, N. J., L. E. Formosa, D. A. Stroud, M. T. Ryan, S. E. Calvo, V. K. Mootha, B. Morar, P. G. Procopis, J. Christodoulou, A. G. Compton and D. R. Thorburn (2019). "A patient with homozygous nonsense variants in two Leigh syndrome disease genes: Distinguishing a dual diagnosis from a hypomorphic protein-truncating variant." Hum Mutat **40**(7): 893-898.

Lake, N. J., B. D. Webb, D. A. Stroud, T. R. Richman, B. Ruzzenente, A. G. Compton, H. S. Mountford, J. Pulman, C. Zangarelli, M. Rio, N. Bodaert, Z. Assouline, M. D. Sherpa, E. E. Schadt, S. M. Houten, J. Byrnes, E. M. McCormick, Z. Zolkipli-Cunningham, K. Haude, Z. Zhang, K. Retterer, R. Bai, S. E. Calvo, V. K. Mootha, J. Christodoulou, A. Rotig, A. Filipovska, I. Cristian, M. J. Falk, M. D. Metodiev and D. R. Thorburn (2017). "Biallelic Mutations in MRPS34 Lead to Instability of the Small Mitoribosomal Subunit and Leigh Syndrome." Am J Hum Genet **101**(2): 239-254.

Lamande, S. R., J. F. Bateman, W. Hutchison, R. J. McKinlay Gardner, S. P. Bower, E. Byrne and H. H. Dahl (1998). "Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency." Hum Mol Genet **7**(6): 981-989.

Latorre-Pellicer, A., A. V. Lechuga-Vieco, I. G. Johnston, R. H. Hamalainen, J. Pellico, R. Justo-Mendez, J. M. Fernandez-Toro, C. Claveria, A. Guaras, R. Sierra, J. Llop, M. Torres, L. M. Criado, A. Suomalainen, N. S. Jones, J. Ruiz-Cabello and J. A. Enriquez (2019). "Regulation of Mother-to-Offspring Transmission of mtDNA Heteroplasmy." Cell Metab **30**(6): 1120-1130 e1125.

Layer, R. M., C. Chiang, A. R. Quinlan and I. M. Hall (2014). "LUMPY: a probabilistic framework for structural variant discovery." Genome Biol **15**(6): R84.

Lee, H. F., H. J. Lee, C. S. Chi, C. R. Tsai, T. K. Chang and C. J. Wang (2007). "The neurological evolution of Pearson syndrome: case report and literature review." Eur J Paediatr Neurol **11**(4): 208-214.

Legati, A., A. Reyes, A. Nasca, F. Invernizzi, E. Lamantea, V. Tiranti, B. Garavaglia, C. Lamperti, A. Ardissone, I. Moroni, A. Robinson, D. Ghezzi and M. Zeviani (2016). "New genes and pathomechanisms in mitochondrial disorders unraveled by NGS technologies." Biochim Biophys Acta **1857**(8): 1326-1335.

Lehtonen, J. M., M. Auranen, N. Darin, K. Sofou, L. Bindoff, O. Hikmat, J. Uusimaa, P. Vieira, M. Tulinius, T. Lonnqvist, I. F. de Coo, A. Suomalainen and P. Isohanni (2020). "Diagnostic value of serum biomarkers FGF21 and GDF15 compared to muscle sample in mitochondrial disease." J Inherit Metab Dis.

Li, H. and R. Durbin (2010). "Fast and accurate long-read alignment with Burrows-Wheeler transform." Bioinformatics **26**(5): 589-595.

Liao, Y., G. K. Smyth and W. Shi (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.

Logsdon, G. A., M. R. Vollger and E. E. Eichler (2020). "Long-read human genome sequencing and its applications." Nat Rev Genet **21**(10): 597-614.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.

Luo, S., C. A. Valencia, J. Zhang, N. C. Lee, J. Slone, B. Gui, X. Wang, Z. Li, S. Dell, J. Brown, S. M. Chen, Y. H. Chien, W. L. Hwu, P. C. Fan, L. J. Wong, P. S. Atwal and T. Huang (2018). "Biparental Inheritance of Mitochondrial DNA in Humans." Proc Natl Acad Sci U S A **115**(51): 13039-13044.

Machiela, M. J. and S. J. Chanock (2015). "LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants." Bioinformatics **31**(21): 3555-3557.

Manichaikul, A., J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale and W. M. Chen (2010). "Robust relationship inference in genome-wide association studies." Bioinformatics **26**(22): 2867-2873.

Matilainen, S., C. J. Carroll, U. Richter, L. Euro, M. Pohjanpelto, A. Paetau, P. Isohanni and A. Suomalainen (2017). "Defective mitochondrial RNA processing due to PNPT1 variants causes Leigh syndrome." Hum Mol Genet **26**(17): 3352-3361.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek and F. Cunningham (2016). "The Ensembl Variant Effect Predictor." Genome Biol **17**(1): 122.

Mendelsohn, B. A., N. K. Bennett, M. A. Darch, K. Yu, M. K. Nguyen, D. Pucciarelli, M. Nelson, M. A. Horlbeck, L. A. Gilbert, W. Hyun, M. Kampmann, J. L. Nakamura and K. Nakamura (2018). "A high-throughput screen of real-time ATP levels in individual cells reveals mechanisms of energy failure." PLoS Biol **16**(8): e2004624.

Menezes, M. J., L. G. Riley and J. Christodoulou (2014). "Mitochondrial respiratory chain disorders in childhood: insights into diagnosis and management in the new era of genomic medicine." Biochim Biophys Acta **1840**(4): 1368-1379.

Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nat Rev Genet **11**(1): 31-46.

Meuwissen, M. E., R. Schot, S. Buta, G. Oudesluijs, S. Tinschert, S. D. Speer, Z. Li, L. van Unen, D. Heijsman, T. Goldmann, M. H. Lequin, J. M. Kros, W. Stam, M. Hermann, R. Willemsen, R. W. Brouwer, I. W. F. Van, M. Martin-Fernandez, I. de Coo, J. Dudink, F. A. de Vries, A. Bertoli Avella, M. Prinz, Y. J. Crow, F. W. Verheijen, S. Pellegrini, D. Bogunovic and G. M. Mancini (2016). "Human USP18 deficiency underlies type 1 interferonopathy leading to severe pseudo-TORCH syndrome." J Exp Med **213**(7): 1163-1174.

Moraes, C. T., S. DiMauro, M. Zeviani, A. Lombes, S. Shanske, A. F. Miranda, H. Nakase, E. Bonilla, L. C. Werneck, S. Servidei, I. Nonaka, Y. Koga, A. J. Spiro, K. Brownell, B. Schmidt, D. Schotland, M. Zupanc, D. C. DeVivo, E. A. Schon and L. P. Rowland (1989). "Mitochondrial DNA deletions in progressive external ophthalmoplegia and Kearns-Sayre syndrome." N Engl J Med **320**(20): 1293-1299.

Morava, E., L. van den Heuvel, F. Hol, M. C. de Vries, M. Hogeveen, R. J. Rodenburg and J. A. Smeitink (2006). "Mitochondrial disease criteria: diagnostic applications in children." Neurology **67**(10): 1823-1826.

Neveling, K., I. Feenstra, C. Gilissen, L. H. Hoefsloot, E. J. Kamsteeg, A. R. Mensenkamp, R. J. Rodenburg, H. G. Yntema, L. Spruijt, S. Vermeer, T. Rinne, K. L. van Gassen, D. Bodmer, D. Lugtenberg, R. de Reuver, W. Buijsman, R. C. Derks, N. Wieskamp, B. van den Heuvel, M. J. Ligtenberg, H. Kremer, D. A. Koolen, B. P. van de Warrenburg, F. P. Cremers, C. L. Marcelis, J. A. Smeitink, S. B. Wortmann, W. A. van Zelst-Stams, J. A. Veltman, H. G. Brunner, H. Scheffer and M. R. Nelen (2013). "A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases." Hum Mutat **34**(12): 1721-1726.

Ohtake, A., K. Murayama, M. Mori, H. Harashima, T. Yamazaki, S. Tamaru, Y. Yamashita, Y. Kishita, Y. Nakachi, M. Kohda, Y. Tokuzawa, Y. Mizuno, Y. Moriyama, H. Kato and Y. Okazaki (2014). "Diagnosis and molecular basis of mitochondrial respiratory chain disorders: exome sequencing for disease gene identification." Biochim Biophys Acta **1840**(4): 1355-1359.

Paila, U., B. A. Chapman, R. Kirchner and A. R. Quinlan (2013). "GEMINI: integrative exploration of genetic variation and genome annotations." PLoS Comput Biol **9**(7): e1003153.

Parikh, S., A. Goldstein, M. K. Koenig, F. Scaglia, G. M. Enns, R. Saneto, I. Anselm, B. H. Cohen, M. J. Falk, C. Greene, A. L. Gropman, R. Haas, M. Hirano, P. Morgan, K. Sims, M. Tarnopolsky, J. L. Van Hove, L. Wolfe and S. DiMauro (2015). "Diagnosis and management of mitochondrial disease: a consensus statement from the Mitochondrial Medicine Society." Genet Med **17**(9): 689-701.

Parikh, S., A. Karaa, A. Goldstein, E. S. Bertini, P. F. Chinnery, J. Christodoulou, B. H. Cohen, R. L. Davis, M. J. Falk, C. Fratter, R. Horvath, M. K. Koenig, M. Mancuso, S. McCormack, E. M. McCormick, R. McFarland, V. Nesbitt, M. Schiff, H. Steele, S. Stockler, C. Sue, M. Tarnopolsky, D. R. Thorburn, J. Vockley and S. Rahman (2019). "Diagnosis of 'possible' mitochondrial disease: an existential crisis." J Med Genet **56**(3): 123-130.

Philippakis, A. A., D. R. Azzariti, S. Beltran, A. J. Brookes, C. A. Brownstein, M. Brudno, H. G. Brunner, O. J. Buske, K. Carey, C. Doll, S. Dumitriu, S. O. Dyke, J. T. den Dunnen, H. V. Firth, R. A. Gibbs, M. Girdea, M. Gonzalez, M. A. Haendel, A. Hamosh, I. A. Holm, L. Huang, M. E. Hurles, B. Hutton, J. B. Krier, A. Misyura, C. J. Mungall, J. Paschall, B. Paten, P. N. Robinson, F. Schiettecatte, N. L. Sobreira, G. J. Swaminathan, P. E. Taschner, S. F. Terry, N. L. Washington, S. Zuchner, K. M. Boycott and H. L. Rehm (2015). "The Matchmaker Exchange: a platform for rare disease gene discovery." Hum Mutat **36**(10): 915-921.

Pinese, M., P. Lacaze, E. M. Rath, A. Stone, M. J. Brion, A. Ameur, S. Nagpal, C. Puttick, S. Husson, D. Degrave, T. N. Cristina, V. F. S. Kahl, A. L. Statham, R. L. Woods, J. J. McNeil, M. Riaz, M. Barr, M. R. Nelson, C. M. Reid, A. M. Murray, R. C. Shah, R. Wolfe, J. R. Atkins, C. Fitzsimmons, H. M. Cairns, M. J. Green, V. J. Carr, M. J. Cowley, H. A. Pickett, P. A. James, J. E. Powell, W. Kaplan, G. Gibson, U. Gyllensten, M. J. Cairns, M. McNamara, M. E. Dinger and D. M. Thomas (2020). "The Medical Genome Reference Bank contains whole genome and phenotype data of 2570 healthy elderly." Nat Commun **11**(1): 435.

Priesnitz, C., N. Pfanner and T. Becker (2020). "Studying protein import into mitochondria." Methods Cell Biol **155**: 45-79.

Pronicka, E., D. Piekutowska-Abramczuk, E. Ciara, J. Trubicka, D. Rokicki, A. Karkucinska-Wieckowska, M. Pajdowska, E. Jurkiewicz, P. Halat, J. Kosinska, A. Pollak, M. Rydzanicz, P. Stawinski, M. Pronicki, M. Krajewska-Walasek and R. Ploski (2016). "New perspective in diagnostics of mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric centre." J Transl Med **14**(1): 174.

Puttick, C., K. R. Kumar, R. L. Davis, M. Pinese, D. M. Thomas, M. E. Dinger, C. M. Sue and M. J. Cowley (2019). "mity: A highly sensitive mitochondrial variant analysis pipeline for whole genome sequencing data." bioRxiv.

Puusepp, S., K. Reinson, S. Pajusalu, U. Murumets, E. Oiglane-Shlik, R. Rein, I. Talvik, R. J. Rodenburg and K. Ounap (2018). "Effectiveness of whole exome sequencing in unsolved patients with a clinical suspicion of a mitochondrial disorder in Estonia." Mol Genet Metab Rep **15**: 80-89.

Rahman, J., A. Noronha, I. Thiele and S. Rahman (2017). "Leigh map: A novel computational diagnostic resource for mitochondrial disease." Ann Neurol **81**(1): 9-16.

Rahman, S., J. Poulton, D. Marchington and A. Suomalainen (2001). "Decrease of 3243 A-->G mtDNA mutation from blood in MELAS syndrome: a longitudinal study." Am J Hum Genet **68**(1): 238-240.

Richards, S., N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding, H. L. Rehm and A. L. Q. A. Committee (2015). "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of

the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." Genet Med **17**(5): 405-424.

Riggs, E. R., E. F. Andersen, A. M. Cherry, S. Kantarci, H. Kearney, A. Patel, G. Raca, D. I. Ritter, S. T. South, E. C. Thorland, D. Pineda-Alvarez, S. Aradhya, C. L. Martin and Acmg (2020). "Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen)." Genet Med **22**(2): 245-257.

Riley, L. G., M. J. Cowley, V. Gayevskiy, A. E. Minoche, C. Puttick, D. R. Thorburn, R. Rius, A. G. Compton, M. J. Menezes, K. Bhattacharya, D. Coman, C. Ellaway, I. E. Alexander, L. Adams, M. Kava, J. Robinson, C. M. Sue, S. Balasubramaniam and J. Christodoulou (2020). "The diagnostic utility of genome sequencing in a pediatric cohort with suspected mitochondrial disease." Genet Med **22**(7): 1254-1261.

Rius, R., M. J. Cowley, L. Riley, C. Puttick, D. R. Thorburn and J. Christodoulou (2019). "Biparental inheritance of mitochondrial DNA in humans is not a common phenomenon." Genet Med **21**(12): 2823-2826.

Rius, R., L. G. Riley, Y. Guo, M. Menezes, A. G. Compton, N. J. Van Bergen, V. Gayevskiy, M. J. Cowley, B. B. Cummings, L. Adams, C. Ellaway, D. R. Thorburn, H. Hakonarson and J. Christodoulou (2019). "Cryptic intronic NBAS variant reveals the genetic basis of recurrent liver failure in a child." Mol Genet Metab **126**(1): 77-82.

Rius, R., N. J. Van Bergen, A. G. Compton, L. G. Riley, M. P. Kava, S. Balasubramaniam, D. J. Amor, M. Fanjul-Fernandez, M. J. Cowley, M. C. Fahey, M. K. Koenig, G. M. Enns, S. Sadedin, M. J. Wilson, T. Y. Tan, D. R. Thorburn and J. Christodoulou (2019). "Clinical Spectrum and Functional Consequences Associated with Bi-Allelic Pathogenic PNPT1 Variants." J Clin Med **8**(11).

Sadedin, S. P., H. Dashnow, P. A. James, M. Bahlo, D. C. Bauer, A. Lonie, S. Lunke, I. Macciocca, J. P. Ross, K. R. Siemering, Z. Stark, S. M. White, A. Melbourne Genomics Health, G. Taylor, C. Gaff, A. Oshlack and N. P. Thorne (2015). "Cpipe: a shared variant detection pipeline designed for diagnostic settings." Genome Med **7**(1): 68.

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

Sadikovic, B., J. Wang, A. W. El-Hattab, M. Landsverk, G. Douglas, E. K. Brundage, W. J. Craigen, E. S. Schmitt and L. J. Wong (2010). "Sequence homology at the breakpoint and clinical phenotype of mitochondrial DNA deletion syndromes." PLoS One **5**(12): e15687.

Sagan, L. (1967). "On the origin of mitosing cells." J Theor Biol **14**(3): 255-274.

Savojardo, C., P. L. Martelli, P. Fariselli and R. Casadio (2015). "TPpred3 detects and discriminates mitochondrial and chloroplastic targeting peptides in eukaryotic proteins." Bioinformatics **31**(20): 3269-3275.

Schaefer, A. M., R. McFarland, E. L. Blakely, L. He, R. G. Whittaker, R. W. Taylor, P. F. Chinnery and D. M. Turnbull (2008). "Prevalence of mitochondrial DNA disease in adults." Ann Neurol **63**(1): 35-39.

Schmittgen, T. D. and K. J. Livak (2008). "Analyzing real-time PCR data by the comparative C(T) method." Nat Protoc **3**(6): 1101-1108.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Methods **9**(7): 671-675.

Shen, L., M. Attimonelli, R. Bai, M. T. Lott, D. C. Wallace, M. J. Falk and X. Gai (2018). "MSeqDR mvTool: A mitochondrial DNA Web and API resource for comprehensive variant annotation, universal nomenclature collation, and reference genome conversion." Hum Mutat **39**(6): 806-810.

Sironi, M., G. Menozzi, L. Riva, R. Cagliani, G. P. Comi, N. Bresolin, R. Giorda and U. Pozzoli (2004). "Silencer elements as possible inhibitors of pseudoexon splicing." Nucleic Acids Res **32**(5): 1783-1791.

Skladal, D., J. Halliday and D. R. Thorburn (2003). "Minimum birth prevalence of mitochondrial respiratory chain disorders in children." Brain **126**(Pt 8): 1905-1912.

Smedley, D., M. Schubach, J. O. B. Jacobsen, S. Kohler, T. Zemojtel, M. Spielmann, M. Jager, H. Hochheiser, N. L. Washington, J. A. McMurry, M. A. Haendel, C. J. Mungall, S. E. Lewis, T. Groza, G. Valentini and P. N. Robinson (2016). "A Whole-Genome Analysis Framework for

Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease." Am J Hum Genet **99**(3): 595-606.

Smith, H. S., J. M. Swint, S. R. Lalani, J. M. Yamal, M. C. de Oliveira Otto, S. Castellanos, A. Taylor, B. H. Lee and H. V. Russell (2019). "Clinical Application of Genome and Exome Sequencing as a Diagnostic Tool for Pediatric Patients: a Scoping Review of the Literature." Genet Med **21**(1): 3-16.

Sohm, B., M. Frugier, H. Brule, K. Olszak, A. Przykorska and C. Florentz (2003). "Towards understanding human mitochondrial leucine aminoacylation identity." J Mol Biol **328**(5): 995-1010.

Sommerville, E. W., X. L. Zhou, M. Olahova, J. Jenkins, L. Euro, S. Konovalova, T. Hilander, A. Pyle, L. He, S. Habeebu, C. Saunders, A. Kelsey, A. A. M. Morris, R. McFarland, A. Suomalainen, G. S. Gorman, E. D. Wang, I. Thiffault, H. Tyynismaa and R. W. Taylor (2019). "Instability of the mitochondrial alanyl-tRNA synthetase underlies fatal infantile-onset cardiomyopathy." Hum Mol Genet **28**(2): 258-268.

Srivastava, S., A. Butala, S. Mahida, J. Richter, W. Mu, A. Poretti, H. Vernon, J. VanGerpen, P. S. Atwal, E. H. Middlebrooks, D. S. Zee and S. Naidu (2019). "Expansion of the clinical spectrum associated with AARS2-related disorders." Am J Med Genet A **179**(8): 1556-1564.

Stark, Z., T. Boughtwood, P. Phillips, J. Christodoulou, D. P. Hansen, J. Braithwaite, A. J. Newson, C. L. Gaff, A. H. Sinclair and K. N. North (2019). "Australian Genomics: A Federated Model for Integrating Genomics into Healthcare." Am J Hum Genet **105**(1): 7-14.

Stenton, S. L. and H. Prokisch (2018). "Advancing genomic approaches to the molecular diagnosis of mitochondrial disease." Essays Biochem **62**(3): 399-408.

Stenton, S. L. and H. Prokisch (2020). "Genetics of mitochondrial diseases: Identifying mutations to help diagnosis." EBioMedicine **56**: 102784.

Stewart, J. B. and P. F. Chinnery (2015). "The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease." Nat Rev Genet **16**(9): 530-542.

Suomalainen, A. and B. J. Battersby (2018). "Mitochondrial diseases: the contribution of organelle stress responses to pathology." Nat Rev Mol Cell Biol **19**(2): 77-92.

Sutovsky, P., R. D. Moreno, J. Ramalho-Santos, T. Dominko, C. Simerly and G. Schatten (1999). "Ubiquitin tag for sperm mitochondria." Nature **402**(6760): 371-372.

Sutovsky, P., K. Van Leyen, T. McCauley, B. N. Day and M. Sutovsky (2004). "Degradation of paternal mitochondria after fertilization: implications for heteroplasmy, assisted reproductive technologies and mtDNA inheritance." Reprod Biomed Online **8**(1): 24-33.

Takakubo, F., P. Cartwright, N. Hoogenraad, D. R. Thorburn, F. Collins, T. Lithgow and H. H. Dahl (1995). "An amino acid substitution in the pyruvate dehydrogenase E1 alpha gene, affecting mitochondrial import of the precursor protein." Am J Hum Genet **57**(4): 772-780.

Tankard, R. M., M. F. Bennett, P. Degorski, M. B. Delatycki, P. J. Lockhart and M. Bahlo (2018). "Detecting Expansions of Tandem Repeats in Cohorts Sequenced with Short-Read Sequencing Data." Am J Hum Genet **103**(6): 858-873.

Taylor, R. W., A. Pyle, H. Griffin, E. L. Blakely, J. Duff, L. He, T. Smertenko, C. L. Alston, V. C. Neeve, A. Best, J. W. Yarham, J. Kirschner, U. Schara, B. Talim, H. Topaloglu, I. Baric, E. Holinski-Feder, A. Abicht, B. Czermin, S. Kleinle, A. A. Morris, G. Vassallo, G. S. Gorman, V. Ramesh, D. M. Turnbull, M. Santibanez-Koref, R. McFarland, R. Horvath and P. F. Chinnery (2014). "Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies." JAMA **312**(1): 68-77.

Taylor, R. W. and D. M. Turnbull (2005). "Mitochondrial DNA mutations in human disease." Nat Rev Genet **6**(5): 389-402.

Theunissen, T. E. J., M. Nguyen, R. Kamps, A. T. Hendrickx, S. Sallevelt, R. W. H. Gottschalk, C. M. Calis, A. P. M. Stassen, B. de Koning, E. N. M. Mulder-Den Hartog, K. Schoonderwoerd, S. A. Fuchs, Y. Hilhorst-Hofstee, M. de Visser, J. Vanoevelen, R. Szklarczyk, M. Gerards, I. F. M. de Coo, D. Hellebrekers and H. J. M. Smeets (2018). "Whole Exome Sequencing Is the Preferred Strategy to Identify the Genetic Defect in Patients With a Probable or Possible Mitochondrial Cause." Front Genet **9**: 400.

Topol, E. J. (2014). "Individualized medicine from prewomb to tomb." Cell **157**(1): 241-253.

Torraco, A., O. Stehling, C. Stumpfig, R. Rosser, D. De Rasmo, G. Fiermonte, D. Verrigni, T. Rizza, A. Vozza, M. Di Nottia, D. Diodato, D. Martinelli, F. Piemonte, C. Dionisi-Vici, E. Bertini, R. Lill and R. Carrozzo (2018). "ISCA1 mutation in a patient with infantile-onset

leukodystrophy causes defects in mitochondrial [4Fe-4S] proteins." Hum Mol Genet **27**(15): 2739-2754.

Tucker, E. J., M. Mimaki, A. G. Compton, M. McKenzie, M. T. Ryan and D. R. Thorburn (2012). "Next-generation sequencing in molecular diagnosis: NUBPL mutations highlight the challenges of variant detection and interpretation." Hum Mutat **33**(2): 411-418.

Tucker, E. J., R. Rius, S. Jaillard, K. Bell, P. J. Lamont, A. Travessa, J. Dupont, L. Sampaio, J. Dulon, S. Vuillaumier-Barrot, S. Whalen, A. Isapof, T. Stojkovic, S. Quijano-Roy, G. Robevska, J. van den Bergen, C. Hanna, A. Simpson, K. Ayers, D. R. Thorburn, J. Christodoulou, P. Touraine and A. H. Sinclair (2020). "Genomic sequencing highlights the diverse molecular causes of Perrault syndrome: a peroxisomal disorder (PEX6), metabolic disorders (CLPP, GGPS1), and mtDNA maintenance/translation disorders (LARS2, TFAM)." Human Genetics **139**(10): 1325-1343.

Wagner, M., R. Berutti, B. Lorenz-Depiereux, E. Graf, G. Eckstein, J. A. Mayr, T. Meitinger, U. Ahting, H. Prokisch, T. M. Strom and S. B. Wortmann (2019). "Mitochondrial DNA mutation analysis from exome sequencing-A more holistic approach in diagnostics of suspected mitochondrial disease." J Inherit Metab Dis **42**(5): 909-917.

Wanders, R. J. A., F. M. Vaz, S. Ferdinandusse, A. B. P. van Kuilenburg, S. Kemp, C. D. van Karnebeek, H. R. Waterham and R. H. Houtkooper (2019). "Translational Metabolism: A multidisciplinary approach towards precision diagnosis of inborn errors of metabolism in the omics era." J Inherit Metab Dis **42**(2): 197-208.

Wang, J., E. S. Schmitt, M. L. Landsverk, V. W. Zhang, F. Y. Li, B. H. Graham, W. J. Craigen and L. J. Wong (2012). "An integrated approach for classifying mitochondrial DNA variants: one clinical diagnostic laboratory's experience." Genet Med **14**(6): 620-626.

Wei, W., A. T. Pagnamenta, N. Gleadall, A. Sanchis-Juan, J. Stephens, J. Broxholme, S. Tuna, C. A. Odhams, C. Genomics England Research, N. BioResource, C. Fratter, E. Turro, M. J. Caulfield, J. C. Taylor, S. Rahman and P. F. Chinnery (2020). "Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans." Nat Commun **11**(1): 1740.

Wightman, P. J., R. Santer, A. Ribes, F. Dougherty, N. McGill, D. R. Thorburn and D. R. FitzPatrick (2003). "MLYCD mutation analysis: evidence for protein mistargeting as a cause of MLYCD deficiency." Hum Mutat **22**(4): 288-300.

Willemsen, M. H., A. T. Vulto-van Silfhout, W. M. Nillesen, W. M. Wissink-Lindhout, H. van Bokhoven, N. Philip, E. M. Berry-Kravis, U. Kini, C. M. van Ravenswaaij-Arts, B. Delle Chiaie, A. M. Innes, G. Houge, T. Kosonen, K. Cremer, M. Fannemel, A. Stray-Pedersen, W. Reardon, J. Ignatius, K. Lachlan, C. Mircher, P. T. Helderman van den Enden, M. Mastebroek, P. E. Cohn-Hokke, H. G. Yntema, S. Drunat and T. Kleefstra (2012). "Update on Kleefstra Syndrome." Mol Syndromol **2**(3-5): 202-212.

Wong, L. C., T. Chen, J. Wang, S. Tang, E. S. Schmitt, M. Landsverk, F. Li, Y. Wang, S. Zhang, V. W. Zhang and W. J. Craigen (2020). "Interpretation of mitochondrial tRNA variants." Genet Med **22**(5): 917-926.

Wortmann, S. B., M. Duran, Y. Anikster, P. G. Barth, W. Sperl, J. Zschocke, E. Morava and R. A. Wevers (2013). "Inborn errors of metabolism with 3-methylglutaconic aciduria as discriminative feature: proper classification and nomenclature." J Inherit Metab Dis **36**(6): 923-928.

Wortmann, S. B., D. A. Koolen, J. A. Smeitink, L. van den Heuvel and R. J. Rodenburg (2015). "Whole exome sequencing of suspected mitochondrial patients in clinical practice." J Inherit Metab Dis **38**(3): 437-443.

Wortmann, S. B., J. A. Mayr, J. M. Nuoffer, H. Prokisch and W. Sperl (2017). "A Guideline for the Diagnosis of Pediatric Mitochondrial Disease: The Value of Muscle and Skin Biopsies in the Genetics Era." Neuropediatrics **48**(4): 309-314.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden (2012). "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction." BMC Bioinformatics **13**: 134.

Yu-Wai-Man, P., P. G. Griffiths, G. Hudson and P. F. Chinnery (2009). "Inherited mitochondrial optic neuropathies." J Med Genet **46**(3): 145-158.

Zambelli, F., K. Vancampenhout, D. Daneels, D. Brown, J. Mertens, S. Van Dooren, B. Caljon, L. Gianaroli, K. Sermon, T. Voet, S. Seneca and C. Spits (2017). "Accurate and comprehensive

analysis of single nucleotide variants and large deletions of the human mitochondrial genome in DNA and single cells." Eur J Hum Genet **25**(11): 1229-1236.

Zhang, C., W. H. Li, A. R. Krainer and M. Q. Zhang (2008). "RNA landscape of evolution for optimal exon and intron discrimination." Proc Natl Acad Sci U S A **105**(15): 5797-5802.

Zhang, H., S. P. Burr and P. F. Chinnery (2018). "The mitochondrial DNA genetic bottleneck: inheritance and beyond." Essays Biochem **62**(3): 225-234.

Zong, S., M. Wu, J. Gu, T. Liu, R. Guo and M. Yang (2018). "Structure of the intact 14-subunit human cytochrome c oxidase." Cell Research **28**(10): 1026-1034.

Appendix A

A. Supplementary information for Chapter 3 Use of Trio GS in patients with suspected mitochondrial diseases

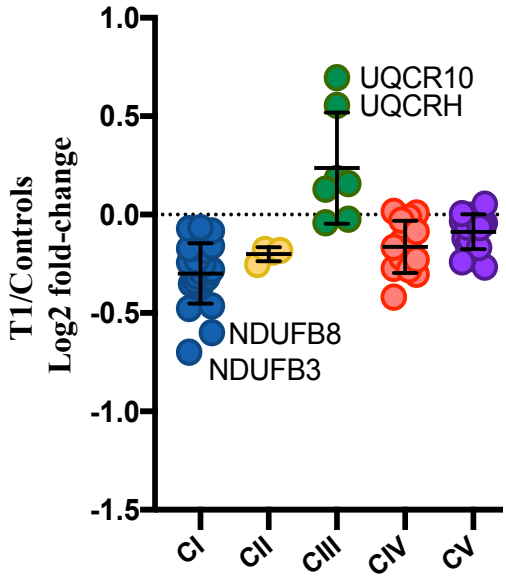

Figure T1. Quantitative Proteomic Analysis of Fibroblasts from T1 and 4 Controls.

Quantitative mass spectrometry performed by Daniella Hock (Bio 21, Melbourne) showing only OXPHOS proteins. Relative to controls (n=4) CI OXPHOS proteins levels are lower in fibroblasts from patient T1.

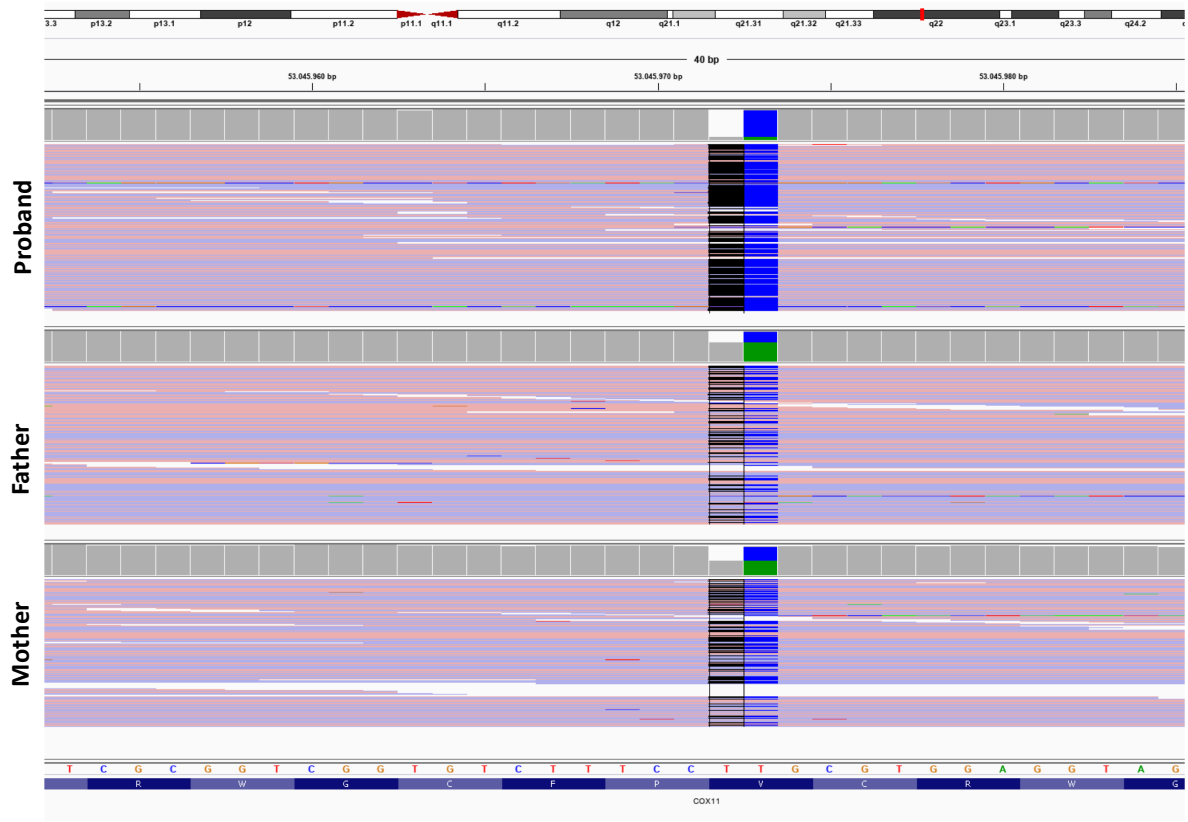

Figure X2 Trio ES detects a homozygous variant in *COX11*.

IGV screenshot depict the NM_004375.4(*COX11*)c.35_36delinsG in a homozygous state in the proband (X2), and in a heterozygous state in the parents.

Appendix B

B. Supplementary information for Chapter 4 Cryptic intronic *NBAS* variant reveals the genetic basis of recurrent liver failure in a child

Table S1. Primers used for *NBAS* cDNA analysis

Fragment	Exons	Primer F	Primer R
1	1-8	5' TAGAGACACTCCTCCGCTGC 3'	5' AGTTCTGCAGACCACTGTGC 3'
2	7-12	5' CACAGGAACTGTGAGGGTGT 3'	5' TGACCCCATTCCCCTTGTTG 3'
3	12-15	5' ACTGAGCATCTGGGCGATTC 3'	5' AGTAATGGTTCGTGGGCGTT 3'
4	15-19	5' ATTTGCACCACCACGGAAAC 3'	5' TCCGTCTACAACGGCAAAGT 3'
5	18-25	5' TCACCACCTGATGAAGAGCC 3'	5' GGAACCATCCACTGGTAGGC 3'
6	25-30	5' GCCTACCAGTGGATGGTTCC 3'	5' CTGGCAGCCAAAACCAAGTC 3'
7	30-34	5' GGAAACCCCACTACAGGGTC 3'	5' ATGAGCTCTTGACGAGTGGC 3'
8	34-40	5' GTTACCAGGACTTGGCCACT 3'	5' GCTGCCAGCTGGAGAGATAA 3'
9	40-44	5' CCAGCTGGCAGCGTATTACT 3'	5' ATAGGCATGAAGCCACTCCG 3'
10	44-47	5' CTTCACCGGAGTGGCTTCAT 3'	5' CTGCTGCAACAACACCTTCC 3'
11	46-52	5' GCCCTCTTGACATCTCACCC 3'	5' GGAGACACACTTCACCAGCA 3'
12	48-52	5' TTTCTGTGCTGATGACGCCT 3'	5' GCTAGCCAAGAGGTGGTCAA 3'

Table S2. Sequence alignment using MUSCLE (Edgar 2004) shows the evolutionary conservation of the NBAS p.Arg873 residue

H.sapiens	NP_056993.2	843	MDWYQTRAEEIEHYARQVDCALSLI**R**LGMERNIPGLLVLCDNLVTLETLV	892
P.troglodytes	XP_001161679.1	843	VDWYQTRAEEIEHYARQVDCALSLI**R**LGMERNIPGLLVLCDNLVTLETLV	892
M.mulatta	XP_002799196.1	843	MDWYQTRAEEIEHYALQVDCALSLI**R**LGMERNIPGLLVLCDNLVTLETLV	892
C.lupus	XP_005630201.1	843	MDWYQTRAEEIEHYARQVDCALSLI**R**LGMERNIPGLLALCDNLVTLEALV	892
B.taurus	NP_001179479.1	842	MDWYGSRAEEIEHYARQVDCALSLI**R**LGMERNIPGLLVLCDNLVTLEALV	891
M.musculus	NP_081982.1	842	MAWYQSRAEDIEHHAGQVDCALSLV**R**LGVERHIPGLLTLCDDLVTLETLV	891
G.gallus	XP_419959.3	847	TDWYLTRAQEIEKYAMQVDCALSLV**R**LGMERNIPGLQVLCDNLITLETVV	896
D.rerio	NP_001038272.1	832	TEWYLTRAQDIESHSRQVDCSLSLV**R**LGKEQNIPGLERLCDDLVTMETLV	881
X.tropicalis	XP_002941418.2	836	KSWYWNRAQEVENYARLVDNALSLV**R**LGIERNIPGLQILSDDLVTLETLV	885

Table S3. *In silico* analysis of the NP_056993.2(NBAS):p.(Arg873Trp)

	Prediction/Score	**Interpretation**
Scaled CADD	33	>30 means that the variant is predicted to be in the top 0.1% of deleterious variants in the genome
MutationTaster	Disease causing (probability: value 0.99)	probability value close to 1 indicates a high security of the prediction
PolyPhen-2 Prediction	Probably damaging (score 1)	0.85-1 is predicted to be damaging
SIFT Prediction	Deleterious (score 0.001)	0-0.15 is predicted to be deleterious
PROVEAN_pred	Deleterious (score -5.25)	<-2.5 is predicted to be deleterious
Grantham score	101	>100 is considered a major amino acid change

Table S4. Relevant prediction algorithms from Human Splicing Finder supporting an alteration of a silencer motif in the *NBAS* gene

Prediction algorithm	cDNA Position	Reference silencer motif (%)	Mutant silencer (%)	Variation
Sironi et al.(Sironi, Menozzi et al. 2004) - Motif 2- (T/G)G(T/A)GGGG	399	agaaggg (75.13)	absent	Site broken (-15.1)
	400	gaagggg (77.13)	gaagcgg (60.23)	-21.92%
	401	aaggggg (63.88)	absent	Site broken (-5.73)
	402	agggggg (80.79)	agcgggg (80.79)	0%
	403	ggggggt (72.48)	absent	Site broken (11.46)
	404	gggggtg (67.64)	cggggtg (60.84)	-10.05%
Fas-ESS hexamers	401	aagggg	absent	Site broken
	402	aggggg	absent	Site broken
	403	gggggg	absent	Site broken
	404	gggggt	absent	Site broken
IIEs from Zhang et al.(Zhang, Li et al. 2008)	402	aggggg	absent	Site broken
	403	gggggg	gcgggg	N/A
	404	gggggt	absent	Site broken
HSF Matrices hnRNP A1 motif	400	gaaggg (68.57)	absent	Site broken (-100)
	401	aagggg (81.91)	aagcgg (66.67)	-18.60%

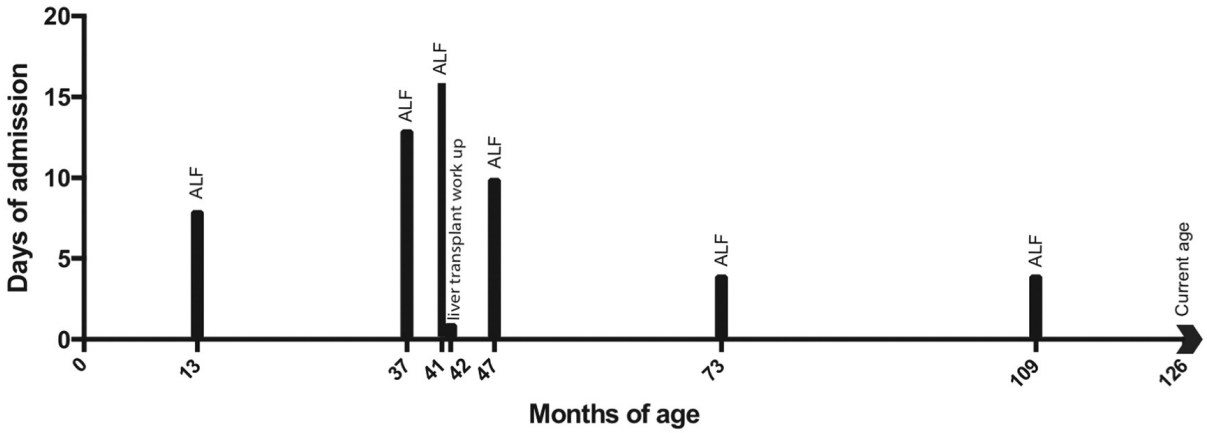

Fig. S1. Natural history of acute liver failure admissions in patient with *NBAS* biallelic variants. The severity and duration of acute liver failure (ALF) reduced with age.

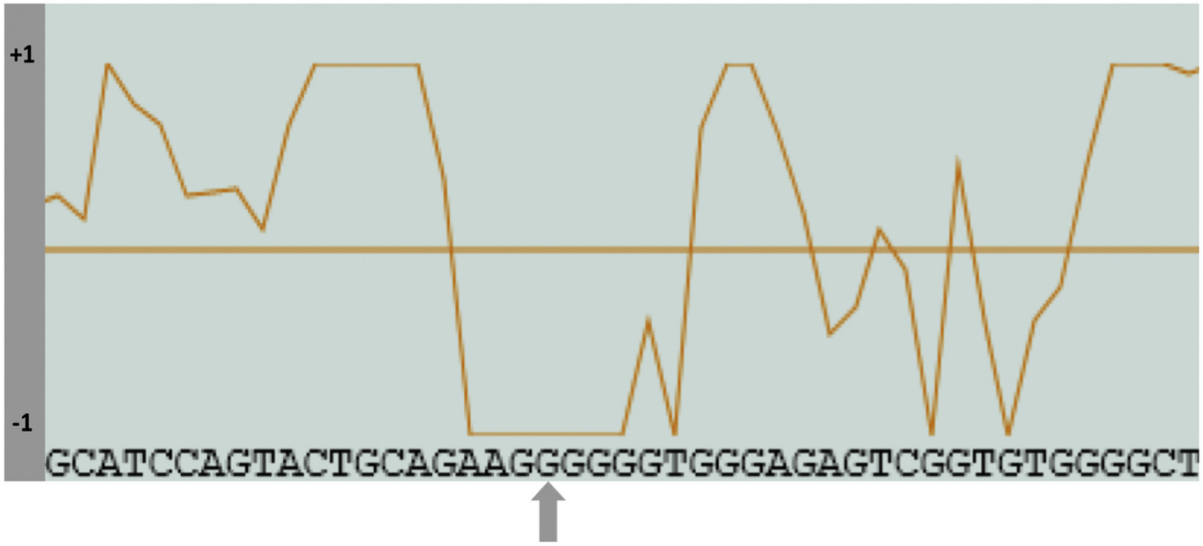

Fig. S2. SROOGLE Browser showing neighbourhood inference scores in deep intronic *NBAS* variant. Negative scores indicate that a given position resembles splicing silencers. The arrow shows the position of the NM_015909.3(*NBAS*):c.2423 + 404G > C variant.

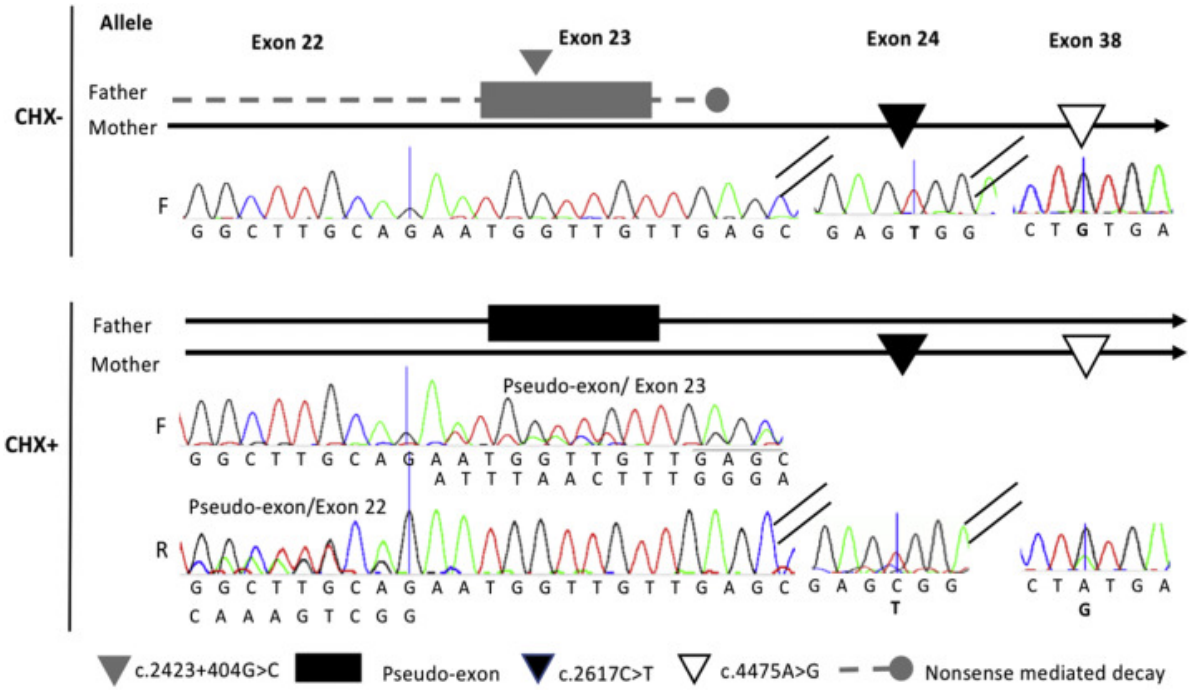

Fig. S3. Schematic representation and partial electropherograms of the RT-PCR products showing the *NBAS* variants identified in the patient as well as their molecular consequences. The deep intronic variant leading to the inclusion of a pseudo-exon is shown as a filled rectangle; unfilled triangles correspond to the likely benign variant c.4475A>G p.(Tyr1492Cys) segregating with the mother. This variant was apparently homozygous in Patient cDNA from cells grown without cycloheximide as was the c.2617C>T p.(Arg873Trp) variant, which is represented as a filled triangle, suggesting that the paternal transcript containing the pseudo-exon is degraded by nonsense mediated mRNA decay. Overlapping sequences can be recognized in the cells treated with cycloheximide, corresponding to the pseudo-exon and the wild type (WT) sequence of exon 23 (Primer forward F) or 22 (Primer reverse R).

Appendix C

C. Supplementary information for Chapter 5 Clinical Spectrum and Functional Consequences associated with Biallelic *PNPT1* variants

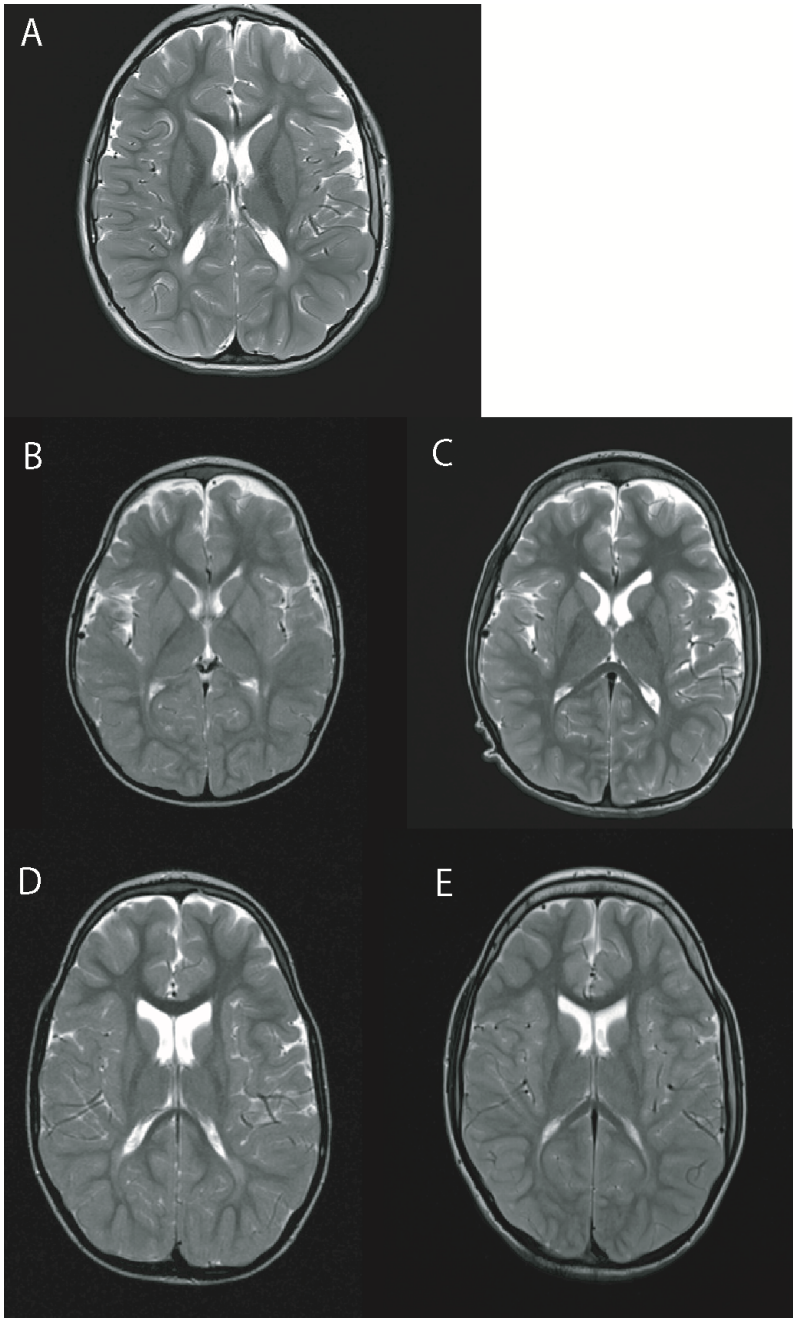

Figure S1. Brain MRI findings in patients with bi-allelic *PNPT1* variants. T2-weighted axial view showing (A) generalized neuroparenchymal volume loss with a posterior predilection in patient 2 at 5 years, (B) mild white matter atrophy in patient 5 at age 3 years and (C) at 11 years, (D) decrease in white matter plus non-specific white matter changes in the periventricular and parieto-occipital regions in patient 6 at 5 years and (E) at 13 years.

Table S1: Clinical phenotype of patients with biallelic *PNPT1* variants

Patient	Family	Reference	Gender	PNPT1 Variants NP_149100.2	Ethnicity	Signs at birth	Microcephaly	Abnormality of the head	Abnormality of the hand	Age at presentation	Presenting symptom(s)	Alive at last follow up	Age at last follow up (years)
1	1	Unpublished	M	p.(Thr531Arg) p.(Ala454Gly)	Anglo-Australian	Feeding difficulties, irritability	Yes	Small anterior fontanelle, dysmorphism in keeping with neurometabolic abnormality Narrow forehead; prominent inferior crus of antihelix	Puffy dorsum of hands and feet; tapering fingers	Birth	Feeding difficulties, irritability	Yes	5.5
2	2	Unpublished	M	p.(Ala507Ser) p.(Val607Lysfs*21)	Anglo-Australian	Foetal distress	No	No	No	3 weeks	Mild to moderate hearing loss	Yes	5.4

3	3	Unpublished	F	p.(Arg136Cys) p.(Pro467His)	Anglo-Australian	No	No	No	No	3 months	Poor head control, feeding difficulties	Yes	3
3.2	3	Unpublished	M	p.(Arg136Cys) p.(Pro467His)	Anglo-Australian	No	No	No	No	5 weeks	Mildly abnormal tone	Yes	1.2
4	4	Unpublished	M	p.(Arg738Cys) p.(Ala684Thr)	Anglo-Australian	No	No	No	No	11 years	Deafness, hearing and visual loss with ptosis and blepharospasm	Yes	78
5	5	Alodaib et al. Eur J Hum Genet. 2016.	M	p.(Gln254Lys) p.(Ala510Pro)	Anglo-Australian	Hypotonia, irritability, dystonic posturing, less severe than sib	No	Hypotonic facies	Small hands, curved finger tips/clubbing	Neonatal	Hypotonia	Yes	19
6	5	Alodaib et al. Eur J Hum Genet. 2016.	M	p.(Gln254Lys) p.(Ala510Pro)	Anglo-Australian	Hypotonia, irritability, poor feeding dystonic posturing,	yes, post natal	Hypotonic facies, triangular head, Coarse hair	Small hands, curved finger tips/clubbing	Neonatal	Hypotonia	Yes	17

7	6	Unpublished	Male	p.(Ala507Ser) p.(Thr531Arg)	European	No	No	No	Normal	Birth	Hypotonia	Yes	6
8	6	Unpublished	Male	p.(Ala507Ser) p.(Thr531Arg)	European	No	No	No	Normal	Birth	Hypotonia	Yes	10
9	7	Slavotinek et al. Clin Genet. 2015;88(5):468-73	M	NM_033109.5:c.404-1G>A (splicing) p.(Ala507Ser)	Northern European	Supplemental oxygen for 1 day; jaundice treated with phototherapy	Yes	Microbrachycephaly; mild bilateral ptosis; downslanting palpebral fissures; Broad, high nasal bridge; tented upper lip	Long fingers	2 months	Feeding and sleeping difficulties, little eye contact	Yes	7.9
10	8	Von Ameln et al. Am J Hum Genet. 2012;91(5):919-27.	M	p.(Glu475Gly)	Moroccan	No	NK	No	No	Childhood	Severe hearing impairment	Yes	24
11	8	Von Ameln et al. Am J Hum Genet. 2012;91(5):919-27.	F	p.(Glu475Gly)	Moroccan	No	NK	No	No	Childhood	Severe hearing impairment	Yes	20

12	8	Von Ameln et al. Am J Hum Genet. 2012;91(5):919-27.	M	p.(Glu475Gly)	Moroccan	No	NK	No	No	Childhood	Severe hearing impairment	Yes	19
13	9	Vedrenne et al. Am J Hum Genet. 2012;91(5):912-8	F	p.(Gln387Arg)	Moroccan	No	NK	No	NK	4 months	Feeding difficulties, vomiting, and swallowing difficulties	Yes	15
14	9	Vedrenne et al. Dhir et al. Nature. 2018;560(7717):238-42	M	p.(Gln387Arg)	Moroccan	No	NK	No	NK	6 months	Sudden motor regression with trunk hypotonia and choreoathetotic movements	Yes	13
15	10	Matilainen et al. Hum Mol Genet. 2017;26(17):3352-61.	F	p.(Arg136His) p.(Pro140Leu)	Vietnamese	Yes	NK	No	NK	1 month	Hypotonia	No, due to infection	2.4
16	11	Sato et al., Clin Genet. 2018;93(2):242-7.	F	p.Gly76Asp) p.(Arg192*)	Japanese	hypotonia	NK	Characteristic facial features, including an upslanted palpebral fissure,	NK	Congenital	Hypotonia	No, due to acute encephalopathy	4

								epicanthus, a depressed nasal ridge and protruding ears.					
17	11	Sato et al. Clin Genet. 2018;93(2):242-7.	M	p.(Gly76Asp) p.(Arg192*)	Japanese	transient tachypnoea	Yes	Upslanted palpebral fissures, epicanthus, a depressed nasal ridge and protruding ears.	NK	Neonatal	Neonatal - transient tachypnoea , hypertonia at 3 months	Yes	3
18	12	Eaton et al. Am J Med Genet A. 2018;176(11):2487-93.	M	p.(Glu584Glyfs*17) p.(Met745Thr)	NK	NK	NK	NK	NK	Neonatal	SNHL	Yes	64
19	12	Eaton et al. Am J Med Genet A. 2018;176(11):2487-93.	M	p.(Glu584Glyfs*17) p.(Met745Thr)	NK	NK	NK	NK	NK	Neonatal	SNHL	Yes	60
20	12	Eaton et al. Am J Med Genet A. 2018;176(11):2487-93.	F	p.(Glu584Glyfs*17) p.(Met745Thr)	NK	NK	NK	NK	NK	Neonatal	SNHL	Yes	40
21	13	Dhir et al. Nature. 2018;560(7717):238-40	M	p.(Arg136His)	NK	Congenital cataract	NK	NK	NK	Neonatal	Cognenital cataract	No, reason not stated	2

22	14	Dhir et al. Nature. 2018;560(7717):238-41	M	p.(Ser70Pro) p.(Asp713Tyr)	NK	NK	NK	NK	NK	NK	NK	Yes	1
23	15	Dhir et al. Nature. 2018;560(7717):238-42	F	p.(Gly499Arg) p.(Ala507Ser)	NK	NK	NK	NK	NK	NK	NK	Yes	7

Patient	**Family**	**Neurodevelopmental abnormality**	**Language**	**Gait disturbance**	**Abnormal muscle tone**	**Abnormal reflexes**	**Abnormality of movement**	**Seizures**	**EEG abnormalities**	**Neuropathy**
1	1	Global developmental delay, profound intellectual disability	Non-verbal	Inability to walk	Central hypotonia with limb hypertonia	Yes, difficult to elicit all reflexes	Dystonia	Infantile Spasms, Tonic head and eye deviation, arm posturing starting between 1-2 years	Abnormal multifocal epileptiform activity with posterior bias	Mild demyelinating
2	2	Global developmental delay, no regression	Non-verbal	Inability to walk	Hypotonia	Reduced	No	Multifocal seizure disorder induced by drowsiness and sleep on EEG, presenting at 2 years clinically	Sleep EEG - multifocal seizure disorder induced by drowsiness and sleep Awake EEG - abnormal background	No

3	3	Global developmental delay, no regression	Non-verbal	Inability to walk	Central hypotonia with peripheral hypertonia	Yes, pathologically brisk	Dystonia, Choreoathetosis	Possible seizure at 8 months	normal	NK
3.2	3	Normal at 14 months	10 words at 14 months	Beginning steps at 14m	Mildly abnormal tone in the neonatal period, normal tone at 14months	Normal	No	No	NK	NK
4	4	Loss of skills/ regression	Yes	Broad based, Ataxia	No	normal	Gait ataxia	Syncope but normal EEG	normal EEG	Large fibre neuropathy
5	5	Severe global developmental delay/no regression, profound intellectual disability	No	Inability to walk	Hypotonia	areflexia	Dystonia (less than brother), Choreoathetosis	No	NK	Sensory and autonomic: self-mutilation lips Nerve biopsy: severe chronic axonal neuropathy
6	5	Severe global developmental delay/ no regression, profound intellectual disability	No	Inability to walk	Hypotonia	Areflexia	Dystonia, generalized but more in arms	No	NK	Sensory and autonomic, bites hands Nerve Conduction-motor normal, decreased in sensory nerves

										Electromyography - no significant activity at rest Muscle biopsy- evidence of denervation possibly due to axonal neuropathy Sural nerve biopsy- consistent with severe axonal neuropathy
7	6	Global developmental delay	1st words at 24 months - fluent at 5 years	Crawled at 2 years; unable to walk at 5 years	Hypotonia	Areflexia	Leg dystonia	No	NK	Sensory
8	6	Global developmental delay	1st words at 2 1/2 years - fluent at 10 years	Crawled at 2 years; unable to walk at 10 years	Hypotonia	Areflexia	Leg dystonia	No	Normal	Sensory

9	7	Global developmental delay, profound intellectual disability	Non-verbal	No	Decreased axial tone; increased appendicular tone	Yes - brisk biceps and patellar reflexes	Dystonia, myoclonic jerks	Tonic seizures at 2 months	Right posterior quadrant epileptiform discharges, poorly organized background, abundant high-amplitude multifocal spikes	No
10	8	No	No	No	No	No	No	No	NK	NK
11	8	No	No	No	No	No	No	No	NK	NK
12	8	No	No	No	No	No	No	No	NK	NK
13	9	Good development until 4 months, at 9 months with loss of purposeful hand movements	No	Inability to walk	Sudden hypotonia at 9 months	Deep-tendon reflexes of the inferior limbs were barely detected	Dystonia at 9m, buccofacial dyskinesia, choreoathetosis, myoclonies	No	Clinically silent erratic myoclonies	Reduced nerve-conduction velocity
14	9	Good development until 4 months, motor regression	No	Inability to walk	Sudden Hypotonia at 6m	NK	Dystonia at 6m, Choreoathetosis, myoclonies	No	Clinically silent erratic myoclonies	NK

15	10	Global developmental delay	No	Inability to walk	Hypotonia	NK	Buccofacial dyskinesia, upper limb movements.	At 2months Infantile Spasms, at 10years daily seizures, controlled with phenobarbital and vigabatrin	NK	NK
16	11	Severe developmental delay	NK	Inability to walk	Hypotonia	NK	NK	Seizures	Hypsarrhythmia consistent with late-onset West syndrome	NK
17	11	Global developmental delay, at 3years (Enjoji Scale of Infant Analytical Development) was 15 months.	NK	Inability to walk	Hypertonia at 3 months that gradually changed to severe hypotonia by 3years	NK	NK	Infantile spasms at 2 years of age, controlled by valproate.	Hypsarrhythmia consistent with late-onset West syndrome	Short-latency somatosensory-evoked potentials showed decreased central conduction velocity and amplitude
18	12	Regression in adulthood, progressive cognitive decline	Sign language	Ataxia at 40 years Wheelc hair-dependent at 47 years	Significantly decreased tone of the lower limbs in adulthood	Diminished in the upper limbs and absent in the lower limbs.	Dystonia, Ataxia	No	NK	NK

19	12	Regression in adulthood, progressive cognitive decline	Sign language	Ataxia at age 45 years, Wheelc hair dependent by age 58 years	NK	Diminished reflexes in all four limbs	Dystonia, Ataxia	No	NK	NK
20	12	Regression in adulthood, progressive cognitive decline	Sign language and a few verbal words.	Progressive balance issues at 30 years, ataxia, wheelch air-dependent by 40 years	Increased tone in the lower extremities	Brisk reflexes and clonus	Dystonia, Ataxia	No	NK	NK
21	13	NK	NK	NK	Truncal hypotonia	NK	NK	NK	NK	NK
22	14	NK	NK	NK	Truncal hypotonia with hypertonia	NK	Dystonia	NK	NK	NK

					of the lower limbs					
23	15	NK	NK	NK	Hypotonia	NK	NK	NK	NK	Sensory neuropathy

Patient	Family	Abnormality of eye movement	Visual impairment	Abnormal fundus morphology	Sensorineural hearing impairment	Feeding difficulties	Abnormality of oesophagus physiology	Episodic vomiting	Constipation	Respiratory problems	Abnormality of the cardiovascular system
1	1	Nystagmus	Yes	Optic atrophy	NK	G-tube fed and removed after complications ; changed to NGT feeds	Gastroesophageal reflux, Dysphagia	Yes	Yes	Yes, required NPA due to upper airway obstruction, nocturnal desaturations and sleep disturbance; recurrent aspiration pneumonia	Patent foramen ovale, Hypertension treated with amlodipine

2	2	Nystagmus	Yes	Optic atrophy	Yes	yes	Gastroesophageal reflux, dysphagia	No	No	No	no
3	3	NK	NK	NK	NK	Yes	Gastroesophageal reflux	No	Yes	Reactive airway disease	no
3.2	3	NK	NK	NK	NK	No	No	No	Yes	No	no
4	4	Progressive ophthalmoplegia	Yes	Optic atrophy	Yes	NK	NK	No	NK	No	No
5	5	Nystagmus	Yes -Moderate, visual acuity 6/45	Optic atrophy	Yes	Gastroparesis	Gastroesophageal reflux	No	Yes	Chronic lung disease due to aspiration	No
6	5	Rotatory and horizontal Nystagmus	Yes	Optic atrophy	Yes	Gastroparesis	Severe Gastroesophageal reflux, required fundoplication	Yes, every 1- 6 weeks lasting hours-days, controlled at 6 years	Yes	Chronic lung disease, Bronchiectasis	No
7	6	Strabismus	No	No	No	No	Gastroesophageal reflux	No	Yes	Obstructive sleep apnoea	Trabeculations
8	6	Nystagmus	Yes	Optic atrophy	No	Yes	Gastroesophageal reflux, Dysphagia	No	No	Obstructive sleep apnoea	Trabeculations

9	7	Nystagmus	Yes	Optic atrophy, chorioretinal defect in left eye	Yes	Yes	Gastroesophageal reflux, Dysphagia	Yes	Yes	Cold-induced asthma	Tachycardia, Hypertension
10	8	No	No	NK	Yes	No	No	No	No	No	No
11	8	No	No	NK	Yes	No	No	No	No	No	No
12	8	No	No	NK	Yes	No	No	No	No	No	No
13	9	No	No	No	NA	Major swallowing difficulties required gastrostomy	Dysphagia	Presented with vomiting at 4 months	NK	NK	NK
14	9	NK	No	NK	NK	NK	NK	NA	NK	NK	NK
15	10	Nystagmus	Yes	NK	No	Yes	NK	Vomiting episodes at 10 months	NK	NK	No
16	11	Nystagmus	NA	NK	No	NK	NK	NK	NK	NK	Ventricular septal defect
17	11	NK	Normal visually evoked potentials	NK	No	NK	NK	NK	NK	NK	NK

18	12	NK	Progressive visual impairment visual acuity of 20/400.	Optic Atrophy	Yes	Swallowing difficulties	NK	NK	NK	NK	Aortic root dilatation and proximal septal thickening
19	12	NK	Progressive Visual impairment visual acuity of 20/400	Optic Atrophy	Yes	Swallowing difficulties	NK	NK	NK	NK	NK
20	12	NK	Yes	Optic Atrophy	Yes	Swallowing difficulties	NK	NK	NK	NK	NK
21	13	Nystagmus	Congenital cataracts	NK	Yes	Yes	NK	NK	NK	NK	NK
22	14	Nystagmus	NK	NK	NK	NK	NK	NK	NK	NK	NK
23	15	Abnormal eye movements	NK	NK	Yes	NK	NK	NK	NK	NK	NK

Patient	Family	Abnormality of the genitourinary system	Sleep disturbance	Scoliosis	Abnormality of the skeletal system	Abnormality of immune system physiology	Other	Raised serum lactate	Raised CSF lactate	Respiratory chain deficiency by spectrophotometry (Tissue /Complex)	MRI abnormalities
1	1	NK	Yes	Yes	Scoliosis, Bilateral hip dislocation that required varus derotation osteotomy at age 4 years	Recurrent infections	NK	No	No	Normal in fibroblasts	Reduction in white matter volume Thin corpus callosum
2	2	NK	Yes	No	No	Recurrent Ear infections	NK	Yes	No	yes, muscle/CIV	Generalized white matter volume loss with a posterior predilection
3	3	NK	NK	NK	NK	Reactive airway disease	NK	NK	NK	NK	Abnormal signal in the pons, cerebellar peduncles, medullary dorsal tracts and peri-aqueductal grey matter
3.2	3	NK	NK	NK	NK	No	NK	NK	NK	NK	NK
4	4	NK	NK	NK	NK	NK	Depression	No	NK	NK	NK

5	5	NK	Yes	yes, milder	Hip dysplasia, Scoliosis (milder than brother)	NK	Less severely affected than brother, increased appetite	No	Yes	Normal in muscle and fibroblasts	Mild white matter atrophy
6	5	Urinary retention	No	yes, progressive	Hip dysplasia, progressive scoliosis	Abnormal histamine response on a skin testing	Bulbar palsy, bruxism, heat intolerance, low sweating, poor temperature regulation	No	No	NK	Non-specific white matter changes in the periventricular and parieto occipital regions
7	6	Urinary incontinence	Yes	No	No	Allergy	NK	No	NK	NK	NK
8	6	Urinary incontinence	Yes	No	No	Allergy	NK	No	NK	NK	Thin corpus callosum
9	7	Renal artery stenosis	Yes	Yes	Bilateral hip dislocation, repaired by open reduction and pelvic osteotomy, progressive scoliosis	Recurrent infections, Multiple hospitalizations for viral illnesses, Cold-induced asthma	Cholelithiasis; precocious puberty (pubic hair noted age 7 y, without other signs of puberty); chronic iron deficiency anaemia	No	NK	NK	Global cerebral volume loss, delayed maturation, simplified gyral pattern, Thin corupus callosum

							treated with iron infusions				
10	8	No	No	No	No	NK	NK	NK	NK	NK	NK
11	8	No	No	No	No	NK	NK	NK	NK	NK	NK
12	8	No	No	No	No	NK	NK	NK	NK	NK	NK
13	9	NK	NK	NK	Major retractions of the right hip and ankles prompted a consideration of multiple tenotomies	NK	NK	Yes	Yes	Yes, liver/CIV and CIII	Bilateral hyperintensities in putamen and caudate nuclei
14	9	NK	NK	NK	NK	NK	NK	Yes	Yes	No	Hyperintensities in the bilateral putamen and caudate nuclei
15	10	NK	NK	NK	NK	NK	NK	Yes	No	NK	Bilateral signal intensity of globus pallidus
16	11	NK	NK	NK	NK	NK	NK	No	No	NK	Delayed myelination
17	11	NK	NK	NK	NK	NK	NK	No	No	NK	Delayed myelination, diminished white matter, thinning of the corpus callosum and reduced volume of the cerebrum and brainstem

18	12	Urinary incontinence	NK	NK	NK	NK	NK	NK	NK	NK	Mild diffuse cerebellar atrophy, cerebral atrophy. Small cystic-like lesions of cerebral peduncles
19	12	Urinary incontinence	NK	NK	NK	NK	Obsessive compulsive disorder	NK	NK	NK	No
20	12	NK	NK	NK	NK	NK	Depression, paranoia	NK	NK	NK	Small hyperintensities in the cerebral hemispheric white matter
21	13	NK	NK	NK	NK	NK	NK	Yes	NK	Multiple deficiency in muscle and fibroblasts	Anomalies of putamen and basal ganglia, lactate peak in MR spectroscopy
22	14	NK	NK	NK	NK	NK	NK	NK	NK	Multiple deficiency in fibroblasts	NK
23	15	NK	NK	NK	NK	NK	NK	NK	NK	Normal in fibroblasts	Delayed myelination, abnormal corpus callosum

NK Not known, NA Not available, SNHLSensorineural hearing loss

Table S2. *In silico* studies and population frequency of *PNPT1* variants identified in new families

Family	Genomic location	Protein NP_149100.2	cDNA NM_033109.4	Population frequency (GnomAD)	Polyphen	SIFT	Mutation Taster	Amino acid change	CADD Score[+]	Conservation	Other
1	chr2:55874492G>C	p.(Thr531Arg)	1592C>G	0.00003560 (10 /28092) 0 homozygous	probably damaging	deleterious	disease causing	71	6.68	very high	Also identified in family 6 in trans with a known pathogenic variant p.(Ala507Ser), listed in ClinVar as Likely pathogenic and Uncertain significance (Variant ID: 209184)
1	chr2:55883346G>C	p.(Ala454Gly)	1361C>G	0.00000398 9 (1/ 250714) 0 homozygous	probably damaging	deleterious	disease causing	60	6.5 6	very high	Submitted to ClinVar as Likely pathogenic (1) (Variant ID: 548987)
2	chr2:55872488A>G	p.(Val607Lysfs*21)	c.1818T>G	not reported	NA	NA	disease causing	NA	0.64	very high	Human Splice Finder: Activation of an exonic cryptic donor site. Potential alteration of splicing. Mutation Taster: splice site changes
2	chr2:55874565C>A	p.(Ala507Ser)	c.1519G>T	0.0002126 (60/282190)	possibly damaging	tolerated	disease causing	99	5.68	very high	Listed in ClinVar as Likely pathogenic

				0 homozygous							(Variant ID: 215010), reported in 4 of the 15 families.
3	chr2:55910967G>A	p.(Arg136Cys)	c.406C>T	0.00001063 (3/282296) 0 homozygous	possibly damaging	deleterious	disease causing	180	7.38	very high	Alternative amino acid change is reported in 2 of the 15 families
3	chr2:55883307G>T	p.(Pro467His)	c.1400C>A	0.000003984 (1/251016) 0 homozygous	probably damaging	deleterious	disease causing	77	6.14	very high	
4	chr2:55863512G>A	p.(Arg738Cys)	c.2212C>T	0.0002601 (73/280624) 0 homozygous	probably damaging	deleterious	disease causing	180	6.84	very high	
4	chr2:55870312C>T	p.(Ala684Thr)	c.2050G>A	0.00001193 (3/251420) 0 homozygous	probably damaging	deleterious	disease causing	58	8.06	very high	
6	chr2:55874565C>A	p.(Ala507Ser)	c.1519G>T	0.0002126 (60/282190) 0 homozygous	possibly damaging	tolerated	disease causing	99	5.68	very high	Listed in ClinVar as Likely pathogenic (Variant ID: 215010), reported in 4 of the 15 families.
6	chr2:55874492G>C	p.(Thr531Arg)	1592C>G	0.00003560 (10 /28092) 0 homozygous	probably damaging	deleterious	disease causing	71	6.68	very high	Also identified in family 1, listed in ClinVar as Likely pathogenic and Uncertain significance (Variant ID 209184)

*GS: Grantham score <65 minor, >65 moderate, >100 major; [+]CADD raw score threshold >1.75 = deleterious