Fakultät für Medizin

Institut für Humangenetik

Mitochondrial disease: elucidating genetic aetiology by variant discovery, validation, and integration with clinical phenotype

Sarah Louise Stenton

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines

Doctor of Philosophy (Ph.D.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Thomas Korn

Betreuer: Prof. Dr. Thomas Meitinger

Prüfer der Dissertation:

1. Prof. Dr. Peter Freisinger

2. Prof. Dr. Thomas Klopstock

3. Prof. Dr. Johannes Zschocke

Die Dissertation wurde am 26.05.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 09.09.2021 angenommen.

Genes are not about inevitabilities; they're about potentials and vulnerabilities.

Robert Sapolsky

in Behave: The Biology of Humans At Our Best and Worst

ZUSAMMENFASSUNG

Mitochondriale Erkrankungen sind eine heterogene Gruppe von Erkrankungen, welche durch Sequenzveränderungen in mehr als 400 Genen ausgelöst werden können und den mitochondrialen Energiestoffwechsel betreffen. Mit einer Prävalenz zwischen 5 und 20 je 100.000 sind sie die häufigsten angeborenen Stoffwechselstörungen und stellen 10% aller Mendelschen Erkrankungen. Mitochondriale Erkrankungen ma- nifestieren sich mit einer mannigfachen klinischen Variabilität, betreffen vorzugsweise Geweben mit hohem Energiebedarf und dekompensieren in metabolischen Stresssituationen. Die Korrelation zwischen Genotyp und Phänotyp ist zumeist wage und es gibt nur wenige klare klinische Syndrome. Nur selten erlaubt der Phänotyp einen Rückschluss auf den zugrundeliegenden Gendefekt and eine gezielte molekulare Diagnostik, insbesondere bei Patienten mit pädiatrischem Krankheitsbe- ginn. Aus diesem Grund hat sich die Sequenzierung der kodierenden Bereiche aller Gene, also des gesamten Exoms (WES), zum bevorzugten diagnostischen Ansatz entwickelt. Dieser erlaubt es, mehr als 80% aller krankheitsverursachenden Varianten zu erfassen. Das erste und übergeordnete Forschungsziel dieser Doktorarbeit war die Bestimmung des diagnostischen Potenzials der WES für mitochondriale Erkrankungen.

In Studie 1 (Kapitel 3.1) wurde der Einsatz von WES bei insgesamt über 2.000 pädiatrischen Patienten mit einer klinisch vermuteten mitochondrialen Erkrankung analysiert. Diese Patienten wurden nach primärer Analyse an lokalen Zentren im Rahmen einer internationalen Kooperation zusammengeführt. Insgesamt wurde eine diagnostische Rate von 54% erreicht, wobei 30% der Patienten sofort mit pathogenen oder wahrscheinlich pathogenen Varianten diagnostiziert werden konnten, 5% durch die Beschreibung von insgesamt 63 neuen Krankheitsgenen definiert wurden, 4% erst im Laufe der Zeit nach Reanalyse mit neuen Erkenntnissen diagnostiziert werden konnten und weitere 15% aufgrund zusätzlicher funktioneller Evidenz geklärt wurden. Die Notwendigkeit für funktionelle Validierungsstudien für die Varianteninterpretation wurde hauptsächlich durch unbeschriebene Missense-Varianten bestimmt. Diese Untersuchungen erforderten neben Blut auch Biopsiematerial vom Patienten und betont die Notwendigkeit einer proaktiven Probenahme in der klinischen Praxis hervor. Insgesamt wurden pathogene Varianten in knapp der Hälfte der 416 derzeit beschriebenen Gene für mitochondriale Erkrankungen erfasst. Für 28 dieser Gene ist eine defektspezifische Behandlung mit

potenziellem Nutzen wie Supplementierung eines kritischen Cofaktors oder einer Ernährungsumstellung beschrieben, dies betrifft 174 Patienten. Die restlichen fast 46% der Fälle erhielten keine genetische Diagnose. Wobei es bei 32% der Patienten keinen Anhaltspunkt aus der genetischen Analyse gab und bei 14% mögliche pathogene Varianten gefunden aber bisher nicht Validiert wurden. Die Studie führte zur Kuratierung einer einmaligen Ressource von standardisierten Patientendatensätzen, die ich zu einem frei zugänglichen Webbrowser GENOMITexplorer entwickelt habe. Diese Ressource umfasst fast 4.000 Patienten-HPO-Datensätze und enthält neben der WES-Studie, Daten von zwei nationalen Patientenregistern und der Literatur. Sie gibt Auskunft über 400 verschiedenen Krankheitsgene.

Studie 2 (Kapitel 3.2) behandelt die Validierung eben beschriebener Varianten unklarer Pathogenität (VUS) in dem mitochondrialen Krankheitsgen FDXR. Diese Missense- und In-Frame-Indel-Varianten wurden in Patienten mit untypischem Phänotyp und Schweregrad identifiziert. Aufgrund der Nichtverfügbarkeit von Biopsieproben der Patienten wurde für die Validierung ein Hefemodell genutzt.

Studie 3 (Kapitel 3.3) beschreibt die Validierung von einem neuen kernkodierten Krankheitsgen, DNAJC30, für die Leberscher Optikusneuropathie (LHON) für die bisher nur die mütterliche Vererbung bekannt war. Bei insgesamt 28 Familien konnten pathogene Varianten gefunden werden. Die Studie belegt damit eine autosomal rezessive Vererbung und plädierte für eine Unterteilung von LHON in mütterliches (mtLHON) und rezessives LHON (arLHON) mit wichtigen Implikationen für die genetische Beratung, einschließlich einer unmittelbaren Indikation für eine vorteilhafte Therapie mit Idebenon. Die Pathogenität einer häufigen Foundermutation wurde durch eine höhere als erwartete Allelfrequenz (0,125% in der Allgemeinbevölkerung) in Frage gestellt und durch unvollständige Penetranz geklärt. Darüber hinaus wurde der Krankheitspathomechanismus durch funktionelle Studien an Fibroblasten von Patienten und die Erzeugung eines DNAJC30-Knock-Out-Zellmodells aufgedeckt. Im Rahmen einer Kollaboration konnte ich einen Defekt in der Reparatur des mitochondrialen Atmungskettenkomplex I nachweisen, der einem besonders hohen Risiko oxidativer Schäden ausgesetzt ist.

Studie 4 (Kapitel 3.4) zeigt, wie mit einem integrativen multi-omischen Ansatz ungelöste WES-Fälle diagnostiziert werden können. Hierbei haben wir genetische Daten mit HPO-kodierte Phänotypen, RNA-Sequenzierungs und quantitative proteomische Daten integrativ ausgewertet. In etwas mehr als 20% der Fälle wurde konnte hiermit eine genetische Diagnose gestellt werden. Pathologische Befunde wurde hauptsächlich durch reduzierte Proteinlevel aufgrund von Protein-destabilisierenden Missense-Varianten und durch Defekte der Genexpression nachgewiesen. Diese Arbeit konzentrierte sich auf die Diagnose von zwei Fällen, in denen der integrativer Ansatz VUS neu priorisierte

und gleichzeitig validierte. In beiden Fällen wurden die Varianten in der WES-Analyse identifiziert aber aufgrund der hohen Allelfrequenz bzw. widersprechender Segregation ausgeschlossen. Der Nachweis der Pathogenität solcher Varianten war ein Bereich von besonderem Interesse für diese Arbeit.

ABSTRACT

Mitochondrial diseases are a heterogeneous collection of diseases due to defects in over 400 genes, with roles converging on mitochondrial energy metabolism. They have an estimated prevalence of 5-20 per 100,000, are the most frequent inborn error of metabolism, and are accountable for ∼10% of Mendelian disease genes. Mitochondrial diseases manifest with remarkable clinical variability, have a predilection for the high-energy demanding tissues, and characteristically decompensate in times of metabolic stress. The association between genotype and phenotype is limited, and only a handful of distinct clinical syndromes arise due to variants in specific genes. This deems targeted sequencing in the majority of patients challenging, particularly those of paediatric-onset with a non-specific presentation. For these reasons, comprehensive capture of disease-causing variation by whole exome sequencing (WES) is rapidly becoming the diagnostic approach of choice, with the capacity to detect over 85% of reported disease-causing variation within the coding regions of the genome. Accordingly, the overarching objective of my doctoral research was to determine the diagnostic power of WES in mitochondrial disease, addressed by four studies presented in this thesis.

Study 1 (Chapter 3.1) analysed the application of WES in an aggregate of over 2,000 paediatric clinically-suspected mitochondrial disease patients. These patients were gathered and analysed in an international collaborative effort. In total, a diagnostic rate of 54% was achieved. This 54% was stratified into 30% with immediately identifiable pathogenic or likely pathogenic variants, 15% in which functional studies were needed to designate variant pathogenicity, 5% involved in one of 63 novel disease gene discoveries, and 4% in which WES reanalysis identified the causative variants. The requirement of functional validation studies for variant interpretation was driven by the predominance of rare missense variants, as candidates for causative DNA variation. These studies required patient-derived bio-material in addition to blood, and highlighted the need for proactive bio-material sampling in clinical practice. In total, pathogenic variants in just under half of the 416 currently described mitochondrial disease genes were captured, of which 28, with defects across 174 patients, were reported to have a defect-specific treatment of potential benefit, such as supplementation of a critical cofactor or dietary modification. The remaining 46% of cases did not receive

a genetic diagnosis. This figure was divided into 14% with variants of uncertain significance (VUS) and 32% in which candidate variants were unable to be prioritised for further exploration. The study led to the curation of a valuable resource of patient-level and gene-level HPO data sets, which I developed into a freely-accessible web browser, GENOMITexplorer. This resource shares almost 4,000 patient HPO data sets from our WES study, patient registries, and the literature, in total spanning genotype-phenotype associations of over 400 disease genes.

Study 2 (Chapter 3.2) validated VUS in the known mitochondrial disease gene, *FDXR*. These missense and in-frame indel variants were prioritised by WES in patients with clinical variability beyond the recognised disease-associated phenotype and reported disease severity. To evaluate pathogenic relevance in the absence of patient-derived bio-material, a yeast model was employed.

Study 3 (Chapter 3.3) discovered a novel nuclear-encoded mitochondrial disease gene, *DNAJC30*, in a total of 28 unrelated patients with Leber's hereditary optic neuropathy (LHON), a disease considered to be exclusively maternally inherited. The study associated an autosomal recessive mode of inheritance to the disease, arguing for sub classification of LHON into maternal LHON (mtLHON) and recessive LHON (arLHON), with important implications for genetic counselling including an immediate indication for a beneficial therapy, idebenone. The pathogenicity of defects in *DNAJC30* was challenged by a higher than expected allele frequency of the causative founder variant (0.125%, in the general population), shared by almost all patients and explained by incomplete penetrance. Moreover, by functional studies of patient-derived cell lines and generation of a *DNAJC30*-knock-out cellular model, the disease pathomechanism was unravelled. Together with a team of collaborators, I could demonstrate impairment of mitochondrial complex I repair due to a defect in the exchange of specific subunits in the assembled complex, exposed to high risk of oxidative damage.

Study 4 (Chapter 3.4) applied an integrative multi-omic diagnostic approach to unsolved WES/WGS cases, encompassing genetic, HPO encoded phenotype, RNA-sequencing, and quantitative proteomic data. A genetic diagnosis was made in just over 20% of cases. Pathogenic variant consequence was primarily detected by protein underexpression, due to protein destabilising missense variants, and by disruption of RNA isoform and abundance. This thesis focused on the diagnosis of two cases in which our integrative approach reprioritised and simultaneously validated VUS that had evaded prioritisation by the standard applied strategy due to features arguing against pathogenicity. Demonstrating the pathogenicity of such variants was an area of interest for this thesis.

Contents

List of figures

List of tables

List of abbreviations

acetyl CoA	acetyl coenzyme A
ACMG	American College of Medicine Genetics
AD	autosomal dominant
ADP	adenosine diphosphate
AR	autosomal recessive
ATP	adenosine triphosphate
AUC	area under the curve
BN-PAGE	blue native polyacrylamide gel electrophoresis
BNE	blue native electrophoresis
bp	base pair
c.	cDNA sequence position
CADD	Combined Annotation Dependent Depletion
CI-V	mitochondrial respiratory chain complex I-V
CNV	copy number variant
COX	cytochrome c oxidase
CPEO	chronic progressive external ophthalmoplegia
CRR	clinically relevant recovery
CS	citrate synthase
CSF	cerebrospinal fluid
DDM	detergent dodecyl maltoside
DEA	differential expression analysis
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotides
E. coli	*Escherichia coli*
ECHO	echocardiography
EDTA	ethylenediaminetetraacetic acid
EEG	electroencephalography
EMA	European Medicines Agency
ETC	electron transport chain
F	forward
FADH2	flavin adenine dinucleotide
FCCP	carbonyl cyanide-p-trifluoromethoxyphenylhydrazone
FGF21	fibroblast growth factor 21
FPKM	Fragments Per Kilobase of transcript per million fragments mapped
GDF15	growth differentiation factor 15
gDNA	genomic DNA

gnomAD	Genome Aggregation Database
GSEA	gene set enrichment analysis
HEK	human embryonic kidney
hg19	human genome assembly GRCh37 UCSC
HGMD	Human Gene Mutation Database
HPO	Human Phenotype Ontology
IBAQ	intensity-based absolute quantification
IEM	inborn error of metabolism
IGV	Integrative Genomics Viewer
IMM	inner mitochondrial membrane
indel	small insertion and deletion variation
kDa	kilodalton
KSS	Kearns-Sayre syndrome
LB	lysogeny broth
LHON	Leber's hereditary optic neuropathy
LogMAR	Logarithm of the Minimum Angle of Resolution
log2fc	log base 2 fold change
m.	mitochondrial DNA sequence position
MAE	mono-allelic expression
MAF	minor allele frequency
MDC	Mitochondrial Disease Criteria
MELAS	mitochondrial myopathy, encephalopathy, lactic acidosis, stroke-like episodes
MERRF	myoclonic epilepsy with ragged-red fibres
MNGIE	mitochondrial neurogastrointestinal encephalopathy syndrome
MRI	magnetic resonance imaging
mRNA	messenger RNA
MRS	magnetic resonance spectroscopy
mtDNA	mitochondrial DNA
NARP	neuropathy, ataxia and retinitis pigmentosa
NGS	next generation sequencing
NHDF	normal human dermal fibroblasts
NM_	RefSeq mRNA-sequence
NMD	nonsense mediated decay
NMDAS	Newcastle Mitochondrial Disease Adult Scale
NP_	RefSeq protein sequence
NPMDS	Newcastle Paediatric Mitochondrial Disease Scale
NUMT	nuclear mitochondrial DNA
OCR	oxygen consumption rate
OMIM	Online Mendelian Inheritance in Man
ORF	open reading frame

OXPHOS	oxidative phosphorylation
p.	protein sequence position
PCR	polymerase chain reaction
pSILAC	pulsed Stable Isotope Labeling by Amino acids in Cell culture
PVDF	polyvinylidene fluoride
R	reverse
RCC	respiratory chain complexes
RGC	retinal ganglion cell
RIN	RNA integrity number
RNA	ribonucleic acid
RNA-seq	RNA-sequencing
RNFL	retinal nerve fiber layer
ROC	receiver operating characteristic
ROS	reactive oxygen species
rRNA	ribosomal RNA
RT	room temperature
RT-PCT	reverse transcription polymerase chain reaction
s.d.	standard deviation
SDS	sodium dodecyl sulphate
SIFT	Sorting Intolerance from Tolerance
SNV	single nucleotide variant
TBST	Tris-buffered saline and Tween 20
TCA	tricarboxylic acid
Tm	melting temperature
TMT	tandem mass tag
Tris	2-amino-2(hydroxymethyl)-1,3-propandiol
tRNA	transfer RNA
U	unit
UTR	untranslated region
v/v	volume per volume
VEP	Ensembl Variant Effect Predictor
VUS	variant(s) of uncertain significance
w/v	weight per volume
WES	whole exome sequencing
WGS	whole genome sequencing
YPD	yeast-extract-peptone-dextrose
YPG	yeast-extract-peptone-glycerol

1 Introduction

1.1 Mitochondria

1.1.1 Mitochondrial function

Mitochondria are critical to cellular metabolism and thereby essential for eukaryotic cells, in turn responsible for the composition of all complex life. They originated from endosymbiosis between prokaryotes approximately four billion years ago [1] and are present in hundreds to thousands per cell. As the "powerhouse of the cell", mitochondria were crucial for the evolution of eukaryotic cells by releasing the bioenergetic constraint on genome size, permitting a 200,000 fold expansion in the number of genes expressed [2].

Mitochondrial energy generation produces the basic unit of cellular energy, ATP, by metabolising carbohydrates, fats, and proteins [3]. Fatty acids are catabolised to acetyl coenzyme A (acetyl CoA) by lipolysis and beta-oxidation, glucose and lactate to pyruvate via glycolysis and the action of lactate dehydrogenase, and proteins by numerous routes of amino acid catabolism resulting in pyruvate, tricarboxylic acid (TCA) cy-

cle intermediates, and acetyl CoA. Their iterative oxidation through the TCA cycle results in electron carriage by the reducing equivalents NADH and $FADH_2$ [4]. The electrons are transported to and utilised by the electron transport chain (ETC), composed of respiratory chain complexes (RCC) I-IV in the inner mitochondrial membrane (IMM). The electron flow drives proton translocation by complex I, III, and IV (CI, III, and IV) across the IMM into the intermembrane space, generating a proton electrochemical gradient utilised by complex V (CV, ATP synthase) for the synthesis of ATP from ADP and inorganic phosphate [3]. Together, electron transfer by mitochondrial CI-IV and ATP synthesis by CV is termed oxidative phosphorylation (OXPHOS).

The metabolic functions of mitochondria reach far beyond bioenergetics. Mitochondria form a highly dynamic network, and cellular functions are tightly linked to coordinated cycles of mitochondrial fission and fusion, motility, positioning, and form [5]. They play a critical role in the biosynthesis of cofactors for cellular processes (e.g., nucleotides, fatty acids, cholesterol, amino acids, glucose, haem, and Fe-S cluster proteins) [6], in the balance of redox equivalents, including reactive oxygen species (ROS), and management of potentially damaging metabolic by-products crucial for maintenance of cellular homeostasis (e.g., lactate, ammonia, and hydrogen sulfide) [3]. ROS are generated, sequestered, and interconverted in the mitochondria in response to cellular pressures such as hypoxia and nutrient availability to facilitate cellular adaptation or death [7]. A major source of mitochondrial ROS is electron leak from the ETC, of which mitochondrial CI, and more specifically the N-module, is a primary site of ROS production and thereby prone to extensive oxidative damage [8, 9]. For this reason, surveillance and replacement of the CI N-module is proposed to be critical to the maintenance of CI function [10].

A schematic representation of the aforementioned mitochondrial functions inclusive of OXPHOS, mitochondrial DNA maintenance, expression, and translation, mitochondrial dynamics, homeostasis, and quality controls, and metabolism of substrates cofactors and toxic compounds is displayed in **Fig. 1.1**.

1.1.2 Mitochondrial genetics

Mitochondria contain their own DNA, the mitochondrial DNA (mtDNA), with properties unique from those of the nuclear DNA. The mtDNA is a circular, 16,569 base pair, double-stranded DNA molecule. During the course of evolution, the majority of the ancestral mitochondrial genome was lost or transferred to the nuclear genome, leaving just 37 mtDNA genes (see **Fig. 1.2**). 13 of these mtDNA encoded gene encode subunits of the mitochondrial RCC and 24 encode mature RNA molecules (22 mitochondrial tRNA molecules and two mitochondrial rRNA molecules) essential for mtDNA-specific

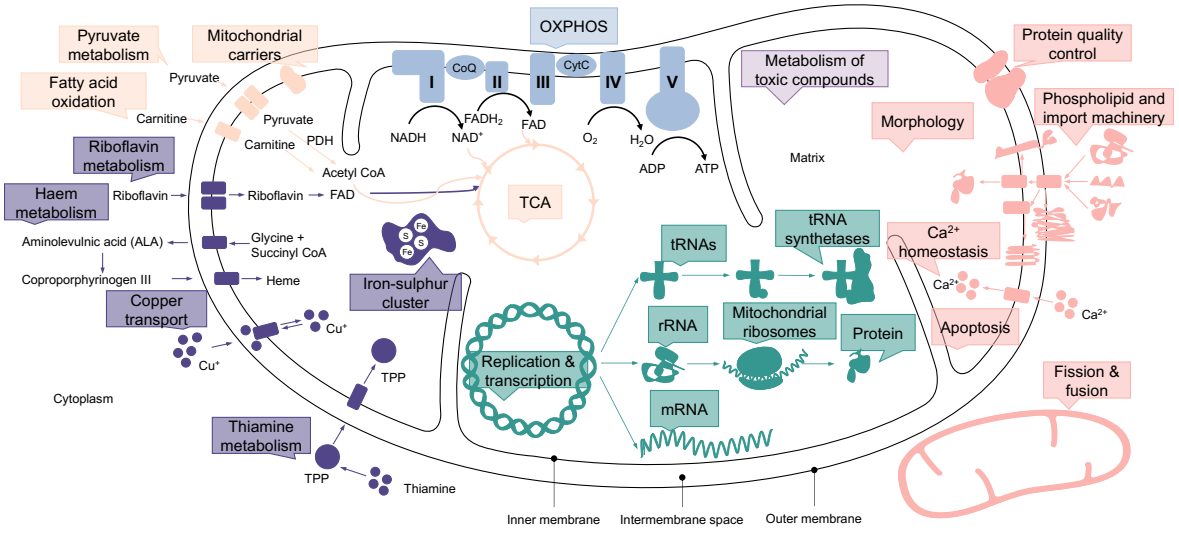


Figure 1.1: Mitochondrial functions.
Key mitochondrial functions separated into the OXPHOS machinery (blue), mitochondrial DNA maintenance, expression, and translation (green), mitochondrial dynamics, homeostasis, and quality controls (pink), and metabolism of substrates (peach), cofactors (deep purple), and toxic compounds (light purple). *Adapted from [11].*

translation of the mtDNA encoded RCC subunits [12]. As the mitochondrial proteome comprises almost 1,150 proteins (MitoCarta3.0, see Chapter 2.2.7), mitochondrial function is reliant upon the nuclear genome hosted genes, encoding the vast majority of mitochondrial proteins (∼99%). These proteins are flanked by mitochondrial targeting sequences and require elaborate protein import machinery to direct them from the cytosol into the mitochondria.

While nuclear DNA is inherited according to Mendelian law, the inheritance of mtDNA is exclusively maternal [13]. Paternal transmission is effectively prevented by dilutional effect (∼100 mtDNA copies in sperm, ∼100,000 mtDNA copies in the oocyte), selective ubiquitination and degradation of paternal mtDNA [14], and the mitochondrial bottleneck, a process reducing and specifying the mtDNA population in the primary oocyte [15]. To date, only a small number of cases of paternal transmission have been reported and remain controversial [16, 17, 18]. These events may be explained by nuclear-mitochondrial DNA segments (NUMTs), paternally transmitted in a Mendelian manner [19].

Eucaryotic cells contain a high copy number (hundreds to thousands) of mtDNA molecules, depending on the cellular energy demand of the cell type or tissue. Their multi-copy nature gives rise to either homoplasmy, where all mtDNA molecules are identical, or heteroplasmy, a blend of wild-type and mutant mtDNA molecules. Heteroplasmy can occur as a consequence of oxidative damage to the mtDNA and high

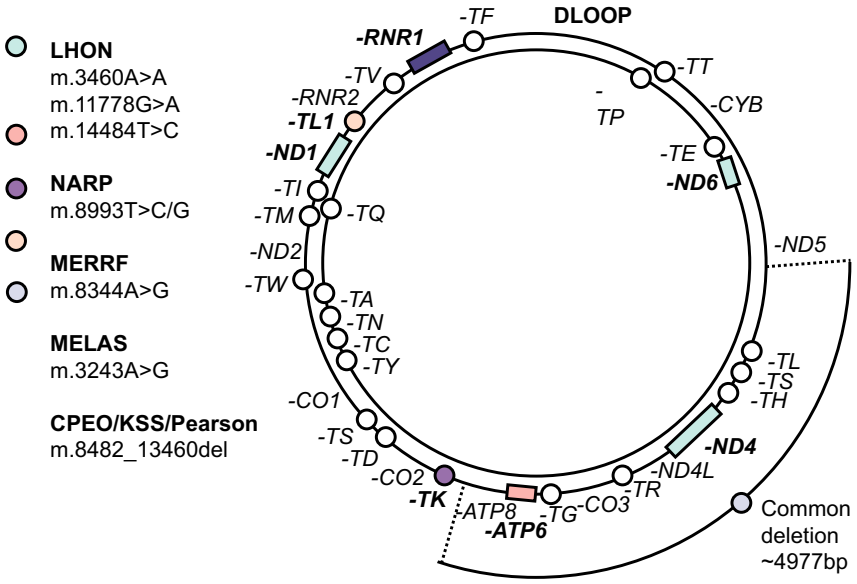


Figure 1.2: mtDNA encoded genes.
mtDNA structure annotated with 37 encoded genes. Reported pathogenic mtDNA variants causative of typical mitochondrial syndromes, including the single large common deletion, are indicated.

levels of mtDNA replication in combination with imperfect repair, allowing mtDNA mutations to arise on a frequent basis [20]. Heteroplasmy level gradually changes with time due to cell proliferation [21] and more dramatically between mother and child as a result of the mitochondrial bottleneck, termed a "heteroplasmy shift" [22]. These unique features become important in the inheritance of pathogenic mtDNA mutations.

1.1.3 Mitochondrial diseases

Mitochondrial diseases result from disruption of mitochondrial function and subsequently impaired energy metabolism. The first mitochondrial disease patient to be clinically defined was in 1959 by Luft et al., with generalized weakness, inability to gain weight despite polyphagia, and excessive perspiration in association with defective OXPHOS function [23]. Since this time, the continual report of mitochondrial disease patients has expanded the associated clinical spectrum and has clouded the water for clinical classification of patients, promoting movement toward a gene agnostic diagnostic approach. The first mitochondrial diseases to be genetically defined in 1988 were Kearns-Sayre syndrome (KSS), due to a large single mtDNA deletion [24, 25], and Leber's hereditary optic neuropathy (LHON) due to a point mutation in *MT-ND4* [26] (see **Fig. 1.2**). These discoveries were shortly followed in 1989 by the first nuclear encoded defect in *PDHA1* [27].

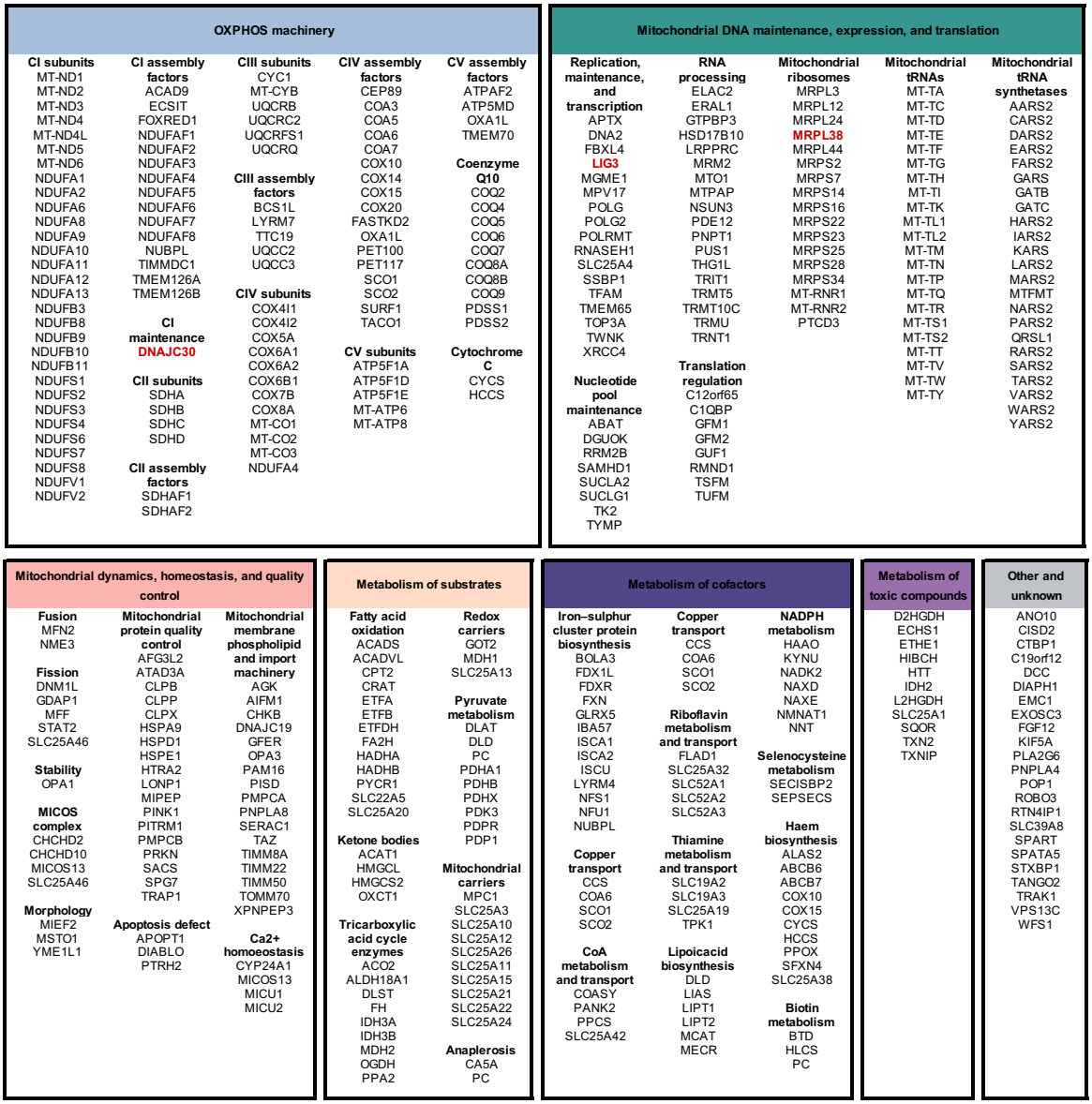


Figure 1.3: Mitochondrial disease genes.
Currently reported mitochondrial disease genes (total number 416) separated by functional category of the encoded gene product. This figures included mitochondrial disease genes discovered as part of the projects in this thesis: *DNAJC30*, *LIG3*, and *MRPL38*, highlighted in red.

To date, over 400 mitochondrial disease genes have been reported [11] (see **Fig. 1.3**). Collectively, they represent ∼10% of disease-associated genes in the OMIM database and are the largest class of inborn error of metabolism (IEM) [28]. The number of disease genes is in constant flux given their continual discovery (see **Fig. 1.4**). Due to the dual genetic origin of the mitochondrial proteome, all modes of inheritance are reported: maternal, autosomal recessive (AR), autosomal dominant (AD), X-linked, and *de novo*. Remarkably, a small number of defects may result in either AR or AD

inheritance due to heterogeneity in the pathogenic consequence of variants (e.g., loss-of-function, gain-of-function). Maternal inheritance is complicated by mtDNA heteroplasmy. A threshold level of heteroplasmy, dependent on the tissue and variant, needs to be reached to result in disease. This is typically reported at 60-80% [29], but can be substantially lower depending on the tissue sampled (such as ∼10% in blood) [30, 31]. In association with certain mtDNA variants, heteroplasmy level can determine the clinical syndrome and is reported to positively correlate with disease severity in the patient, such as for the common *MT-TL1* variant m.3243A>G [32], *MT-TK* variant m.8344A>G [33], and *MT-ATP6* variant m.8993T>G [34, 35, 36].

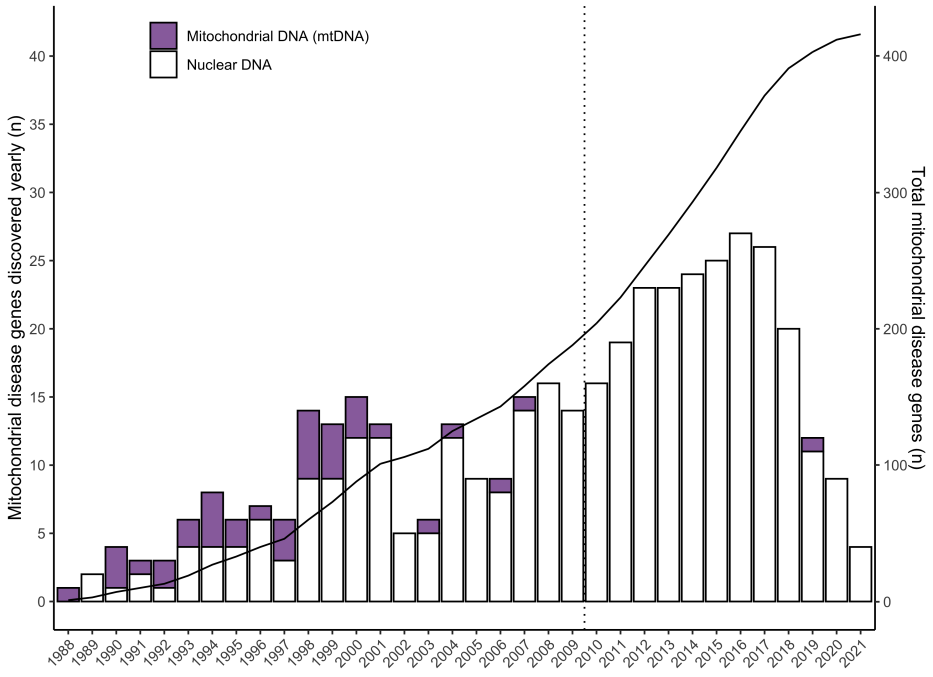


Figure 1.4: Mitochondrial disease gene discovery.
Number of mitochondrial disease genes discovered per year, and cumulatively, since 1988. 2010 marks the advent of the NGS era and an acceleration in the rate of disease gene discovery. *Adapted from [37].*

The overall prevalence of mitochondrial disease is reported as 5-20 per 100,000 [38], stratified into 5-15 per 100,000 for adult-onset disease (majority mtDNA encoded, ∼80%) and 10 per 100,000 for paediatric-onset disease (majority nuclear encoded, ∼70-85%) [39, 40]. By far the most frequent mitochondrial disease reported to date is LHON with a prevalence of 3.22 per 100,000 [41], followed by single large mtDNA deletion syndromes (1.5 per 100,000) [40]. Though a number of mitochondrial syndromes, such as LHON, are readily identifiable by distinctive conforming clinical features (see **Tab. 1**), extensive phenotypic heterogeneity in the majority of mitochondrial diseases means that many patients evade clinical diagnosis. For these reasons, the prevalence

figures given are likely to be an underestimate. Approaches utilising population allele frequencies may provide a clearer picture, estimating a collective lifetime risk of approximately 30 per 100,000 for 249 of the nuclear encoded mitochondrial diseases [42]. However, these approaches are susceptible to overestimation given the possibility of variants leading to *in utero* lethality and variants with incomplete penetrance. Amongst the nuclear encoded mitochondrial disease genes, incomplete penetrance has been reported for a limited number of variants in for example *POLG*. In contrast, incomplete penetrance is frequently reported for mtDNA variants, exemplified by homoplasmic mtDNA variants causing LHON (50% penetrance in males, 10% penetrance in females) [43] (see **Fig. 1.2** and **Tab.1**).

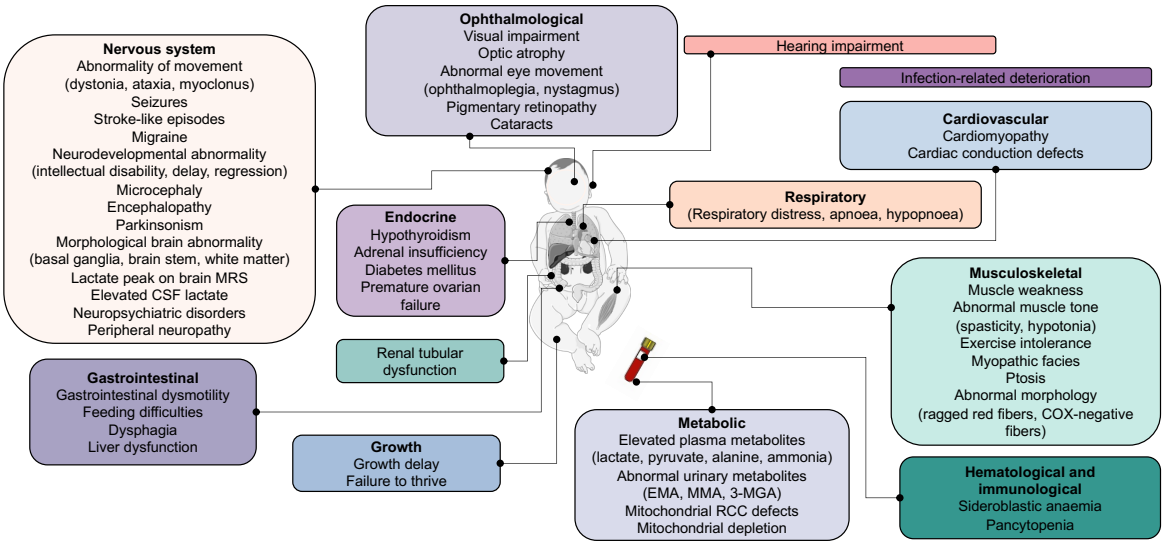


Figure 1.5: Phenotypic manifestations of mitochondrial disease.
Frequent mitochondrial disease-associated phenotypes indicated by organ system. MRS, magnetic resonance spectroscopy; CSF, cerebrospinal fluid; EMA, ethylmalonic aciduria; MMA, methylmalonic aciduria; 3-MGA, 3-methylglutaconic aciduria.

1.2 Clinical and biochemical diagnosis of mitochondrial diseases

1.2.1 Clinical phenotype

Given cellular dependency on maintaining efficient energetic status, genes encoding the mitochondrial proteome are almost ubiquitously expressed across tissues [44]. Subsequently, defects within these genes can present with remarkable phenotypic variability across tissues (see **Fig. 1.5**), with a predilection for the high-energy demanding tissues (e.g., brain, skeletal muscle, heart, liver), and characteristically decompensating

Paediatric onset mitochondrial disease syndromes	Disease gene(s)	Encoding genome(s)
Leigh syndrome	>90 disease genes	mtDNA and nDNA
Alpers-Huttenlocher syndrome (AHS)	*POLG*	nDNA
Childhood myocerebrohepatopathy spectrum (MCHS)	*POLG*	nDNA
Ataxia neuropathy spectrum (ANS)	*POLG*	nDNA
Myoclonic epilepsy myopathy sensory ataxia (MEMSA)	*POLG*	nDNA
Senger syndrome	*AGK*	nDNA
MEGDEL syndrome	*SERAC1*	nDNA
Pearson syndrome	Single large mtDNA deletion	mtDNA

Juvenile or adult onset mitochondrial disease syndromes	Disease gene(s)	Encoding genome
Leber's hereditary optic neuropathy (LHON)	*MT-ND1, -ND4, -ND6*	mtDNA
Kearns-Sayre syndrome (KSS)	Single large mtDNA deletion	mtDNA
Mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS)	*MT-TL1, -TF, -TV, -TQ*	mtDNA
Myoclonic epilepsy with ragged-red fibres (MERRF)	*MT-TK, -TF, -TL1, -TI, -TP*	mtDNA
Neurogenic muscle weakness, ataxia, and retinitis pigmentosa (NARP)	*MT-ATP6*	mtDNA
Chronic progressive external ophthalmoplegia (CPEO)	>15 disease genes	nDNA
Mitochondrial neurogastrointestinal encephalopathy syndrome (MNGIE)	*TYMP, RRM2B, POLG*	nDNA

Table 1: Syndromic manifestations of mitochondrial disease.
Early paediatric, juvenile, and adult-onset syndromes presenting with typical constellations of symptoms listed with their corresponding genetic defect(s).

in times of metabolic stress (e.g., infection).

Given the clinical heterogeneity of mitochondrial disease, reaching a genetic diagnosis on the basis of phenotype is challenging. Patients with a shared genetic diagnosis can present differently, such as single mtDNA mutations spanning adult-onset isolated organ involvement in LHON, to infantile-onset Leigh syndrome, otherwise known as infantile subacute necrotising encephalopathy, and clinically identified by symmetrical lesions in the basal ganglia or the brain stem on magnetic resonance imaging (MRI), elevated serum lactate, and neurodevelopmental abnormalities [45, 46, 47]. Equally, distinct clinical syndromes can be associated with many different disease genes, such as Leigh syndrome with defects reported in over 90 different disease genes (according to The Leigh Map, see Chapter 2.2.7) [48]. In juvenile and adult-onset mitochondrial disease, recognisable syndromes can pinpoint the genetic diagnosis to a single gene, or a handful of genes as depicted in **Fig. 1.2** and **Tab.1** [38]. Examples include LHON, mitochondrial myopathy, encephalopathy, lactic acidosis, stroke-like episodes (MELAS), myoclonic epilepsy with ragged-red fibres (MERRF), neuropathy, ataxia and retinitis pigmentosa (NARP), and mitochondrial neurogastrointestinal encephalopathy syndrome (MNGIE), in addition to the single large mtDNA deletion syndromes, chronic progressive external ophthalmoplegia (CPEO) and KSS. In contrast, in paediatric-onset mitochondrial disease, only the minority of patients present with a distinct clin-

ical syndrome. Limited examples include *POLG*-related syndromes, Senger syndrome, MEGDEL syndrome, and Pearson syndrome.

As many patients present with non-specific symptoms, it is generally suggested to suspect a mitochondrial disease upon impairment of seemingly unrelated tissues (e.g., brain and heart involvement) [49]. To quantify the likelihood of a patient suffering from a mitochondrial disease, clinical criteria have been developed, such as the Mitochondrial Disease Criteria (MDC) described by Morava et al., [50] and Witters et al., [51]. These criteria allocate scores to specific phenotypes to provide the clinician with an indication of mitochondrial disease likelihood, and subsequently aid diagnostic decisions. However, as the phenotypes of high weighting in these criteria have equally high genetic heterogeneity, the specificity of the MDC to identify patients truly suffering from a genetically defined mitochondrial disease is limited. Likewise, clinical scales have been developed to quantify disease severity, such as the Newcastle Paediatric Mitochondrial Disease Scale (NPMDS) [52] and the Newcastle Mitochondrial Disease Adult Scale (NMDAS) [53]. The scores provided by these scales are useful objective measures for use as primary or secondary endpoints in clinical trials to determine disease course and treatment efficacy.

1.2.2 Metabolic and biochemical measures

Metabolic and biochemical measures at the forefront of mitochondrial disease diagnostics are mitochondrial RCC activities, immunohistochemical and histoenzymatic assays of muscle (e.g., in the detection of ragged-red fibres and COX-negative fibres), plasma and urine metabolites resulting from OXPHOS impairment (e.g., lactate, ammonia, and alanine) [54], and more recently, increased serum FGF21 and GDF15 [55, 56]. Though these biomarkers provide a useful indication of mitochondrial disease presence, they are generally subject to low sensitivity and/or specificity. This is exemplified by elevated serum lactate, a biomarker that can be overlooked due to only transient elevation in a primary mitochondrial disease, and falsely prioritised by a non-specific elevation detected in an acutely unwell child or as an artifact in a blood sample taken from a struggling child or due to tourniquet use. Subsequently, to date, there is no single reliable metabolic or biochemical biomarker for the identification of a mitochondrial disease.

While measurement of mitochondrial RCC activity was considered the gold standard for the clinical diagnosis of a mitochondrial disease, a number of drawbacks, in combination with increasing uptake of gene agnostic diagnostic approaches, have steadily reduced their uptake [57]. Mitochondrial RCC activity measurement necessitates an invasive biopsy (typically of muscle or of the disease affected tissue), is ideally per-

formed on fresh specimens (often unfeasible in practice), and is analysed across centres without a universally accepted protocol or reference range for quantification, hindering comparison [58]. Moreover, specificity is limited by the plethora of diseases resulting in secondary OXPHOS dysfunction and preanalytical artifact. Despite these drawbacks, detection of a mitochondrial RCC defect illuminates a biochemical signature of mitochondrial disease [59], and remains useful in the provision of functional evidence of pathogenicity following detection of VUS in a mitochondrial disease gene. To reduce the burden on the patient of an invasive muscle biopsy, mitochondrial RCC defects can be measured reliably in patient-derived fibroblast cell lines, obtained by a minimally invasive skin biopsy. In fibroblast cell lines, most defects measured on muscle biopsy are recapitulated [60, 61], and occasionally, patients with normal RCC enzyme activity in muscle show reduced enzyme activity in fibroblasts [58].

1.3 Genetic diagnosis of mitochondrial diseases

1.3.1 Motivation

Despite the aforementioned hurdles to the diagnosis of a mitochondrial disease, motivation to identify patients and provide a genetic diagnosis remains high. Detection of the disease-causative variant(s) is essential to genetic counselling, in providing accurate recurrence risk estimates to families based on inheritance pattern, and in disease prevention (e.g., by preimplantation genetic diagnostics or mitochondrial transfer) [62]. Where the natural history of the disease is well documented, the diagnosis also provides valuable information on prognosis and stratification of disease risk to determine the need for anticipatory care, such as ECHO surveillance for cardiomyopathy [63]. Moreover, pinpointing the genetic cause provides an opportunity to individually target treatment. Prime examples of treatments used to ameliorate disease include replenishment of a critical cofactor or vitamin (in cofactor- and substrate-metabolism defects), or dietary modification (such as a ketogenic diet in *PDHA1* defect, and valine restriction in *HIBCH* and *ECHS1* defect to prevent the accumulation of toxic metabolites). In total, 63 mitochondrial disease gene defects are considered to have a specific therapeutic option based on expert opinion in the literature [64, 65] (see **Fig. 1.6**), and genetic definition of disease is fast-becoming the prerequisite for clinical trials to establish defect specific treatments and guidelines. Notably, these treatments include idebenone therapy for LHON, a potent antioxidant and electron donor bypassing mitochondrial CI to restore electron flow and respiration. To date, this is the only EMA licenced treatment for a mitochondrial disease [66].

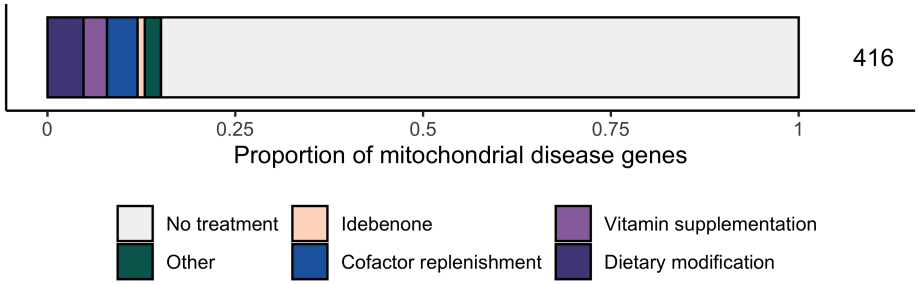


Figure 1.6: Therapeutic options for mitochondrial disease.
Therapeutic options are reported for 63/416 mitochondrial disease genes based on expert opinion in the literature. Idebenone is licensed only for the clinical indication of LHON in association with a confirmed pathogenic variant.

1.3.2 Targeted sequencing

Single gene sequencing

Before the introduction of next generation sequencing (NGS), single gene sequencing was the principal genetic diagnostic approach for mitochondrial disease, whereby deep clinical phenotyping and measurement of mitochondrial RCC activity directed the selection of a single gene(s) for Sanger sequencing [57]. Sanger sequencing was introduced in 1977 [67]. It transformed genetic diagnostics by allowing pathogenic mutations to be identified with single base resolution (unlike the earlier method of linkage analysis using restriction fragment length polymorphism, that identified only the rough genomic location) [68, 69]. In Sanger sequencing, random incorporation of chain-terminating dideoxy nucleotides by DNA polymerase during DNA replication is used to determine DNA sequence (up to 800 nucleotides). This method was essential to the first complete sequencing of the mtDNA in 1981 [70] and the nuclear DNA in 2003 [71]. In total, over 200 mitochondrial disease genes have been identified by single gene sequencing, often in combination with linkage analysis and homozygosity mapping. On average 7.5 mitochondrial disease genes were discovered per year between 1988 and the introduction of NGS in 2010 (see **Fig. 1.4**), with a preponderance in the discovery of pathogenic mtDNA variants in the earlier years due to the earlier definition of this 16 kilobase target region [37].

Single gene sequencing is useful in diagnosing highly recognisable clinical syndromes caused by only a single gene, or small number of genes (e.g., LHON, caused by three mtDNA point mutations in approximately 90% of familial cases, see Chapter 1.2.1 and **Fig. 1.5**). In cases not presenting with a distinct mitochondrial disease syndrome, the overall diagnostic yield of single gene sequencing is reported at approximately 10% [72]. This low diagnostic yield drove the use of high-throughput NGS methods of the

mtDNA and nuclear DNA, in the form of whole mtDNA sequencing, whole exome sequencing (WES), and whole genome sequencing (WGS) for wider angle capture of disease-associated genes.

Whole mtDNA sequencing

NGS of the mtDNA allows detection of variants along the entire mtDNA sequence, within which 95 pathogenic mutations have been reported to date (according to MitoMap, see Chapter 2.2.7), and simultaneous measurement of heteroplasmy level. Whole mtDNA sequencing is often undertaken as an initial step prior to WES or WGS in diagnostic centres specialising in the diagnosis of mitochondrial disease, especially for adult-onset patients where mtDNA aetiology is more frequent. Whole mtDNA sequencing of DNA from easily accessible tissues, such as the blood and urinary epithelial cells, is a minimally invasive starting point. However, mtDNA heteroplasmy varies from tissue to tissue and numerous pathogenic mtDNA variants are restricted to the disease-affected tissue(s). Therefore, a negative result does not exclude an mtDNA mutation. Moreover, preferential selection for wild-type mtDNA over mutant mtDNA overtime in the blood dilutes the diagnostic value of DNA extracted from blood with age in the detection of some mtDNA variants [73]. In adult-onset mitochondrial disease, sampling of the muscle, as a high energy demanding post-mitotic tissue, is often needed, particularly in the detection of single large deletions and mtDNA depletion in CPEO and KSS [74]. In paediatric-onset mitochondrial disease, mtDNA variants are more readily detectable in the blood and urinary epithelial cells [75], negating the need for an invasive biopsy, though skin biopsies are recommended [57].

Panel sequencing

Selection of a NGS panel requires an *a priori* suspicion of a specific disease. In suspected mitochondrial disease, panels are targeted towards known mitochondrial disease genes and candidate genes predicted to be involved in vital mitochondrial functions, ranging from hundreds to thousands of genes. Applying a panel of 300 genes (MitoSure300, see Appendices) in a heterogenous cohort reaches a diagnostic rate of approximately 30%. Depending on panel size diagnostic rates are reported in the literature between 10-30% [76, 77, 78, 79, 80, 81]. NGS panels have the advantage of high sequencing depth (500-1,000 times or higher) and faster analysis time in comparison to WES and WGS. Limitations include a limited shelf-life given the ever growing number of reported disease genes and blinkering to unexpected and potentially treatable genetic diagnoses beyond the targeted regions.

1.3.3 Whole exome sequencing

The diagnostic shortfall of targeted sequencing approaches, caused by the clinical heterogeneity of mitochondrial disease and significant phenotypic overlap with other genetic diseases (e.g., IEM, neurodevelopmental, neurodegenerative, and neuromuscular disease), rendered the need for an untargeted approach. WES probes the coding, exonic, regions, accounting for ∼2% of the 3.2 billion nucleotide genomic sequence. The exonic region contain ∼85% of identified disease-causing variants (according to HGMD and ClinVar, see Chapter 2.2.7). Briefly, in WES, genomic DNA (gDNA) is sheared into small fragments and the exonic regions are selectively captured by hybridization to oligonucleotide baits, pulled down by magnetic beads, amplified by polymerase chain reaction (PCR), and sequenced by massively parallel sequencing [82]. In this way, approximately 97% of protein-coding regions are reliably covered (at least 20 times) with high concordance to Sanger sequencing [83].

The first mitochondrial disease gene to be identified by WES was *ACAD9* in 2010 [84]. Since this time, WES has become the gold standard for disease gene discovery, accelerating the rate of mitochondrial disease gene discovery to approximately 20-25 per year (see **Fig. 1.4**), and resulting in the discovery of over 300 novel Mendelian disease genes per year [85]. In recent years, WES has translated into routine clinical practice, becoming a first-tier genetic diagnostic investigation in suspected mitochondrial disease [57]. WES reaches diagnostic rates of 25-50% across broad genetic indications [86], and of 35-70% in suspected mitochondrial disease [81, 87, 88, 89, 90, 91, 92, 93, 94]. In these mitochondrial disease WES studies, the diagnostic rate was highly dependent on patient selection. Small, deeply characterised, paediatric-onset, biochemically confirmed cohorts achieved the highest rates. Conversely, large, less clearly defined cohorts (arguably those more reflective of everyday clinical practice) report more modest diagnostic rates (see **Fig. 1.7**). Importantly, the mtDNA can be analysed in parallel with the nuclear DNA by WES. mtDNA reads are captured "off-target" with high recall, precision, and comparable estimation of heteroplasmy levels to whole mtDNA NGS [31].

Variant detection and annotation

WES data, in the form of short sequence reads, are stored as a FASTQ file. Sequences are aligned to the reference genome and single nucleotide variants (SNVs) and small insertions and deletions (indels) are called based on differences between the input sequence and reference sequence, stored in a VCF (variant call format) file. The methods used for alignment, SNV and indel variant calling, and copy number variant (CNV) detection in this thesis are detailed in Chapter 2.2.1. The variant calls are filtered for minimal read depth and quality. High quality variants are subsequently

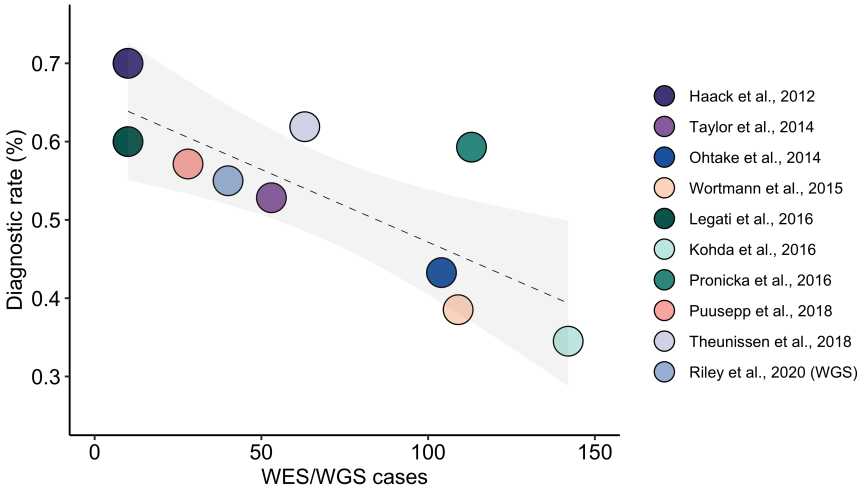


Figure 1.7: Diagnostic rate of WES and WGS in suspected mitochondrial disease. Nine reported studies analysing the use of WES and one of WGS in reaching a genetic diagnosis for patients with suspected mitochondrial disease. The size of the study is depicted in relation to the diagnostic rate. *Adapted from [37].*

annotated with their location within the gene (5'-UTR, exonic, splice site, intronic, 3'-UTR, or intergenic), allele frequency in population sequence databases (gnomAD, see Chapter 2.2.7), functional consequence on the gene product (e.g., synonymous, missense, in-frame indel, frameshift, stop-gain, and stop-loss), and their predicted pathogenicity according to a range of *in silico* prediction tools, such as the Combined Annotation Dependent Depletion (CADD) [95] and Sorting Intolerance from Tolerance (SIFT) [96].

Variant prioritisation

WES has shifted the primary challenge in human genetics from variant discovery to variant interpretation, given the vast number of variants called (specifically VUS). Approximately 50,000 high quality variants are called and annotated per individual. The diagnostic objective, often referred to as searching for the "needle in the haystack" [97], is to identify the one to two variants causative of a patient's monogenic disease by variant prioritisation, as depicted in **Fig. 1.8**. The burden of variant analysis is eased by a number of standard prioritisation steps. First, synonymous variants (excluding direct splice site variants) are removed, given that they do not change the encoded amino acid and are likely benign. This reduces the number of variants to be analysed to approximately 11,500. Second, stringent filtering for variant frequency, using a minor allele frequency (MAF) threshold, is applied. In rare genetic disease a MAF of $\leq$0.1% is routinely applied and results in $\sim$250 variants for analysis. Selection of the MAF threshold must take into account the prevalence of the disease. All but one

nuclear encoded mitochondrial disease associated pathogenic variant are associated with a MAF $\leq$0.1%. The one exception is a European founder mutation with a MAF of 0.11% in *MTFMT* (c.626C>T, NM_139242.4) [98, 99]. Completely penetrant pathogenic variants with an MAF >0.1% would result in disease prevalence higher than expected for an individual mitochondrial disease. This filter, however, may need to be relaxed to detect pathogenic variants with incomplete penetrance. Our ability to assess population allele frequencies has been made possible by the development of resources aggregating and harmonizing exome and genome sequencing data from population genetic studies, such as gnomAD, containing 125,748 whole exome sequences and 15,708 whole genome sequences from unrelated individuals [100].

In the next step, variants are filtered based on the expected pattern of inheritance of disease. The vast majority of mitochondrial diseases, especially those of paediatric-onset, are inherited in an AR manner. Limiting variant analysis to homozygous or potentially compound heterozygous variants further reduces the number of variants for analysis to between five and 25. When trio-based WES is available (index, mother, and father) variants can be phased to immediately confirm their biallelic nature. Moreover, *de novo* and AD variants can be prioritised. Though *de novo* and AD variants are expected to only account for a small number of mitochondrial disease diagnoses [101, 102, 103], they are a frequent finding in neurodevelopmental diseases, arising in patients with extensively overlapping phenotypes to mitochondrial disease. Trio-sequencing reduces the number of potentially *de novo* variants ten-fold [104], resulting in approximately 1-2 variants for closer inspection. Considering mtDNA variant detection, there is one study in the literature analysing the mtDNA from WES of DNA extracted from blood in over 2,000 individuals of broad inclusion to a genetic diagnostic centre. The study detected a confirmed pathogenic mtDNA variant (according to MitoMap, see Chapter 2.2.7) in approximately 2% of patients, and a VUS in a further 0.7% of patients [31], of which approximately 20% arose *de novo* [105]. This study was however, enriched for suspected mitochondrial disease patients, and thereby, ascertainment bias could be a confounder. Overall, analysis of the mtDNA from WES can be expected to only add a limited number of additional variants for consideration when called in parallel with nuclear variants.

Known pathogenic variants can be identified by their presence in curated disease variant databases such as ClinVar, DECIPHER, HGMD, and MitoMap (see Chapter 2.2.7), leading to an immediate diagnosis. To date, ClinVar reports 147,071 pathogenic (P) or likely pathogenic (LP) variants (see **Fig. 1.9**). Beyond these conditions, novel variants in known disease genes, or novel variants in novel disease genes require careful consideration. For the analysis of novel variants in known disease genes, a number of recommendations for variant classification are provided by the American College of

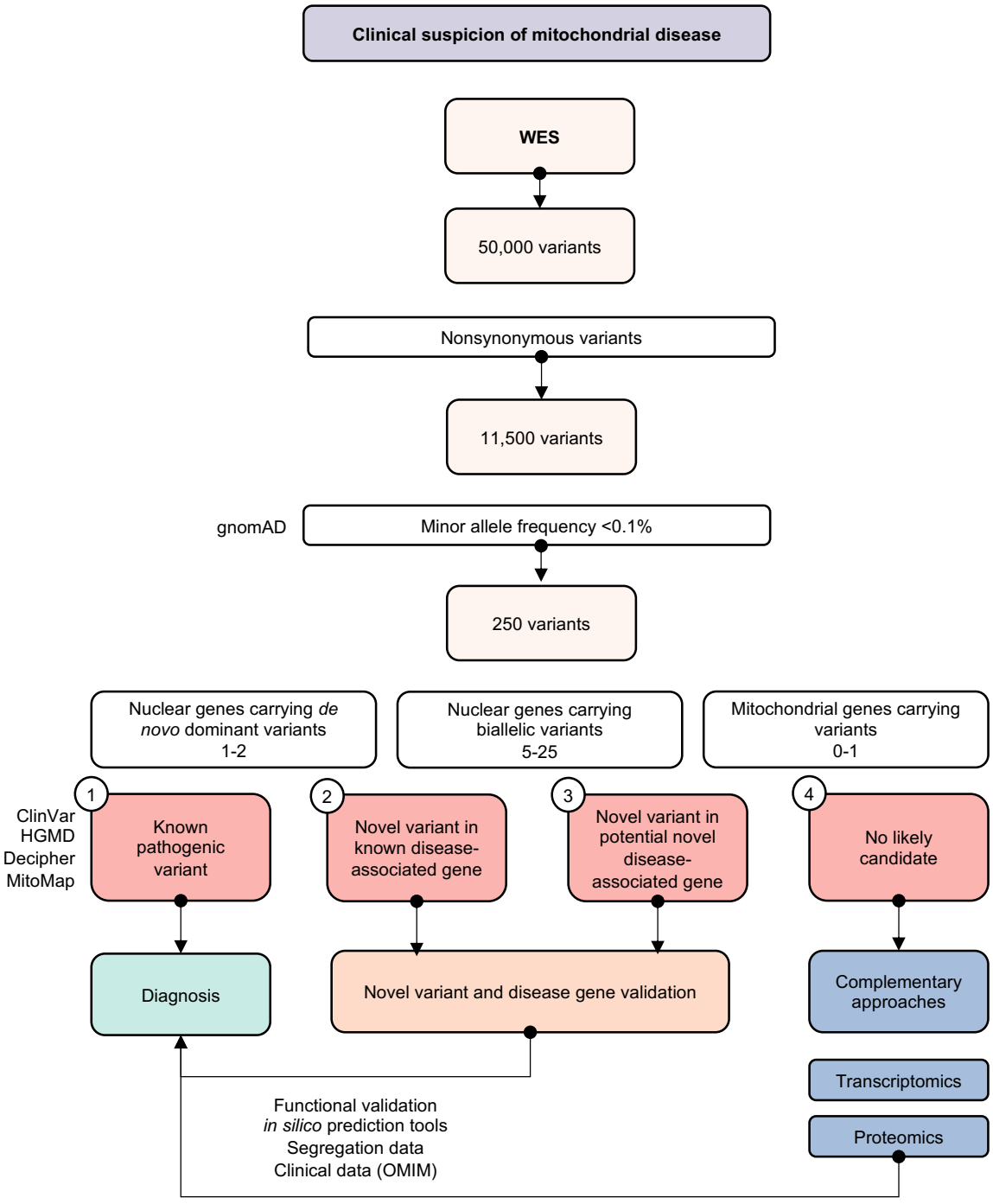


Figure 1.8: WES analysis workflow in suspected mitochondrial disease. Variants are filtered according to a number of criteria for variant function, frequency, and expected mode of inheritance. In the case of novel variant(s) or disease gene detection, further evidence is required to determine pathogenicity. Numbers 1-4 mark the diagnostic objectives addressed in study 1, 2, 3, and 4 of this thesis, respectively.

Medicine Genetics and Genomics (ACMG) and the Association for Molecular Pathology, here referred to as the ACMG criteria. This classification implements a number of criteria based on population, computational, functional, segregation, and allelic data, to position a variant within a five tier classification (benign, likely benign, uncertain significance, likely pathogenic, and pathogenic) [106] (see **Tab. 2**). The analysis of variants in this rigorous manner can be eased by computational algorithms such as InterVar (see Chapter 2.2.1) [107]. For the analysis of novel variants in potential novel mitochondrial disease genes, prioritisation can first query the mitochondrial localisation of the encoded protein. The MitoCarta database (MitoCarta3.0, see Chapter 2.2.7) reports 1,136 proteins with a high probability to localise to the mitochondria, of which to date just over 400 are known to be disease causing (see **Fig. 1.3**). The ∼750 genes encoding the remaining mitochondrial proteins, so far not associated with disease, leaves ample room for further novel disease gene discovery in the future. Evidence for pathogenicity is gathered by functional validation, *in silico* prediction, confirmation of segregation, and similarity between the patient's phenotype and the phenotype reported in the literature or in curated databases (such as provided by Orphanet, DECIPHER, and OMIM, see Chapter 2.2.7). With regards to phenotype similarity, WES is increasingly identifying novel genotype-phenotype associations in mitochondrial disease, by the detection of patients presenting atypically for a disease. This has led to a continuous flow of studies expanding the clinical phenotype of reported disease genes (e.g., most recently for *MT-ATP6*) [34, 35, 36]. Candidate disease genes can be shared via collaborative platforms such as GeneMatcher [108] and Matchmaker Exchange [109], and via mitochondrial disease networks (e.g., mitoNET in Germany, and GENOMIT across Europe and Asia) to collect patients with potentially pathogenic variants in the same gene across research centres (see Chapter 2.2.7). If no candidate variants are identified, complementary approaches such as RNA-sequencing (RNA-seq) and quantitative proteomics can be considered for the detection and simultaneous functional validation of variants (see Chapter 1.3.4). WES data should also be revisited periodically to detected pathogenic variants in known disease genes and disease genes only reported to be disease-associated after the initial WES analysis. WES reanalysis has been demonstrated to provide a genetic diagnosis to up to 10-30% of unsolved cases just 1-3 years after initial analysis [110, 111, 112, 113, 114, 115, 116].

Variant validation

Variant validation can take many forms and is essential to allocate pathogenicity to VUS. These variants are typically missense, near-splice, and in-frame indel variants without predicted protein-truncating (loss-of-function) effect, associated with nonsense, splice, and frameshift variants (see **Fig. 1.9**). Splice and near-splice variants are likely to affect RNA splice pattern, and can be analysed by RT-PCR or RNA-seq

	ACMG criteria
PVS1	Null variant in a gene where loss of function is a known mechanism
PS1	Same amino acid change as a pathogenic variant
PS2	Confirmed *de novo* variant in a patient with no family history
PS3	Well established *in-vitro* or *in-vivo* function study
PS4	Prevalence in affected individuals is significantly higher than in controls
PM1	Located in a mutational hot-spot or functional domain
PM2	Absent from controls in the 1000 Genomes Project, gnomAD, and ExAC
PM3	In *trans* with a pathogenic variant if autosomal recessive
PM4	Protein length changes as a result of the variant
PM5	Missense change in an amino acid where a different missense variant is pathogenic
PM6	Assumed *de novo* variant without confirmation in the parents
PP1	Segregation of the variant with multiple affected family members
PP2	Missense variant in a gene with a low rate of benign missense variation
PP3	Multiple lines of computational evidence
PP4	Patient's phenotype or family history highly specific for the disease
PP5	Reputable source reports pathogenicity
BP1	Missense variant in a gene with primarily truncating variants
BP2	Observed in *cis* with a pathogenic variant if autosomal recessive
BP3	In-frame deletions/insertions in a repetitive region
BP4	Multiple lines of computational evidence indicate no impact of the variant
BP5	Variant found in cases with an alternative molecular basis
BP6	Reputable source recently reports the variant as benign
BP7	A synonymous variant where no splice aberration is predicted
BS1	Allele frequency greater than expected for the disease
BS2	Observed in healthy adults for autosomal recessive (homozygous), autosomal dominant, and X-linked dominant
BS3	Established functional studies show no damaging effect
BS4	Lack of segregation in affected family members
BA1	Allele frequency >5% in 1000 Genomes Project, gnomAD, and ExAC

Table 2: ACMG criteria for variant classification.
Criteria with a P prefix (red-peach) indicate the variant to be pathogenic (PV, very strong; PM, strong, PS, supporting indication, respectively). Criteria with a B prefix (blue-green) indicate the variant to be benign (BP, supporting; BS, strong; BA, stand-alone indication, respectively).

	Total	B/LB	US	LP/P
CADD>30	74,326 (9%)	1,265 (<1%)	20,160 (5%)	52,785 (36%)
CADD>25	172,810 (20%)	6,344 (2%)	88,008 (21%)	78,248 (53%)
CADD>20	309,294 (36%)	21,945 (8%)	194,036 (47%)	92,976 (63%)
Present in gnomAD	394,279 (46%)	181,483 (62%)	190,795 (46%)	217,39 (15%)
ClinVar submitters				Single 79,105 (54%) Multiple 19,986 (14%) Expert panel 7,552 (5%) Submitted as part of haplogroup or genotype19,986 (27%)

Table 3: ClinVar reported variants.
Number of ClinVar reported variants according to CADD score thresholds, presence of the variant in the gnomAD database, and number of ClinVar submitters. Enrichment for more deleterious (higher) CADD *in silico* prediction is demonstrated in LP/P variants. Variant presence in gnomAD provides an indication of variant MAF (absence from gnomAD indicates a MAF$<3.5 \times 10^{-6}$). The vast majority of LP/P variants are reported by a single submitter only. B/LB, benign/likely benign; US, uncertain significance; LP/P, likely pathogenic/pathogenic.

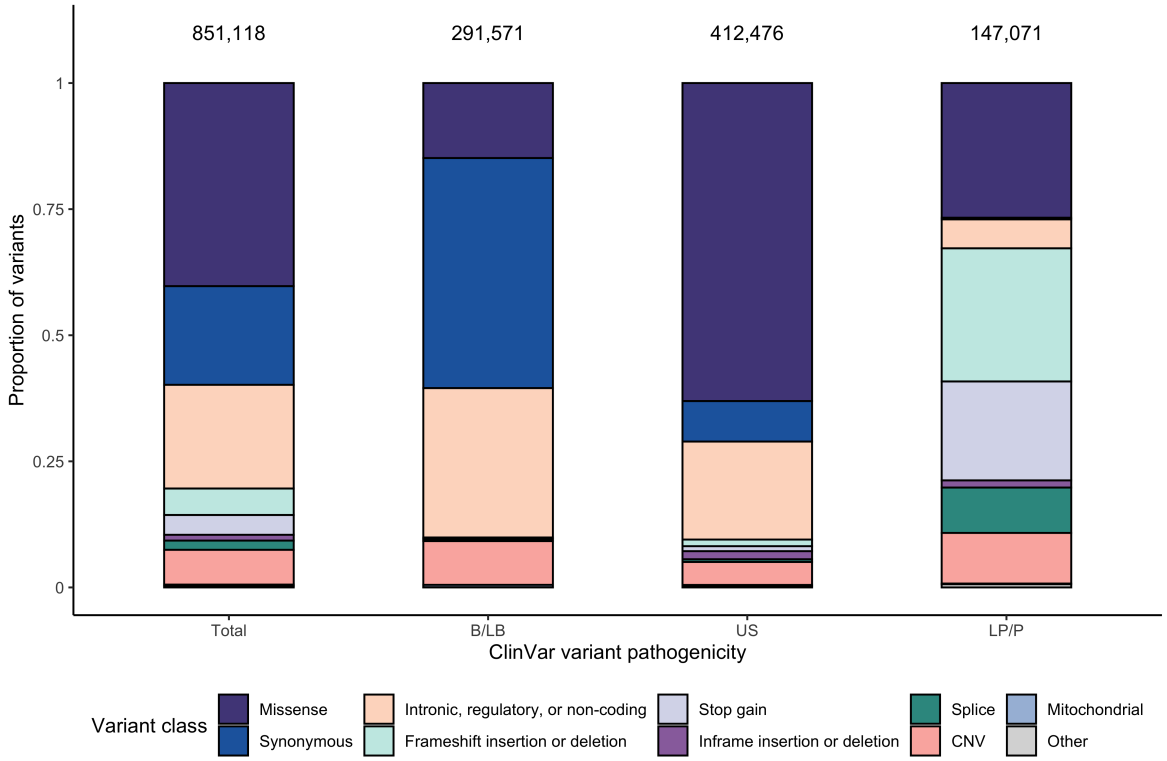


Figure 1.9: ClinVar reported variants.
Number and proportion of ClinVar reported variants according to variant pathogenicity and function, respectively. B/LB, benign/likely benign; US, uncertain significance; LP/P, likely pathogenic/pathogenic; CNV, copy number variant.

[117]. Missense variants are more challenging to interpret as they can affect protein stability or protein function. Missense (and other) variants significantly altering the protein structure and stability, and thereby leading to degradation, can be analysed by immunoblotting or quantitative proteomics [117, 118]. Missense variants in functional domains that leave the protein intact, cannot be interpreted in this way, however can still lead to dysfunction. Mitochondrial protein dysfunction may be detectable as defects in enzymatic activity (such as decreased mitochondrial RCC activity) [59], by decreased of oxygen consumption rate (as a read out of mitochondrial respiration) [119], by measurement of disrupted metabolite level by metabolomics [117, 120], and by enzyme specific assays, amongst others. BN-PAGE and complexome analyses determine mitochondrial complex assembly and abundance, and can be utilised to detect impaired assembly of the mitochondrial RCC and supercomplexes and complex degradation [121]. Rescue of these functional abnormalities by over-expression of a wild-type copy of the defective gene provides strong evidence for causality [122]. The vast majority of these functional assays are united by the need for patient-derived biomaterial, such as a tissue biopsy or cell line (see **Tab. 4**). Where no such material is available, cellular models can be generated (e.g., a knock-out model by CRISPR-Cas)

[123] and model organisms can be employed, exemplified by the yeast model given that >200 mitochondrial disease associated genes have a yeast ortholog [37, 124, 125].

Functional assay	Mitochondrial disease gene capture	Patient-derived material
Transcriptomics	~85% of mitochondrial disease genes (10-15,000 expressed genes detected)	Blood, tissue biopsy, or cell line
Proteomics	~65% of mitochondrial disease genes (5-8,000 expressed proteins detected)	Blood, tissue biopsy, or cell line
Metabolomics	>80 cofactor and substrate metabolism genes	Blood, urine, tissue biopsy, or cell line
BN-PAGE and complexome	>30 OXPHOS assembly factors >50 OXPHOS subunits	Tissue biopsy or cell line
Mitochondrial RCC activity	>50 OXPHOS subunits	Tissue biopsy or cell line
Oxygen consumption rate	>400 genes involved in OXPHOS function	Tissue biopsy or cell line
Yeast model	>200 mitochondrial disease genes with yeast orthologs	No requirement

Table 4: Functional validation assays for variants in mitochondrial disease genes. Assays for capturing the functional consequence of VUS in a large number of mitochondrial disease genes and their corresponding requirement for patient-derived bio-material.

1.3.4 Approach to inconclusive whole exome sequencing

Whole genome sequencing

Variants eluding detection by WES include mosaic variants (requiring higher sequencing depth), genomic alterations such as large deletions and insertions, structural variants including chromosomal rearrangements (e.g., inversion, translocation, deletion, and duplication), repeat expansions, as well as deep intronic and regulatory variants. This shortfall may be overcome WGS, particularly long-read sequencing. WGS captures all regions of the 3.2 billion nucleotide genome, within which approximately four million variants are called. Evidence suggests that as many as 30% of non-coding variants impact gene expression [126]. Therefore, though non-coding variants are currently only reported to be responsible for ∼6% of pathogenic variants (see **Fig. 1.9**), this figure will undoubtedly increase in the future as genetic diagnostic focus shifts from the coding to the non-coding regions.

So far, the increase in diagnostic rate by WGS in comparison to WES has been modest, reported at ∼2% across diverse genetic indications [86]. A single study of WGS in suspected mitochondrial disease reported a diagnostic rate of 55%, with all diagnoses made within the coding region (e.g., missense, frameshift, start-loss, nonsense, and mitochondrial) and thereby also detectable by WES [127]. The interpretation of detected intergenic, deep intronic, and regulatory variants will require integration of functional evidence, given the uncertain consequence of these variants. Due to their vast numbers, this evidence is most efficiently provided by high-throughput genome-

wide "multi-omic" functional validation, by adding a simultaneous functional readout of variant consequence on the gene product (e.g., RNA and protein) [128].

RNA-sequencing

RNA-sequencing (RNA-seq) has proven valuable in identifying the genetic diagnosis in up to 35% of inconclusive WES cases across diverse indications [117, 129, 130, 131], and 10-15% in suspected mitochondrial disease [117]. RNA-seq determines coding and non-coding variant consequence on the abundance and form of the RNA transcript. The detection of three aberrant events aids in variant interpretation, (i) aberrant expression level, (ii) aberrant splicing, and (iii) mono-allelic expression (MAE) of rare variants. Dedicated pipelines for these analyses have now been established [132, 133]. RNA-seq is thereby useful in determining the pathogenicity of splice and near-splice variants and can bring to light the pathogenic nature of presumably benign variants (e.g., regulatory, deep intronic, and synonymous variants) which may lead to splice aberration and nonsense mediated decay (NMD). The principal considerations in RNA-seq are tissue selection and data normalisation. First, the selection of tissue must consider differential expression levels and splice patterns across tissues. Determining the suitability of a certain tissue for disease gene detection is aided by tools such as PAGE (Panel Analysis of Gene Expression), GTEx, the Expression Atlas, and MAJIQ-CAT [131, 134, 135, 136]. Second, normalisation of the data is essential for detection of aberrant disease-causing events, for which machine learning algorithms have been developed to remove known and unknown confounders from the data prior to outlier calling [137].

RNA-seq analysis of patient-derived fibroblast cell lines is reported to detect 10,000-15,000 gene transcripts [117, 138, 118] reliably covering up to 90% of mitochondrial, and 75% of OMIM disease genes [138] (see **Tab. 4**), thereby providing comprehensive coverage of disease-associated genes.

Quantitative proteomics

Quantitative proteomics can evaluate the consequence of diminished RNA expression on the protein level and provide additional insight into the consequence of protein destabilising missense variants, a variant with a functional consequence intractable to detection by RNA-seq. Missense variants are the most frequent class of disease-causing variant in mitochondrial disease and are the most frequent ClinVar reported P/LP variant in Mendelian disease (∼27%). Moreover, they account for approximately 50% of ClinVar reported VUS (see **Fig. 1.9**). An estimated 40% of missense variants lead to protein destabilisation [139], theoretically detectable by quantitative proteomics if the protein of interest is covered. Therefore, the functional interpretation of missense VUS

by proteomics is valuable in diagnostics. Moreover, approximately 20% of mitochondrial, and 45% of OMIM disease genes encode proteins forming part of a protein complex, according to the >2,350 protein complexes listed in CORUM (see Chapter 2.2.7). A defect in one protein of a complex can lead instability of the entire complex and lead to a reduction in the abundance of protein interaction partners as a downstream consequence [138]. In this way, the proteomics assay can not only provide evidence of the protein destabilising nature of a variant of interest, but can provide evidence of downstream functional consequence on the rest of the protein complex (e.g., a defect in a mitochondrial ribosomal protein leading to reduced levels of other proteins in the mitoribosomal complex and subsequent reduction in the synthesis of OXPHOS subunits encoded by the mtDNA [140, 141]), to elucidate underlying disease mechanism [142], and to provide cumulative functional evidence to WES/RNA-seq candidate variants (e.g., a mitochondrial RCC assembly factor defect leading to an incompletely assembled complex that is subsequently degraded [117]).

Proteomic analysis of patient-derived fibroblast cell lines is reported to detect 5,000-8,000 proteins and cover up to 65% of mitochondrial, and 50% of OMIM disease genes [138, 118] (see **Tab. 4**). While proteomics has been used to provide additional functional evidence of pathogenicity in the diagnosis of mitochondrial disease [117, 140, 141, 142], the utility of the systematic application of quantitative proteomics, and integration with other omics such as RNA-seq into a diagnostic pipeline is yet to be analysed in the literature (addressed in Study 4 of this thesis).

Due to the genome-wide nature of "multi-omic" analyses, they may in the future foreseeably replace single gene product studies (e.g., RT-PCR and immunoblotting, for RNA and protein, respectively). The "multi-omic" approaches are optimally used in a complementary manner to achieve the highest level of sensitivity (i.e., disease gene coverage) and to provide independent validation for one another to confidently designate a VUS as pathogenic in order to direct genetic counselling.

1.4 Objectives of this thesis

This thesis concerns the genetic diagnosis of clinically suspected mitochondrial disease patients. Focus is placed on the application of WES and the description of challenging diagnoses due to variants evading detection by routinely applied diagnostic strategies, given features arguing against pathogenicity. The thesis can be formulated into four studies with the following central objectives:

1. To analyse the diagnostic power of WES in >2,000 paediatric suspected mitochondrial disease patients, genetically and clinically characterise the study population by integration of HPO encoded phenotype descriptors, and explore the value of functional data integration from patient-derived bio-material.

2. To validate novel variants in a known disease gene, *FDXR*, by employing a yeast model as a valuable alternative in the absence of patient-derived bio-material.

3. To validate a novel mitochondrial disease gene, *DNAJC30*, and provide insight in a novel biological pathway by characterisation of the disease pathomechanism in patient-derived cell lines and a cellular model.

4. To integrate WES, HPO encoded phenotype descriptors, RNA-seq, and quantitative proteomic data to discover and simultaneously validate the genetic diagnosis in inconclusive WES/WGS cases.

2 | Materials and methods

This chapter outlines the materials, experimental methods, and data analyses utilised to detect and validate disease-associated variants, and to unravel a novel disease pathomechanism.[1] Patient-derived bio-materials (blood and fibroblast cell lines) were essential for the majority of the methods. When these bio-materials were not available or a secondary method of functional validation was required, the model organism *S. cerevisiae* was utilised or a cellular model was generated, respectively.

2.1 Materials

2.1.1 Nucleic acids

DNA

Genomic DNA extracted from blood and/or fibroblast cell lines of patients and healthy controls was utilised for sequencing. This DNA collection belongs to the Institute of

[1] All methods presented in this chapter were personally performed unless otherwise indicated.

Human Genetics (Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany). Informed consent was obtained for diagnostic or research purposes. Sequencing data and human phenotype ontology (HPO) terms (one to 40 per patient) were entered into a local database (The Exome Variant Annotation Database, EVAdb, see Chapter 2.2.7), containing only pseudonymised data.

Oligonucleotides

Oligonucleotide primers were designed for use in Sanger sequencing, site-directed mutagenesis, and lentiviral transduction of human and plasmid DNA, and were synthesised by Metabion (Martinsried, Germany). Listed below are the utilised primers by indication:

Sanger sequencing

	Direction	Sequence
DNAJC30 Part A	F	5'-CAACCGACTCCTCTCATTGG-3'
DNAJC30 Part A	R	5'-GTGGGTCCGAGAGGTGG-3'
DNAJC30 Part B	F	5'-GCCTACGTGGTGCTGGG-3'
DNAJC30 Part B	R	5'-CAAGGGTTCAGAGGCAGG-3'
BFG-1 plasmid	F	5'-CAGATCATCAAGGAA-3'
BFG-1 plasmid	R	5'-AGCACCACCACCAGT-3'

Site-directed mutagenesis

	Direction	Sequence
FDXR c.35C>G	F	5'-GGGCTGGTGGGCGTGGCCTCGGACCCGGCTGCCTCCCGC-3'
FDXR c.35C>G	R	5'-GCCACGCCCACCAGCCCCACCAGCGCCAGCAGCGCGAAGCCAT-3'
FDXR c.325G>A	F	5'-GCCTTCTGGAGCAACGTGGAGGTGGGCAGGGACGTGACGGTGC-3'
FDXR c.325G>A	R	5'-CCACGTTGCTCCAGAAGGCACAGCGGCCAGAATGGGCCGTCT-3'
FDXR c.332T>C	F	5'-TGTGCCTTCCGGGGCAACGTGGAGGTGGGCAGGGACGTGACGG-3'
FDXR c.332T>C	R	5'-CGTTGCCCCGGAAGGCACAGCGGCCAGAATGGGCCGTCTGGG-3'
FDXR c.368G>A	F	5'-GGAGCTGCAGGAGGCCTACCACGCTGTGGTGCTGAGCTAC-3'
FDXR c.368G>A	R	5'-AGGCCTCCTGCAGCTCCGGCACCGTCACGTCCCTGCC-3'
FDXR c.576delTCTGGACGTGGC	F	5'-GAACGTGGCCCGCATCCTACTGACCCCACCTGAGCACCTGG-3'
FDXR c.576delTCTGGACGTGGC	R	5'-AGGATGCGGGCCACGTTCCCCTGCCCCAGAATCACGGCTGT-3'
FDXR c.632C>T	F	5'-CGAAGGCAGTCCTGGGTGTACTGAGGCAGAGTCGAGTGAAGACAGTG-3'
FDXR c.632C>T	R	5'-ACACCCAGGACTGCCTTCGTGATGTCCGTTCTCTCCAGGTGCTCAG-3'
FDXR c.683G>T	F	5'-AGTGGGCCTGCGTGGACCCCTGCAAGTGGCCTTCACCATTAAG-3'
FDXR c.683G>T	R	5'-GTCCACGCAGGCCCACTAGCCACACTGTCTTCACTCGACTC-3'
FDXR c.724C>T	F	5'-AAGGAGCTTTGGGAGATGATTCAGTTACCGGGAGCCCGG-3'
FDXR c.724C>T	R	5'-TCATCTCCCAAAGCTCCTTAATGGTGAAGGCCACTTGCAGG-3'
FDXR c.1058G>A	F	5'-ACCTCCCTTATGGGCTGGTGCTCAGCAGCATTGGGTATAAGAGCC-3'
FDXR c.1058G>A	R	5'-ACCAGCCCATAAGGGAGGTGAAGGTACAGAGGGCACCCGTGACG-3'
FDXR c.1156C>T	F	5'-GTGGAGGGCTGGGTTATGGATGTGCCAGGCCTCTACTGCAGCG-3'
FDXR c.1156C>T	R	5'-CCATAACCCAGCCCTCCACATTGGGGATGACCCCAAGCTTGGAGTCA-3'
FDXR c.1343G>A	F	5'-TCAGCAGCCAAGGGGTCCGGCCAGTCTCTTTCTCAGACTGGGAGAAGCT-3'
FDXR c.1343G>A	R	5'-CGGACCCCTTGGCTGCTGACGTCCCGGACCTACCGACGCATCG-3'

Lentiviral transduction

	Direction	Sequence
DNAJC30 Kozak + ORF	F	5'-CACGATGGCAGCCATGCGCTGGC-3'
DNAJC30 Kozak + ORF	R	5'-TTAAATATAAAAGCCGATGATGATGAA-3'

The melting temperature of all primers was determined by the NEB T_m calculator (see Chapter 2.2.7). The forward (F) and reverse (R) primer pairs were designed to have a maximum difference in melting temperatures of 5°C to allow the selection of an appropriate annealing temperature. The specificity of the primer pair for the DNA segment of interest was determined by the UCSC In-Silico PCR tool (see Chapter 2.2.7) to ensure only the region of interest was amplified.

2.1.2 Cells

Primary patient-derived fibroblast cell lines were utilised in functional validation studies of variant pathogenicity and in studies of disease pathomechanism. These cell lines derived from fresh skin biopsies and were established in culture in the Institute of Human Genetics (Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany) and the Institute of Human Genetics or Neurogenomics (Helmholtz Zentrum München, Neuherberg, Germany). Informed consent was obtained for diagnostic and research purposes. Normal human dermal fibroblasts (NHDF) from neonatal tissue were utilised as a control cell line and were purchased commercially (Lonza, Basel, Switzerland). Human embryonic kidney (HEK) 293FT cells, purchased commercially (Thermo Fisher Scientific, Waltham, USA), were utilised in the generation of a knock-out (KO) cell line and for viral production in functional complementation by lentiviral transduction. One Shot Stbl3 chemically competent *E. coli* were utilised for the amplification of plasmid DNA and were purchased from Thermo Fisher Scientific (Waltham, USA). In Study 2, validating VUS in *FDXR*, For the functional validation of VUS in the mitcohondrial disease gene *FDXR*, a yeast *Arh*-null mutant strain, shared by Paul et al., [124] (Institut IMAGINE, Paris, France), was employed. This strain was transformed with either a BFG-1 *S. cerevisiae* expression plasmid with a wild-type human *FDXR* insert or an empty BFG-1 *S. cerevisiae* expression plasmid.

2.1.3 Chemicals and solutions

Chemicals and solutions were obtained from Sigma-Aldrich (St. Louis, USA) or Merck (Darmstadt, Germany), unless otherwise stated in the methods (see Chapter 2.2).

2.1.4 Antibodies

Antibodies utilised in immunoblotting for the analysis of protein abundance in patient-derived and control fibroblast cell lines, were as follows:

Antibody	Manufacturer	Number
FDXR	Abcam (Cambridge, UK)	ab204310
Anti-β actin	Abcam (Cambridge, UK)	ab8227
Anti-rabbit conjugated	Jackson Immuno Research Laboratories (West Grove, PA, USA)	111-036-045

2.2 Methods

2.2.1 DNA and RNA analysis

These methods were undertaken to determine human and/or plasmid DNA and/or RNA-sequence.

Polymerase chain reaction

Target DNA sequences were amplified by polymerase chain reaction (PCR) using the Qiagen Taq DNA Polymerase Kit (Qiagen, Hilden, Germany) and a PeqStar thermal cycler (PeqLab Biotechnology, Erlangen, Germany) following standard procedures.

The PCR reactions were performed as follows:

	Volume for 10 µL reaction	Final concentration
HPLC water	4.56 µL	
DNA (50 ng/µL)	1 µL	5 ng/µL
Taq DNA Polymerase (5 U/µL)	0.04 µL	0.2 U
10x PCR Buffer	1 µL	1 x
2 mM dNTPs	1 µL	0.2 mM
Q solution	2 µL	1 x
10 µM Primer (F)	0.2 µL	0.2 µM
10 µM Primer (R)	0.2 µL	0.2 µM

The amplification conditions were as follows:

	Temperature (°C)	Time
Heat lid	110	
Denature	95	10 min
Start cycle (35 cycles)		
Denature	95	30 sec
Anneal	Primer $T_m - 5$	30 sec
Extend	72	1 min per kilobase
Close cycle		
Extend	72	10 min
Cool down	20	1 min

The resultant PCR product was subjected to agarose gel (1.5%) electrophoresis (100 V, 30 min) for the assessment of product quality and size.

Sanger sequencing

Following PCR amplification of the target DNA sequence 7.5 µL PCR product was purified using the MultiScreen PCR 96 Filter Plate (Millipore, Merck KGaA, Darmstadt, Germany) in accordance with the manufacturer's protocol. The purified PCR product was used for subsequent cycle sequencing using the ABI BigDye Terminator v.3.1 Cycle Sequencing kit (Life Technologies, Carlsbad, USA).

The sequencing reaction was performed as follows:

	Volume for 5 ul reaction	Final concentration
BigDye Terminator v.3.1 Ready Reaction mix	0.5 µL	
BigDye Terminator 5x Sequencing Buffer	1.5 µL	1.5 x
10 µM primer (F or R)	1 µL	2 mM
Purified PCR product	1 µL	
HPLC water	1 µL	

The sequencing programme was performed as follows:

	Temperature (°C)	Time
Heat lid	110	
Denature	96	1 min
Start cycle (25 cycles)		
Denature	96	10 sec
Anneal	50	5 sec
Extend	60	1 min 30 sec
Close cycle		
Cool down	20	1 min

For purification, the sequencing reaction was precipitated with 25 µL 100% ethanol (EtOH) for 15 min in the dark followed by centrifugation at 3,000 g for 30 min at 10°C. The pellet was washed with 125 µL 70% EtOH, centrifuged at 2,000 g for 10 min at 10°C, and placed at RT in the dark for 10 min for the EtOH to evaporate. The pellet was subsequently resuspended in 40 µL HPLC water, transferred to a microtiter plate, and placed into the automated ABI 3730 sequencer. Resulting sequences were analysed using the SnapGene software (see Chapter 2.2.7).

Whole exome sequencing

Whole exome sequencing data analysed in this thesis were either generated at the Institute of Human Genetics or Neurogenomics (Helmholtz Zentrum München, Neuherberg, Germany) (as below) or were imported from external collaboration partners. Exonic regions were enriched using the SureSelect Human All Exon kit from Agilent (Agilent Technologies, Santa Clara, USA) followed by sequencing as 100 bp paired-end runs on an Illumina HiSeq2,000, Illumina HiSeq2500, and Illumina HiSeq 4000 (AG_50MB_v4, AG_50MB_v5, AG_50MB_v5, and AG_60MB_v6 exome kit samples) (Illumina, San Diego, USA). An integrated analysis tool incorporating a primary analysis pipeline and an expert system for variant interpretation was utilised (EVAdb, see Chapter 2.2.7). For both in-house and imported data, variant calling and annotation incorporated numerous publicly available bioinformatics tools and customized software. Briefly, reads were aligned to the human reference genome (UCSC Genome Browser build hg19) using Burrows-Wheeler Aligner (v.0.7.5a). Single-nucleotide variants and small insertions and deletions (indels) were detected with SAMtools (version 0.1.19) and the Genome Analysis Toolkit (GATK). CNVs were identified with ExomeDepth. Runs of homozygosity (RoH) were detected by BCFtools/RoH, [143]. As per ACMG guidelines [144], consanguinity is suspected when RoH accounted for 10% of the genome, consistent with a first- or second-degree parental relationship. Variants were called from the mitochondrial genome (mtDNA) as described in [31]. The reported variants were annotated with their genomic coordinates, allele frequency (gnomAD database,

gnomad.broadinstitute.org/), and functional consequence on the gene product accoding to the Ensembl Variant Effect Predictor (VEP, see Chapter 2.2.7).[2] EVAdb (see Chapter 2.2.7) was used to explore variants. Variants were prioritised by predicted deleterious effect on the protein by the CADD score [95], SIFT [145] score, and PolyPhen [146] (see Chapter 2.2.7).

In study 1, analysing WES in >2,000 patients and in study 4, integrating multi-omics for the diagnosis of mitochondrial disease patients, variants determined to be potentially disease-causing were formally classified by the ACMG recommendations [106] using the Python package "InterVar" [107] and were subjected to co-segregation analyses. A symmetric semantic similarity of $\geq$2 was accepted as a phenotype match (criteria PP4) in accordance with [130]. A mitochondrial RCC defect on muscle biopsy or a maximal respiration rate <71.6% in patient-derived fibroblast cell lines were accepted as functional evidence for variants in MD genes (ACMG criteria PS3) in accordance with [60]. In autosomal and X-linked dominant disease genes, a pathogenic variant (P, ACMG class 5) or a likely pathogenic variant (LP, class 4) was reported as a definite genetic diagnosis. In autosomal and X-linked recessive disease genes, biallelic P/LP variants were reported as definite genetic diagnoses. Patients with mitochondrial depletion or multiple mtDNA deletions and no identified causative variant(s) were not considered to be solved. All disease-causing variants were in the process of submission to the ClinVar database (see Chapter 2.2.7) at the time of writing this thesis.

In study 1, analysing WES in >2,000 patients two automated variant analyses were undertaken. First, an automated reanalysis for the identification of homozygous ACMG classified P variants in AR and XLR OMIM disease genes, with the objective to detected overlooked diagnoses and dual diagnoses. Second, an automated search for potential novel mitochondrial disease genes. In this analysis, all rare variants (MAF<0.1%) in genes encoding predicted mitochondrially-localised proteins (according to MitoCarta3.0, see Chapter 2.2.7) and not yet known to be disease-associated according to OMIM were considered. Variants were filtered to be missense (with a CADD>25), frameshift, or direct splice site in predicted function, followed by filtering to be potentially biallelic (i.e., by detection of homozygous variant or by deletion of two variants with the potential to be compound heterozygous).

RNA-sequencing

RNA was isolated from whole-cell lysates using the AllPrep RNA Kit (Qiagen, Hilden, Germany) and RNA integrity number (RIN) was determined with the Agilent 2100

[2] All methods prior to variant prioritisation and interpretation were undertaken by a team of coworkers.

BioAnalyzer (RNA 6000 Nano Kit, Agilent Technologies, Santa Clara, USA). For library preparation, 1 µg of RNA was poly(A) selected, fragmented, and reverse transcribed with the Elute, Prime, Fragment Mix (Illumina, San Diego, USA). A-tailing, adaptor ligation, and library enrichment were performed as described in the TruSeq Stranded mRNA Sample Prep Guide (Illumina, San Diego, USA). RNA libraries were assessed for quality and quantity with the Agilent 2100 BioAnalyzer and the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, USA). RNA libraries were sequenced as 150 bp paired-end runs on an Illumina HiSeq4000 platform. The STAR aligner (v 2.4.2a) [147] with modified parameter settings (–twopassMode=Basic) was used for split-read alignment against the human genome assembly hg19 (GRCh37) and UCSC knownGene annotation. To quantify the number of reads mapping to annotated genes HTseq-coun (v0.6.0) was used [148]. If the 95th percentile of the coverage across all samples was below 10 reads the gene was considered "not expressed" and discarded from later analysis. Fragments Per Kilobase of transcript per million fragments mapped (FPKM) values were calculated using custom scripts of the Institute of Human Genetics (Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany) and the Institute of Neurogenomics (Helmholtz Zentrum München, Neuherberg, Germany).[3]

In study 3, validating *DNAJC30* as a novel mitochondrial disease gene, RNA-seq data underwent normalisation to control for known and unknown confounders using OUTRIDER [137] and differential expression analysis was performed using the R Bioconductor package DESeq2 [149]. Gene set enrichment analysis (GSEA) was performed in the R Bioconductor package clusterProfiler [150]. The MitoPathways3.0 dataset was used as input to the GSEA (MitoCarta3.0 genes annotated into a hierarchy of 149 biological pathways) (see Chapter 2.2.7). The Benjamini Hochberg method [151] was applied to correct the p values for multiple testing in the GSEA.

In study 4, integrating multi-omics for the diagnosis of mitochondrial disease patients, RNA-seq analysis was performed using the DROP pipeline [132], a workflow integrating quality controls, data normalisation, and expression outlier calling with OUTRIDER [137], splicing outlier calling with FRASER [133], and MAE with a negative binomial test [117].[4]

[3] All methods prior to RNA-seq data analysis were undertaken by a team of coworkers.
[4] RNA-seq data analyses in the DROP pipeline undertaken by Dimitrii Smirnov.

2.2.2 Protein analysis

Immunoblotting

Immunoblotting was performed to confirm reduced abundance of the FDXR protein in patient-derived fibroblast cell lines. Cell lysates were prepared by resuspension of frozen fibroblast pellets in 300 µL RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% (v/v) NP-40, 0.1% (w/v) SDS, 0.5% (w/v) deoxycholate) supplemented with 1:100 Protease Inhibitor Cocktail Set III, Animal-free (Calbiochem, an affiliated of Merck, Darmstadt, Germany). Samples were incubated for 1 h at 4°C on rotation. The mixture was centrifuged for 10 min at 15,000 g at 4°C. The supernatant was recovered and disrupted by five strokes with a 0.30 mm x 8 mm syringe (Becton, Dickinson and Company, Franklin Lakes, USA). The whole protein amount of the recovered supernatant was quantified by Bradford assay using Protein Assay Dye Reagent Concentrate (Bio-Rad, Hercules, USA) on a Jasco V- 550UV/VIS Spectrophotometer. Subsequently, samples were adjusted to 1.5 µg protein/µL in 1x Laemmli buffer (5% (w/v) SDS, 250 mM Tris-HCl pH 6.8, 50% (v/v) glycerol, 500 mM ß-mercaptoethanol, 0.025% (w/v) bromphenol blue) and heated for 10 min at 50°C. 30 µg of protein per sample was loaded onto a precast gel (Lonza, Basel, Switzerland). Electrophoresis in 1x ProSieve EX Running buffer (Lonza, Basel, Switzerland) ran at 50 V for 30 min followed by 120 V for 45 min. Proteins were subsequently transferred semi-dry to PVDF membranes (GE Healthcare Life Sciences, Chalfont St. Giles, UK) using 1x ProSieve EX Western Blot Transfer buffer (Lonza, Basel, Switzerland) at a constant voltage of 25 V for 15 min. Protein transfer was confirmed by staining with Ponceau S solution. The membranes were blocked in 5% non-fat milk (Bio-Rad, Hercules, USA) in TBST (150 mM NaCl, 30 mM Tris base, pH 7.4, 0.1% Tween 20) for 1 h. Prior to incubation with the primary antibody (FDXR 1:250, β actin 1:1,000), the membrane was divided between the molecular weight of the target protein and the loading control (FDXR 54 kDa, β actin 42 kDa). The membranes were incubated on a roller either overnight at 4°C or for 1 h at RT, for the target protein and loading control respectively. Signals were detected by subsequent incubation with HRP-conjugated goat anti-rabbit secondary antibody (1:10,000, Jackson Immuno Research Laboratories, West Grove, US) for 1 h and visualized using ECL (GE Healthcare Life Sciences, Chalfont St. Giles, UK).

TMT-labelled quantitative proteomics

Quantitative proteomics was performed at the BayBioMS core facility of the Technical University of Munich (Freising, Germany). Fibroblast cell pellets containing 0.5 million cells were lysed under denaturing conditions in urea containing buffer and quantified

using BCA Protein Assay Kit (Thermo Scientific, Waltham, USA). 15 µg of protein extract was further reduced, alkylated, and the tryptic digest was performed using Trypsin Gold (Promega, Madison, USA). Digests were acidified, desalted and TMT-labelling was performed according to [152] using TMT 10-plex labelling reagent (Thermo Scientific, Waltham, USA). Each TMT-batch consisted of eight patient samples and two reference samples which allowed for data normalisation between batches. Each TMT 10-plex peptide mix was fractionated using trimodal mixed-mode chromatography as described by [153]. LC-MS measurements were conducted on a Fusion Lumos Tribrid mass spectrometer (Thermo Scientific, Waltham, USA), which was operated in data-dependent acquisition mode and multi-notch MS3 mode. Peptide identification was performed using MaxQuant (version 1.6.3.4) [154] and protein groups obtained. Data was processed by sample-wise normalisation to account for different sample loadings across channels in a TMT 10-plex experiment, and by protein-wise normalisation to enable comparisons of samples across-TMT 10-plex experiments.[5]

In study 3, validating *DNAJC30* as a novel mitochondrial disease gene, differential expression analysis was performed using the R Bioconductor package limma [155]. GSEA was performed in the R Bioconductor package clusterProfiler [150]. The MitoPathways3.0 dataset was used as input to the GSEA (MitoCarta3.0 genes annotated into a hierarchy of 149 biological pathways, see Chapter 2.2.7). The Benjamini Hochberg method [151] was applied to correct the p values for multiple testing in the GSEA.

In study 4, integrating multi-omics for the diagnosis of mitochondrial disease patients, aberrant protein expression was identified using the algorithm PROTRIDER. PROTRIDER was developed in collaboration with the research group of Prof. Julien Gagneur to estimate deviations from expected protein intensities while controlling for known and unknown sources of proteome-wide variation as described in [118].[6]

Blue native electrophoresis and complexome analysis

In study 3, validating *DNAJC30* as a novel mitochondrial disease gene, blue native electrophoresis and complexome analysis were utilised to determine mitochondrial complex assembly and abundance in DNAJC30 defect patient-derived fibroblast cell lines, and in the *DNAJC30*-KO HEK cellular model. Sample preparation and blue native electrophoresis (BNE) of cultured cell pellets was performed as described in [156]. Each lane of the BNE was cut into equal fractions for complexome profiling, performed as described in [157].[7]

[5] All methods prior to proteomic data analysis were undertaken by a team of coworkers.
[6] Proteomic data analyses in the PROTRIDER pipeline undertaken by Dimitrii Smirnov.
[7] Undertaken by Dr. Ilka Wittig (Goethe University of Frankfurt, Frankfurt am Main, Germany).

Measurement of protein turnover in mitochondrial complexes

In study 3, validating *DNAJC30* as a novel mitochondrial disease gene, pulsed Stable Isotope Labeling by Amino acids in Cell culture (pSILAC) metabolic labeling was used to allow measurement of protein turnover in mitochondrial complexes in DNAJC30 defect patient-derived fibroblast cell lines, and in the *DNAJC30*-KO HEK cellular model. For pSILAC, cell culture medium was exchanged with medium containing $^{13}C_6$,$^{15}N_4$-L-Arginine and $^{13}C_6$,$^{15}N_2$-L-Lysine isotopes (Silantes, Munich, Germany) to allow quantification of newly synthesised protein at 0, 6, 8, 10, and 12 h. Mitochondrial membranes were solubilised with either digitonin or detergent dodecyl maltoside (DDM). Protein complexes were separated by BNE and the bands representing either the supercomplex and CV or the individual complexes (CI-V), in digitonin and DDM solubilization respectively, were extracted for mass spectrometry. Procedures and parameters for mass spectrometry and data analysis were summarized at PRIDE-proteomics identification database with the identifiers PXD021385, PXD021386, PXD-021500, PXD022340, PXD022339, and PXD021548.[8]

Protein turnover was calculated using mass spectrometry intensity-based absolute quantification (IBAQ) values as:

$$\frac{\text{Heavy IBAQ}}{\text{Light IBAQ} + \text{Heavy IBAQ}}$$

Heavy IBAQ is a quantification of newly synthesised proteins containing isotope labled amino acids. Light IBAQ is a quantification of proteins synthesised prior to the cell culture medium exchange to culture medium containing the $^{13}C_6$,$^{15}N_4$-L-Arginine and $^{13}C_6$,$^{15}N_2$-L-Lysine isotopes. Pymol version 2.3.3 was used to generate turnover heatmaps on the mouse complex I structure (PDB 6g2) [158].

For the analysis of protein complex turnover, all detected proteins were mapped onto individual protein complexes in the CORUM (v3.0) database [159] using the "coreComplexes.txt" file downloaded from the CORUM database (see Chapter 2.2.7) in addition to a number of manually curated subgroups of interest of mitochondrial CI subunits (e.g., the CI N-module). The number of quantified subunits in each CORUM complex with $\geq$2 subunits were counted. To identify specific protein complexes with differential turnover, the turnover rate of all proteins in a CORUM complex was compared between the control and *DNAJC30*-KO HEK cell line at 12 h.

[8] Undertaken by Dr. Ilka Wittig (Goethe University of Frankfurt, Frankfurt am Main, Germany).

The delta turnover of proteins between the control and the *DNAJC30*-KO HEK cell line was calculated as:

$$\left(\frac{\text{Heavy IBAQ in control}}{\text{Light IBAQ + Heavy IBAQ in control}}\right) - \left(\frac{\text{Heavy IBAQ in KO}}{\text{Light IBAQ + Heavy IBAQ in KO}}\right)$$

Subsequently, the delta turnover for each complex was calculated by the mean of the delta turnover values of all identified proteins in the complex.

2.2.3 Bacterial culture and techniques

Bacterial transformation

One Shot Stbl3 chemically competent *E. coli* cells were used in the amplification of plasmid DNA. Briefly, the cells were thawed on ice. 1-2 µL plasmid DNA was mixed gently with 25 µL bacterial cells, followed by incubation on ice for 30 min. Heat-shock at 42°C was performed for 30 sec in a water bath, followed by incubation on ice for 2 min. 250 µL S.O.C. medium was added to the cells and incubated with horizontal shaking (45% in a Heiz-Thermomixer HTML 133) at 37°C for 1 h. 100 µL bacterial cell suspension was plated on pre-warmed LB plates containing a plasmid specific selection antibiotic (100 mg/mL, 1:1,000). Bacterial plates were incubated overnight at 37°C. The following day colonies were selected for colony PCR.

Colony PCR

20 µL HPLC water was inoculated with bacteria from individual colonies and denatured at 95°C for 20 min in the thermocycler, of which 2.5 µL were utilised in a 10 µL colony PCR reaction containing the following:

	Volume for 10 µL reaction	Final concentration
HPLC water	4.05 µL	
DNA extracted from bacterial colony	2.5 µL	
Taq DNA Polymerase	0.05 µL	0.25 U
10x PCR Buffer	1 µL	1 x
2 mM dNTPs	1 µL	0.2 mM
Q solution	1 µL	1 x
10 µM Primer (F)	0.2 µL	0.2 µM
10 µM Primer (R)	0.2 µL	0.2 µM

The amplification conditions were as follows:

	Temperature (°C)	Time
Heat lid	110	
Denature	95	5 min
Start cycle (25 cycles)		
Denature	95	1 min
Anneal	Primer T_m - 5	30 sec
Extend	72	4 min
Close cycle		
Extend	72	5 min
Cool down	20	1 min

2.5 µL the resultant PCR product was subjected to agarose gel (1.5%) electrophoresis (100 V, 30 min) for the assessment of product quality, size, and where applicable (e.g., when cloning a DNA sequence into an expression vector), orientation of DNA insertion into the expression vector.

2.2.4 Yeast culture and techniques

In study 2, validating VUS in *FDXR*, an *Arh1*-null (human *FDXR* ortholog) *S. cerevisiae* strain was used as a disease model. This yeast strain was provided by Paul et al., [124] and was transformed with a pCM189 plasmid expressing *Arh1* and conferring resistance to kanamycin.

Yeast culture

Yeast was cultured either in YPD (yeast-extract-peptone-dextrose) medium supplemented with kanamycin (1:1,000) in a shaking incubator at 30°C 200 rpm, or on YPD agar plates supplemented with kanamycin (1:1,000) at 30°C to maintain the pCM189 plasmid.

Competent yeast cell preparation

To produce competent yeast cells for efficient transformation, cells were cultured over night in a nutrient rich medium (1:1 volume of YPD medium and 80% glycerol) supplemented with kanamycin (1:1,000) in a shaking incubator at 30°C and 150 rpm. The cell density of the overnight culture was determined by OD_{600} measured by the Jasco V- 550UV/VIS Spectrophotometer. 50 mL of fresh medium was inoculated to OD_{600}=0.15 and allowed to grow at 30°C until OD_{600}=0.5. The cells were subsequently harvested by centrifugation (2,000 g for 2 min at RT) and washed by resuspension in 25 µL sterile H_2O followed by a further centrifugation and wash by resuspension in

6.25 µL LiSorb (100,µM Li-actetat, 10 µM Tris/HCl pH8, 1 µM EDTA, 1 M Sorbitol). The cells were cultured at RT for 5 min before a final centrifugation with subsequent resuspension in 300 µL LiSorb with the addition of 30 µL carrier DNA. 50 µL aliquots were prepared and stored at -80°C in preparation for transformation.

Site-directed mutagenesis

Site-directed mutagenesis was used to create specific targeted changes to a double stranded BFG-1 plasmid, a vector used for expression in *S. cerevisiae*. This plasmid, expressing full length *FDXR*, was provided by Paul as colleagues [124]. The protocol for site-directed mutagenesis was adapted from [160]. In brief, primers pairs were designed to share an overlapping complementary region at the 5' end (∼25 b) and a non-overlapping region at the 3' end (∼35 b). The desired mutation was positioned in the center of the overlapping region. The melting temperature (T_m) of the non-overlapping 3' regions was designed to be 5-10 °C higher than the T_m of the overlapping 5' region.

The mutagenesis PCR reactions were performed as follows:

	Volume for 50 µL reaction	Final concentration
HPLC water	28.5 µL	
Template DNA (2 ng/µL)	1 µL	2 ng
Phusion DNA polymerase (2 U/µL)	0.5 µL	1 U
5 x Phusion HF Buffer	10 µL	1 x
2 mM dNTPs	5 µL	0.2 mM
10 µM Primer mix (F and R)	5 µL	1 µM

The mutagenesis PCR conditions were as follows:

	Temperature (°C)	Time
Heat lid	110	
Denature	95	5 min
Start cycle (12 cycles)		
Denature	95	1 min
Anneal	T_m non-overlapping region - 5	30 sec
Extend	72	1 min per 500 bp
Close cycle		
	T_m non-overlapping region - 5	1 min
Extend	72	30 min
Cool down	20	1 min

1 µL Dpnl restriction enzyme (20 U/µL) (New England Biolabs, Ipswich, USA) was added to the resultant PCR product and incubated for 1 h at 37°C. Dpnl selectively digests methylated DNA to ensure only unmethylated DNA created by the mutagenesis

PCR reaction remained. 10 µL the PCR product was subjected to low percentage agarose gel (1%) electrophoresis (100 V, 30 min) for the assessment of product quality and size. The desired result was a single band at ∼8 kb, corresponding to the full-length BFG-1 plasmid (6.5 kb) and *FDXR* insert (1,476 b). 2 µL the PCR product was used to transform One Shot Stbl3 chemically competent *E. coli* cells for amplification of the plasmid (see Chapter 2.2.3). 100 µL bacterial cell suspension was plated on the pre-warmed LB plates containing ampicillin (100 mg/mL, 1:1,000) for which the BFG-1 plasmid conferred resistance. Bacterial plates were incubated overnight at 37°C. The following day, five colonies from each site-directed mutagenesis reaction were selected for colony PCR using a primer combination for the *FDXR* sequence inserted into the plasmid (BFG-1 plasmid (F) + BFG-1 plasmid (R)) (see Chapter 2.2.3). 2.5 µL the resultant PCR product was subjected to agarose gel (1.5%) electrophoresis (100 V, 30 min) for the assessment of product quality and size. 2.5 µL the PCR product was Sanger sequenced (see Chapter 2.2.1) to select plasmids with a desired mutation in the *FDXR* sequence. Selected colonies were grown overnight in 4 mL of LB medium supplemented with ampicillin (100 mg/mL, 1:1,000). Plasmid isolation was performed with the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. The concentration of the isolated plasmid DNA was measured on NanoDrop OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, USA) and the presence of the desired mutation in the *FDXR* sequence was once more confirmed by Sanger sequencing (see Chapter 2.2.1) prior to yeast transformation.

Yeast transformation

5 µL plasmid DNA was added to 10 µL competent yeast cells and gently mixed with 300 µL LiPEG (100 mM Li-actetat, 10 mM Tris/HCl pH8, 1 mM EDTA, 40% PEG4000), followed by incubation for 30 min at 30°C. 35 µL DMSO was added and the cells were subjected to a heat shock for 20 min at 42°C in a water bath. Cells were gently centrifuged (1500 g for 2 min at RT) and resuspended in medium for plating and incubation for 2-3 days at 30°C.

Yeast growth assay

The *Arh1* expressing pCM189 plasmid was eliminated by 5-FOA treatment to evaluate the growth capability of the *Arh1*-null cells transformed with BFG-1 plasmids expressing wild-type (control) and mutated *FDXR*, as created by site-directed mutagenesis. Cells were spotted onto YPD or YPG (yeast-extract-peptone-glycerol) plates. Drop dilution growth tests were performed at 1:5 dilution steps (initially with 2,000,000 seeded cells), and plates were incubated for three days at either 30°C or 35°C. Four conditions were

selected for yeast culture to progressive stress the cells, (i) YPD at 30°C, (ii) YPD at 35°C, (iii) YPG at 30°C, and iv) YPG at 35°C.[9]

2.2.5 Cell culture and techniques

In study 3, validating *DNAJC30* as a novel mitochondrial disease gene, to understand the role of DNAJC30 in mitochondrial function and the consequence of DNAJC30 defects, patient-derived fibroblast cell lines and a *DNAJC30*-KO cellular model were employed.

Cell culture

Primary patient-derived fibroblast cell lines, NHDF, and HEK 293FT cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies, Carlsbad, USA), supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, and 200 µM uridine (Life Technologies, Carlsbad, USA). The culture medium was replaced once every 2-3 days. Cells were grown at 37°C in the presence of 5% CO_2. All patient-derived fibroblast cell lines were tested negative for mycoplasma contamination using the MycoAlert Mycoplasma detection kit (Lonza, Basel, Switzerland). Where required (e.g., for measurement of cellular oxygen consumption rate), the cell count was determined by the Scepter Handheld Cell Counter with 60 µM Scepter Sensors following manufacturer's protocol.

Measurement of cellular oxygen consumption rate

Cellular oxygen consumption rate was used as a readout of mitochondrial function. The Seahorse XF96 Extracellular Flux Analyzer (Seahorse Bioscience, Agilent Technologies, Santa Clara, USA) "Mito Stress Test" was utilised to measure glucose-dependent respiration in fibroblast and HEK cell lines according to the manufacturer's protocol. Briefly, one day prior to measurement, fibroblast or HEK cells were seeded in 80 µL cell culture medium in a XF 96-well cell culture microplate (Seahorse Bioscience, Agilent Technologies, Santa Clara, USA), at 20,000 or 15,000 cells per well, respectively. The culture microplate was incubated at 37°C in the presence of 5% CO_2 overnight. The sensor cartridge was rehydrated in the XF96 utility plate by adding 200 µL XF calibrant solution per well. On the day of measurement, cells were gently washed with bicarbonate-free DMEM (Life Technologies, Carlsbad, USA) and subsequently incubated in 180 µL bicarbonate-free DMEM at 37°C for 30 min prior to measurement. The oxygen consumption rate (OCR) was measured using the XF96 Extracellular Flux Analyzer (Seahorse Biosciences, Billerica, USA). After an initial calibration step of

[9] Method undertaken together with Robert Kopajtich and Lea Kulterer.

30 min, the OCR was measured in iterative cycles of three mixing and three measuring steps. The cycle was initially performed with no additions, followed by the sequential addition of oligomycin (1 µM), FCCP (0.4 µM), rotenone (2 µM), and antimycin (2.5 µM), resulting in three measurement points per condition per well. Individual measurement points were normalised to mean non-mitochondrial respiration per well. Maximal respiration rate was normalised to the control cell line, included on all plates, to facilitate comparison across experiments.

Functional complementation by lentiviral transduction

Functional complementation by lentiviral transduction was used to determine whether re-expression of wild-type *DNAJC30* could rescue measured cellular defects, as evidence of pathogenicity for the potential novel disease gene. Lentivirus-mediated expression of the full-length *DNAJC30* in fibroblast cell lines was performed using the ViraPower HiPerform Lentiviral TOPO Expression Kit (Thermo Fisher Scientific, Waltham, USA) according to [117]. In brief, full length *DNAJC30* gDNA (one exon) from a healthy control was first amplified by PCR. Primers were designed to flank the ORF and to introduce a 5' Kozak consensus sequence (5'-CACG-3') immediately ahead of the ATG start codon to provide a site for protein translation initiation. The Platinum Taq DNA Polymerase High Fidelity Kit (Thermo Fisher Scientific, Waltham, USA) was used to ensure a high degree of replication accuracy.

The 10 µL PCR reaction contained the following:

	Volume for 10 µL reaction	Final concentration
HPLC water	6.16 µL	
Template DNA	1 µL	5 ng
Platinum® Taq DNA Polymerase High Fidelity (5 U/µL)	0.04 µL	0.2 U
10x High Fidelity PCR Buffer	1 µL	1 x
2 mM dNTPs	1 µL	0.2 mM
50 nM $MgSO_4$	0.4 µL	2 mM
10 µM Primer (F)	0.2 µL	0.2 µM
10 µM Primer (R)	0.2 µL	0.2 µM

Amplification was performed under the following conditions:

	Temperature (°C)	Time
Heat lid	110	
Denature	95	2 min
Start cycle (30 cycles)		
Denature	95	30 sec
Anneal	60	30 sec
Extend	68	1 min per kilobase
Close cycle		
Extend	68	10 min
Cool down	20	1 min

The PCR product was analyzed using agarose gel (1.5%) electrophoresis (100 V, 30 min). 1 µL was subsequently cloned into the pLenti6.3/V5-TOPO expression vector and transformed into One Shot Stbl3 Competent *E. coli* cells (see Chapter 2.2.3). 100 µL bacterial cell suspension was plated on the pre-warmed LB plates containing Ampicillin (100 mg/mL, 1:1,000) for which the plasmid conferred resistance. Bacterial plates were incubated overnight at 37°C. 12 colonies were selected for colony PCR (see Chapter 2.2.3). The colony PCR utilised primer pairs allowing confirmation of the orientation of the DNA insertion (*DNAJC30* Kozak+ORF (F) + pLenti6.3/V5-TOPO V5 and pLenti6.3/V5-TOPO CMV + *DNAJC30* Kozak+ORF (R)). 2.5 µL the resultant PCR product was subjected to agarose gel (1.5%) electrophoresis (100 V, 30 min) for the assessment of product quality, size, and orientation of DNA insertion into the pLenti6.3/-V5-TOPO expression vector. Three colonies with correctly orientated *DNAJC30* were selected for overnight growth in 4 mL of LB medium supplemented with ampicillin (100 mg/mL, 1:1,000) and subsequent plasmid isolation with the QIAprep Spin Midiprep Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. The concentration of the isolated plasmid DNA was measured on NanoDrop OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, USA). The *DNAJC30* sequence insert in the pLenti6.3/V5-TOPO expression vector was confirmed by Sanger sequencing and one sequence validated clone was selected for subsequent transfection reactions.

DNAJC30 was re-expressed in fibroblast cell lines using the Lenti-Pac FIV Expression Packaging Kit (GeneCopoeia, Rockville, USA). The expression vector was cotransfected with a packaging plasmid mix into HEK 293FT cells using Lipofectamine 2,000. The transfection mix was replaced with high glucose DMEM supplemented with 10% FBS 24 h after transfection. 72 h after, the supernatant containing the viral particle was collected and used to transduce a control NHDF and a patient-derived fibroblast cell line. Cells stably expressing *DNAJC30* were selected by medium supplementation with blasticidin 5 µg/mL (Thermo Fisher Scientific, Waltham, USA). Cells remained

in selection for a minimum of two weeks prior to experimentation to ensure only cells stably expressing *DNAJC30* remained.

CRISPR-Cas knock-out cell line generation

To provide a second disease model in addition to patient-derived fibroblast cell lines, a *DNAJC30*-KO HEK cellular model was generated. The commercially available Origene KN2.0 non-homology mediated CRISPR-Cas *DNAJC30* knock-out kit was utilized according to the manufacturer's protocol to generate a *DNAJC30*-KO HEK cell line. The method targeted a specific guide RNA (gRNA) to cut the genome and allow integration of a linear donor DNA containing a selection cassette on one allele with introduction of an indel on the second allele of *DNAJC30*, resulting in biallelic knock-out. In brief, 24 h prior to transfection, HEK 293FT cells were plated in a 6-well culture plate, 200,000 cells per well, with 2 mL of culture medium. Upon reaching 50-70% confluency, the transfection reagents (20 µL gRNA vector, 250 µL Opti-MEM I, 1 µg donor DNA, 8.0 6 µL Turbofectin) were gently combined in the prescribed order, incubated for 15 min at RT, and subsequently added drop-wise to the plated cells. Following 48 h of incubation, cells were split 1:10, grown for three days, and split 1:10 a further 1-3 times as required over a period of two weeks. Cells with successful integration of the selection cassette, conferring antibiotic resistance, were selected by the addition of puromycin 0.5 µg/mL to the culture medium. The dosage of puromycin was determined with a kill curve, to be the lowest dose to kill non-transfected cells by 5-7 days. Single puromycin resistant cell colonies were isolated by limiting dilution by plating a single cell suspension in a 96-well culture plate. Following 1-2 weeks of growth, the presence of single cell colonies was confirmed by direct observation under the brightfield microscope. Wells with single cell colonies were split 1:10 and 1:2 into two duplicate 96-well plates, for continued expansion and for DNA extraction, respectively. DNA was extracted once cells were confluent using QuickExtract™ DNA Extraction Solution (Lucigen, Middleton, USA). The extracted DNA was subject to PCR amplification of *DNAJC30* (see Chapter 2.2.1). The PCR resulted into two bands per sample. First, a band at ∼3.4 kb reflecting amplification of the allele containing the selection cassette. Second, a band at ∼700 b reflecting amplification of the allele containing indels. The ∼700 b bands were subsequently extracted from the gel using the Monarch DNA Gel Extraction Kit (New England Biolabs, Ipswich, USA). The presence of indels was determined by Sanger sequencing (see Chapter 2.2.1). One colony containing a c.158_159insA (NM_032317.2) frameshift mutation leading to a premature stop codon within the functional DNAJ domain of the DNAJC30 protein (p.Leu53Leufs*58], NP_115693.2) was selected for expansion and further experiments. Biallelic knock-out of DNAJC30 was further confirmed by TMT-labelled quantitative proteomics (see Chapter 2.2.2).

2.2.6 Data analysis

Data restructuring and automation of analysis

In study 1, analysing WES in >2,000 patients, data analysis focused on data restructuring, automation of downstream statistical analyses, data visualisation, and the development of an interactive online resource GENOMITexplorer with "Snakemake" [161] and "wBuild" [162]. Automation of these analyses was necessitated by the dynamic inclusion and continual collection of data. For these purposes, an analysis pipeline coded in R (version 3.6.1) and Python (version 3.8.2) was developed. This pipeline encompassed: (i) human phenotype ontology (HPO) data restructuring with the R package "OntologyX" [163] to facilitate comparison of patient HPO terms at different levels on the ontology tree (reflecting different levels of phenotyping depth), (ii) symmetric semantic similarity (SS) calculation with the R package "PCAN" [164] to facilitate comparison of patient HPO terms with established disease gene associated HPO terms, and (iii) variant annotation and automated ACMG classification with the Python package "InterVar" [107] to designate variant pathogenicity in a robust manner. These codes were subsequently utilised in study 4, integrating WES, HPO phenotype, RNA-seq, and proteomic data for the diagnosis of mitochondrial disease patients.

In study 3, validating *DNAJC30* as a novel mitochondrial disease gene, R codes were developed for the automated analysis of OCR data (e.g., in normalisation, calculation of maximal respiration rate, statistical analysis, and data visualisation), and for the automated analysis of proteomic data (e.g., in normalisation, calculation of protein turnover, statistical analyses, and data visualisation).

Statistical analysis

Statistical analyses used R version 3.6.1. The choice of statistical test was determined by the distribution of the data. Where the assumptions of normality and equal variance were met, the parametric two-sided Student's t-test was used to compare the means of two groups, and where not, the Wilcoxon rank-sum test was used. Enrichment analyses and categorical data analysis used Fisher's exact tests with one degree of freedom. Bonferroni multiple-testing adjusted $p \leq 0.05$ were considered significant. Correlation analyses used Pearson's or Spearman's rank correlation, depending on the distribution of the data. When multiple comparisons were made to a control, the p values were corrected with the Dunnett's test. To determine the performance of MDC scores and biomarkers, receiver operator characteristic (ROC) curves and the area under the curve (AUC) were calculated using the R package "ROCR". The p values are annotated in all figures as follows: $p > 0.05$ (NS), $p \leq 0.05$ (*), $p \leq 0.01$ (**), $p \leq 0.001$ (***), and $p \leq 0.0001$ (****).

Analysis of clinical data

In study 1, analysing WES in >2,000 patients, and Chapter 3.4, integrating multi-omics for the diagnosis of mitochondrial disease patients, patient phenotypes were curated from medical records and/or mitochondrial disease registry entries and/or case reports in the literature, and were subsequently digitilised as HPO terms. These HPO terms were integrated into a local database (EVAdb) and shared with all collaboration partners via the development of the interactive web resource GENOMITexplorer (see Chapter 2.2.7). As aforementioned, all ancestral HPO terms were derived from the ontology using the R package "OntologyX" [163] to facilitate comparison between patients with differing depth of reported phenotype. The number of informative (non-redundant) HPO terms was reported as the number of HPO terms collected per patient. Symmetric SS was calculated between patient HPO terms and disease-gene associated HPO terms using the R package "PCAN" [164] and the HPO "genes to phenotype" annotation file (human-phenotype-ontology.github.io/downloads, downloaded January 2021). The analysis was limited to genes with an associated mode of inheritance in the OMIM database i.e., known Mendelian disease genes (omim.org, accessed January 2021). A phenotype match between the patient and the reported gene-associated phenotype was defined as a SS $\geq$2 in accordance with [130]. Affected organ systems were visualized with the R package "gganatogram" [165]. To retrospectively determine the clinical likelihood of mitochondrial disease in our cohort of >2,000 paediatric suspected mitochondrial disease patients (see Chapter 3.1), the Mitochondrial Disease Criteria developed by Morava et al., [50] and by Witters et al.,[51] were modified for use with HPO terms (see Appendices).

In study 2, validating VUS in *FDXR*, to objectively determine disease severity in patients with defects in *FDXR* and to allow comparison to previously reported cases, the Newcastle Paediatric Mitochondrial Disease Scale (NPMDS) [52] was modified to provide an objective measure of disease severity comparable between patients from different pediatric age groups. This resulted in a multi-systemic scoring system encompassing 20 phenotypes, each with a possible score of 0 (normal), 1 (mild), 2 (moderate), and 3 (severe), and one phenotype category "developmental" with a possible score of 0 (normal), 1-2 (isolated motor, speech, language, or hearing delay with or without developmental progression), 5-6 (global developmental delay with or without progression), and 7 (developmental regression) (see Appendices). Patient case reports were utilized to score patients within this framework.

2.2.7 Resources

Resources utilised or referred to in this thesis were as follows:

Bioplex Interactome database	bioplex.hms/harvard.edu/
CADD calculator	cadd.gs.washington.edu/
ClinVar database	ncbi.nlm.nih.gov/clinvar/
CORUM database	mips.helmholtz- muenchen.de/corum/
DECIPHER database	deciphergenomics.org/
DIDA database	dida.ibsquare.be/
Ensembl Variant Effect Predictor	grch37.ensembl.org/Homo_sapiens/Tools/VEP/
EVAdb	github.com/mri-ihg/EVAdb/
GeneMatcher	genematcher.org/
Genome Aggregation Database	gnomad.broadinstitute.org/
GENOMIT	genomit.eu/
GENOMITexplorer web resource	wes-d9159.web.app/
Human Gene Mutation Database	hgmd.cf.ac.uk/
Human Phenotype Ontology	purl.obolibrary.org/obo/hp.obo/
Leigh Map	mseqdr.org/leighmap.php/
Matchmaker Exchange	matchmakerexchange.org/
MitoCarta3.0	broadinstitute.org/mitocarta/
MitoMap	mitomap.org/MITOMAP/
MitoPathways3.0 database	broadinstitute.org/mitocarta/
NEB T_m calculator	tmcalculator.neb.com/
omicsDiagnostics web resource	prokischlab.github.io/omicsDiagnostics/
OMIM database	omim.org/
Orphanet database	orpha.net/
PolyPhen calculator	genetics.bwh.harvard.edu/pph2/
ProtPhylo	protphylo.org/
SIFT calculator	sift.jvic.org/
SnapGene software	snapgene.com/
UCSC In-Silico PCR tool	genome.ucsc.edu/cgi-bin/hgPcr/

3 | Results

3.1 WES is a powerful diagnostic approach in mitochondrial disease

This study explored the diagnostic power of WES analysis in over 2,000 paediatric clinically suspected mitochondrial disease patients gathered in an international collaborative effort between centres specialising in the diagnosis of mitochondrial disease in Europe and Asia, and developed an interactive web resource for almost 4,000 patient and gene level HPO data sets (wes-d9159.web.app/).

The content of this study was subject to a manuscript with the working title "Diagnosing paediatric mitochondrial disease: lessons from 2,000 exomes" by **Stenton S.L** et al., at the time of writing this thesis.[10]

Throughout this study, mitochondrial disease will be abbreviated to MD.

[10] This project was led by myself and Dr. Holger Prokisch. All analyses presented in this chapter were personally performed unless otherwise indicated.

3.1.1 Study outline

In this study, WES and HPO-encoded phenotype data from 2,035 paediatric patients with a clinical suspicion of mitochondrial disease were assimilated and analysed. These patients were sequenced and initially analysed between October 2010 and January 2021 at one of 11 specialist MD diagnostic centres. The data collection was supported by an international collaboration initiated by the European Network for Mitochondrial Diseases (GENOMIT). An exome-wide search for disease-causing variation in known and potential novel disease genes was undertaken, complemented by functional studies. Variant classification followed ACMG guidelines, and the P/LP designation of all variants was confirmed using the algorithm "InterVar" to incorporate supportive evidence of variant pathogenicity on the phenotype and functional level (see Chapter 2.2.1).

3.1.2 Study population

Paediatric-onset (<18 years) clinically suspected MD patients were eligible for inclusion if there was no genetic diagnosis established by targeted candidate gene sequencing (including complete mtDNA sequencing in some centres) or no genetic testing prior to WES. In total, 2,035 patients were included, 1,960 index patients plus 75 affected siblings. 292/2,035 patients (14%) had previously been reported as part of discrete research studies or novel disease gene descriptions. WES data from single (1,791/2,035, 88%) or trio-based (244/2,035, 12%) analyses were collected. Data on the age of onset, clinical signs and symptoms, neuroimaging, metabolic investigations, muscle histology, and mitochondrial RCC measurements were collected from medical records and/or patient registries for all 2,035 patients and were subsequently encoded as HPO terms.

Across the 2,035 patients, a total of 949 independent informative HPO terms were extracted, a median of eight HPO terms per patient (range 2-40) across a median of four organ systems (range 1-12) (see **Tab. 5** and **Fig. 3.1**). 321/2,035 patients (16%) presented with a constellation of symptoms indicative of Leigh syndrome [47] (see Appendices for HPO criteria) and only 37/2,035 patients (2%) with symptoms suggestive of MELAS syndrome (Mitochondrial Encephalopathy, Lactic Acidosis, and Stroke-like episodes), Alpers syndrome, Kearns-Sayre syndrome, MERRF (myoclonic epilepsy with ragged red fibers), and LHON (Leber's hereditary optic neuropathy). The remaining 1,677/ 2,035 patients (82%) did not present with a distinct MD syndrome.

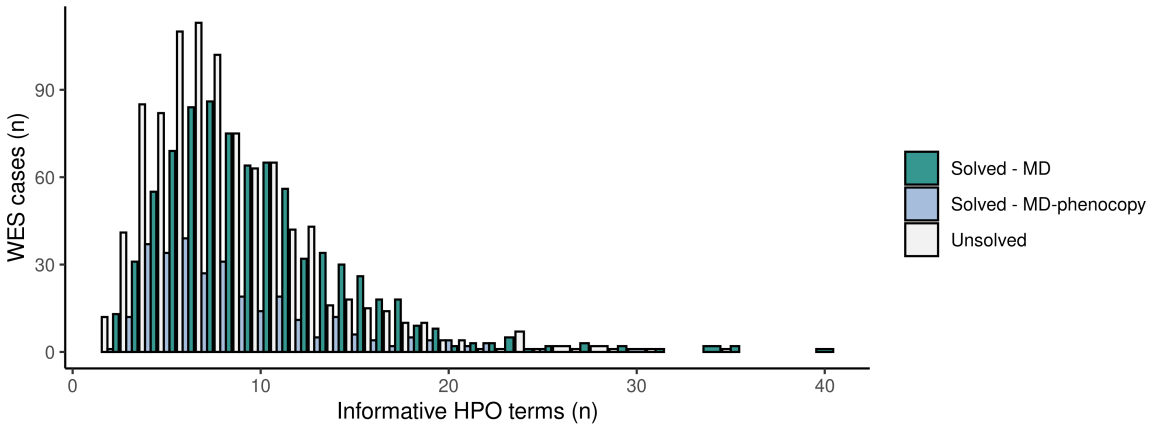


Figure 3.1: Number of HPO terms extracted per patient.
Number of HPO terms extracted per patient stratified by genetic diagnosis.

Key patient demographics are displayed in **Tab. 5**. 1,075/2,035 patients were male (53%). The age-of-onset ranged from neonatal to juvenile, with the majority of patients presenting in infancy (714/2,035, 35%).

3.1.3 Diagnostic rate

A summary of the analysis workflow and study resources is depicted in **Fig. 3.2**.

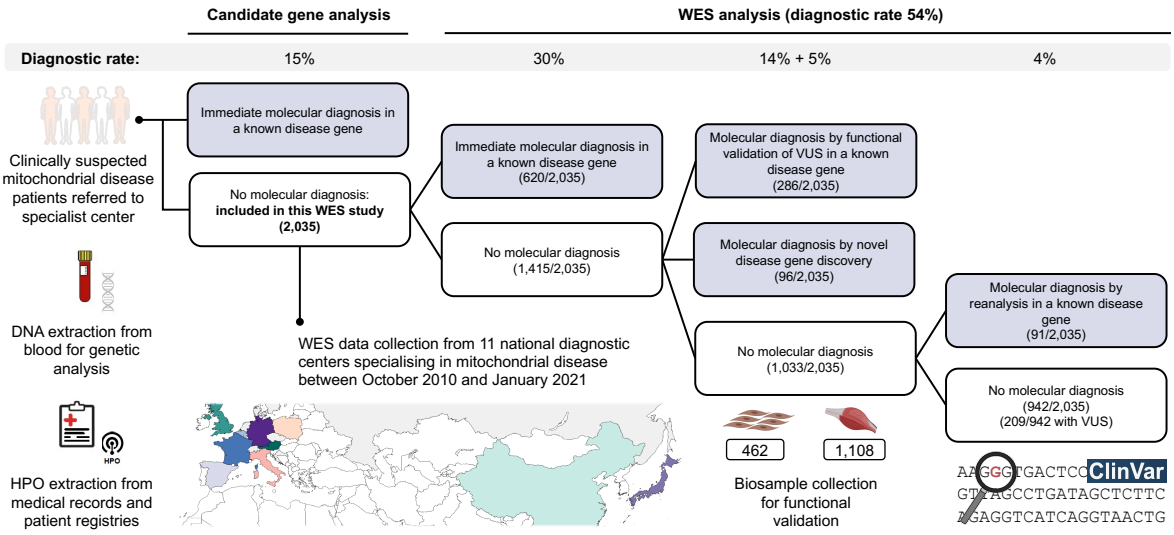


Figure 3.2: Diagnostic workflow for clinically suspected mitochondrial disease.
Based on the experience of the specialised mitochondrial disease diagnostic centres, an estimated 15% of paediatric patients were reported to have received a genetic diagnosis by targeted candidate gene sequencing (≤5 individual genes or mtDNA sequencing), the majority of which have mutations in the mtDNA. This study focused on the 2,035 patients investigated by WES.

		Total	MD	MD-phenocopy	Unsolved
Patients		2035	800/2035 (39%)	293/2035 (14%)	942/2035 (46%)
Gender	Male	1075/2035 (53%)			
Onset	Neonatal	554/2035 (27%)			
	Infantile	714/2035 (35%)			
	Early childhood	452/2035 (22%)			
	Childhood	214/2035 (11%)			
	Juvenile	110/2035 (5%)			
Diagnostic rate	Study population	1093/2035 (54%), 94/1093, 9% *de novo*	800/2035 (39%), 33/800, 4% *de novo*	293/2035 (14%), 61/293, 21% *de novo*	
	Proband-only WES, *de novo*	972/1791 (54%), 60/972, (6%) *de novo*			
	Trio-based WES, *de novo*	121/244 (50%), 34/121, (28%) *de novo*			
Median HPO phenotypes (range)		8 (2-40)	8 (2-40)	7 (2-30)	8 (2-35)
Median systems involved (range)		4 (1-12)	4 (1-12)	4 (1-11)	4 (1-11)
Abnormal RCC activity		894 (44%)	397 (50%)	99 (34%)	398 (42%)
Elevated lactate (serum and/or CSF)		1026 (50%)	507 (63%)	88 (30%)	431 (46%)
MRI abnormality (BG and/or BS)		526/2035 (26%)	274/800 (34%)	41/293 (14%)	211/942 (22%)
MDC score	Morava et al., 2006	6.0 ± 2.3	6.7 ± 2.4	5.2 ± 2.2	5.8 ± 2.2
	Witters et al., 2019	4.9 ± 2.0	5.4 ± 2.1	4.5 ± 1.8	4.7 ± 2.0

Table 5: Key demographic, genetic, clinical, and biochemical features.
Features stratified by genetic diagnosis. Neonatal (≤28 days), infantile (>28 days, <2 years), early childhood (≥2 years, <5 years), childhood (≥5 years, <12 years), juvenile (≥12 years, <18 years). BG, basal ganglia; BS, brain stem.

In autosomal dominant (AD) and X-linked dominant (XLD) disease genes, a pathogenic variant (P, ACMG class 5) or a likely pathogenic variant (LP, class 4) was reported as a definite genetic diagnosis. In autosomal recessive (AR) and X-linked recessive (XLR) disease genes, P/LP variants confirmed to be biallelic by segregation analysis were reported as definite genetic diagnoses. Patients with mitochondrial depletion or multiple mtDNA deletions and no identified causative variant(s) were not considered to be solved. This resulted in a diagnostic rate by WES of 54%[11], with a total

[11] The initial WES analysis was undertaken by clinicians and researchers at the respective diagnostic centres, for which a table of patients and their respective genetic diagnoses were provided on the variant level. Subsequent formal classification of all identified variants with the ACMG criteria was personally performed.

of 1,093/2,035 patients receiving a genetic diagnosis (see **Fig. 3.2** and **Fig. 3.3**). A dual diagnosis was identified in seven patients (0.3%). As indicated in **Fig. 3.2**, the diagnostic rate of 54% was reached by four different means:

1. Diagnosis by the immediate detection of ACMG classified P/LP variants in a known disease gene (620/2,035, 30%).
2. Diagnosis by functional validation of VUS in a known disease gene resulting in ACMG P/LP designation (286/2,035, 14%).
3. Diagnosis by the discovery and validation of variants in a novel disease gene resulting in ACMG P/LP designation (96/2,035, 5%).
4. Diagnosis by the reanalysis of WES data given detection of ACMG classified P/LP variants in a known disease gene, that was only reported to be disease-associated after the patient's initial WES data analysis, and was thereby only made possible be reanalysis of the data at a later time point (91/2,035, 4%).

VUS validation and novel disease gene discovery typically required patient-derived biomaterials for functional validation studies (e.g., patient-derived fibroblast cell lines). As depicted in **Fig. 3.2**, within our cohort, 462 patient-derived fibroblast cell lines (23% of the cohort) and 1109 muscle biopsies (54% of the cohort) were available for study and in total, 63 novel disease gene discoveries were made in the cohort over the 10 year period of the study.

Reanalysis was not undertaken at a single time point or after a set time interval following the initial WES data analysis, but rather *ad hoc* upon integration of WES data sets from a collaborating centre into the WES database of the Institute of Human Genetics (Munich, Germany), upon the description of a new disease-gene association in the literature prompting a database search for P/LP variants in unsolved cases, or as a byproduct of the analysis of more recently sequenced cases sharing P/LP variants in the same disease gene.[12] We can thereby only retrospectively indicate in which cases reanalysis was needed, by identifying cases with P/LP variants in a disease gene only described to be disease causing after the initial WES data analysis. On average, a 2.6 year interval (mean ± 1.7 year s.d.) was needed following the initial analysis to make a genetic diagnosis by reanalysis.

[12] Unsolved cases were reanalysed in this manner at the Institute of Human Genetics (Munich, Germany) by myself, Dr. Bader Alhaddad, Tekla Wolstein, Robert Kopajtich, and Mirjana Gusic. An automated reanalysis of all cases for homozygous ACMG classified P variants in AR and XLR OMIM disease genes was also personally undertaken (see Chapter 2.2.1) to diagnose five additional cases. No further dual diagnoses were detected in this manner.

To provide a theoretical comparison between targeted panel sequencing and WES diagnostic rate in the study, we applied a filter restricting the observed pathogenic variation to genes covered by 13 "mitochondrial disease" panels (see Appendices). The selected panels covered a total of 1,099 different disease genes and candidate disease genes (mean 285 ± 157 s.d. genes per panel) and consequently identified the causative variant(s) in 27% of patients (± 7% s.d.) (see **Fig. 3.3**), highlighting the benefit of selecting a exome-wide approach.

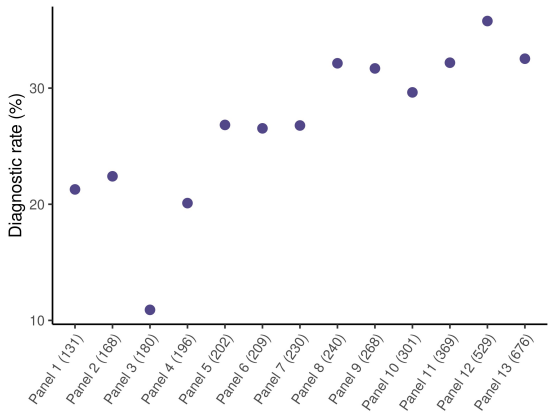


Figure 3.3: Theoretical diagnostic rate of targeted gene panels.
Theoretical diagnostic rate achievable by targeted gene panel sequencing, calculated by applying a filter restricting the observed pathogenic variation in the 2,035 patients to genes covered by 13 "mitochondrial disease" panels.

3.1.4 Molecular genetic aetiologies

In this study, MD genes were defined as per Schlieben and Prokisch 2020 (413 genes) [11], with the addition of three recently identified MD disease genes within this cohort (*DNAJC30*, *LIG3*, and *MRPL38*) [123, 118] (see **Fig. 1.3**). Disease genes not included in this definition within which defects were detected in the study, were subsequently termed "mitochondrial disease phenocopy" (MD-phenocopy) disease genes, given that the patients manifested clinically with MD, as such that the clinician suspected a MD genetic diagnosis.

In the solved cases, disease-causing P/LP variants were detected in 200 different MD genes (150 known and 50 novel) in 800/1,093 (73%) of the solved cases (see **Fig. 3.4**). While, in the remaining 293/1,093 solved cases (27%), the genetic diagnoses were identified in genes coding proteins for other organelles or compartments, thereby a MD-phenocopy gene (170 known and 13 novel). These MD-phenocopy disease genes were primarily implicated in neurodevelopmental disease, neuromuscular disease, or inborn

errors of metabolism. The remaining 942/2,035 (47%) cases remained unsolved. The unsolved cases can be further subdivided into 290/2,035 cases (14%) with prioritised variants of uncertain significance (VUS) in an OMIM disease gene or VUS in a candidate disease gene,[13] and 652/2,035 patients (32%) within which no variant could be prioritised for further consideration.

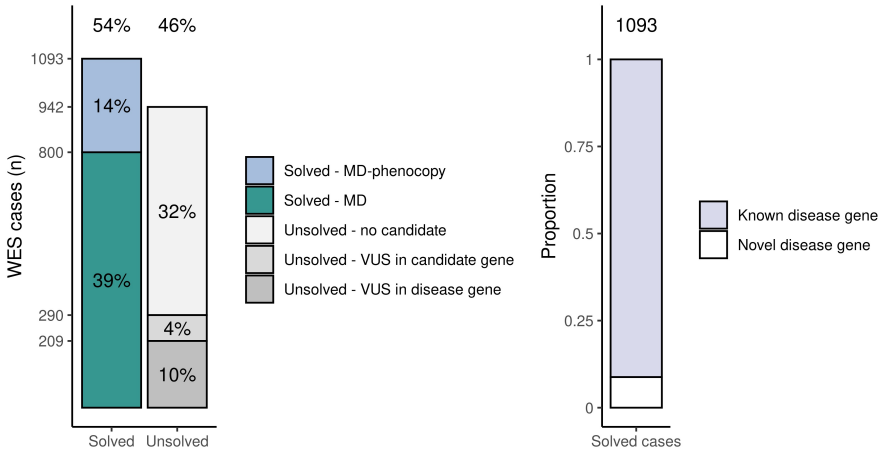


Figure 3.4: Diagnostic rate of WES stratified by genetic diagnosis.
Left, diagnostic rate in the 2,035 patients investigated by WES stratified by genetic diagnosis. Right, proportion of genetic diagnoses made in known disease genes or discovered and subsequently validated in novel disease genes.

The genetic defects spanned 383 genes and 1,150 unique variants in total. The most frequent MD gene hits were *PDHA1*, *ECHS1*, *ACAD9*, and *MT-ATP6*, each with ≥24 cases. Amongst the MD genes were 28 genes with defects across 174 patients where vitamin supplementation, replenishment of a critical cofactor, dietary modification, or idebenone therapy offer the potential for disease modifying treatment [64, 65]. The most frequent MD-phenocopy gene hits were *MORC2*, *MECP2*, and *POLR3A*, each with ≥6 cases. However, these larger collections of patients were the exception, and the majority of the genetic diagnoses (201/383, 52%) were reported in single cases only (see **Fig. 3.5**).

In 85% of confirmed MD patients (682/800), the causative gene was encoded in the nuclear DNA and the majority of genetic defects were inherited in an autosomal recessive (AR) manner (604/800 cases, 76%) (see **Fig. 3.6**). Of these AR cases, 51% cases carried homozygous variants (306/604). When searching for runs of homozygosity (see Chapter 2.2.1), a first- or second-degree parental relationship was suspected in

[13] An automated search for potential novel mitochondrial disease genes (see Chapter 2.2.1) was personally undertaken, variants in 12 genes were detected in >1 case and a further 30 genes in one case.

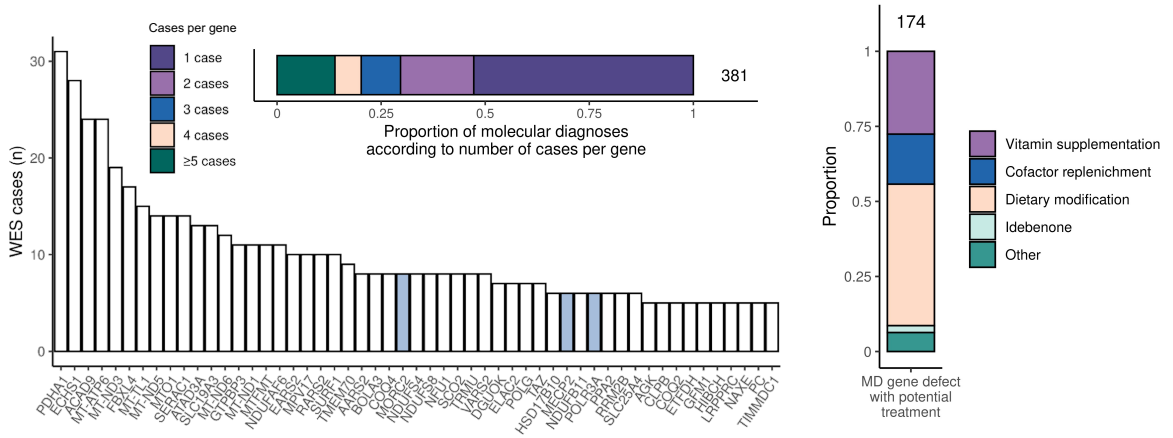


Figure 3.5: Molecular genetic aetiology of 1,093 solved cases.
Upper, proportion of genetic diagnoses according to the number of cases per gene. Lower, gene defects detected in ≥5 cases individually displayed with their frequency. The MD-phenocopy diagnoses are highlighted in blue. Right, 174 MD patients with defects in one of 28 gene reported to have a potentially beneficial treatment option, stratified by treatment type. Note, idebenone is only indicated in genetically solved patients presenting clinically with LHON.

42/2,035 cases, accounting for 35/306 cases with homozygous P/LP variants in MD genes (11%). 15% (118/800) of solved MD cases had a pathogenic mtDNA variant, detected in the WES as an off-target (see Chapter 2.2.1). Pathogenic mtDNA variants were mostly heteroplasmic (70/119 cases, 59%) and are reported with the individual heteroplasmy level in GENOMITexplorer. In MD-phenocopy patients, *de novo* dominant variants were responsible for disease in significantly more patients than in MD patients (p value 5×10^{-16}, Fisher's exact) (see **Tab. 5**). Notably, *de novo* variants were also recurrently detected in the MD gene *PDHA1* (14 cases) and in the mtDNA-encoded genes (seven cases).

In total, 1,471 P/LP variants were reported amongst our 1,093 solved cases. Only a small number of variants (324/1,471, 22%) were already reported to be P/LP in ClinVar or MITOMAP at the time the patient's WES data was analysed, facilitating an immediate genetic diagnosis (see **Fig. 3.7**).

Variant function spanned missense, near-splice, in-frame indel, protein-truncating (nonsense, splice, and frameshift), CNV, and mtDNA point or single deletion variants (see **Fig. 3.8**). Missense variants accounted for the largest proportion (54%) of all identified P/LP variants (799/1,471 variants), covering a total of 631 unique missense variants, of which the majority (84%) were reported in a single case only (533 variants). In 48% of solved cases at least one missense variant was involved in the genetic diagnosis (523/1,093) and 27% of solved cases carried biallelic missense variants (296/1,093). Missense variants are a challenging variant class in which to interpret pathogenicity,

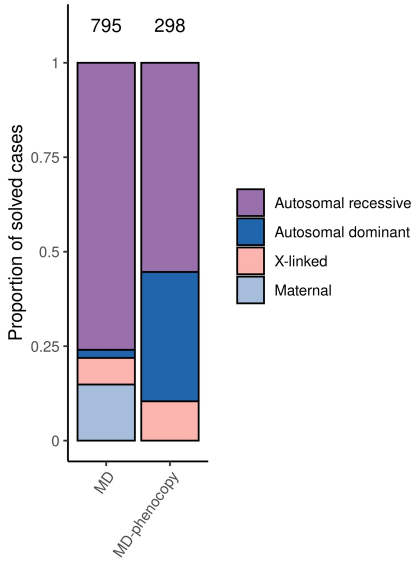


Figure 3.6: Inheritance pattern.
1,093 solved MD and MD-phenocopy cases stratified by inheritance pattern.

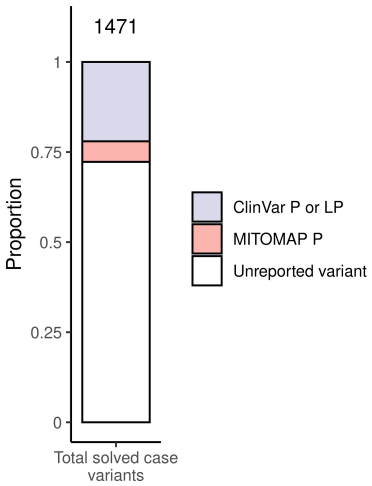


Figure 3.7: Detection of P/LP variants listed in ClinVar or MITOMAP.
Proportion of variants listed in ClinVar or MITOMAP as P or LP at the point of genetic diagnosis (324/1,471 variants).

especially when biallelic. The interpretation of missense variants thereby often necessitates functional studies for P/LP designation as depicted in **Fig. 3.8**. In contrast, biallelic predicted protein-truncating variants were identified in just 165/1,093 solved cases (15%). In 31/1,093 solved cases, pathogenic CNVs were identified, of which in four cases, the CNV was detected in *trans* with an SNV. All disease-causative variants are detailed on the patient-level with their ACMG classification on GENOMITexplorer.

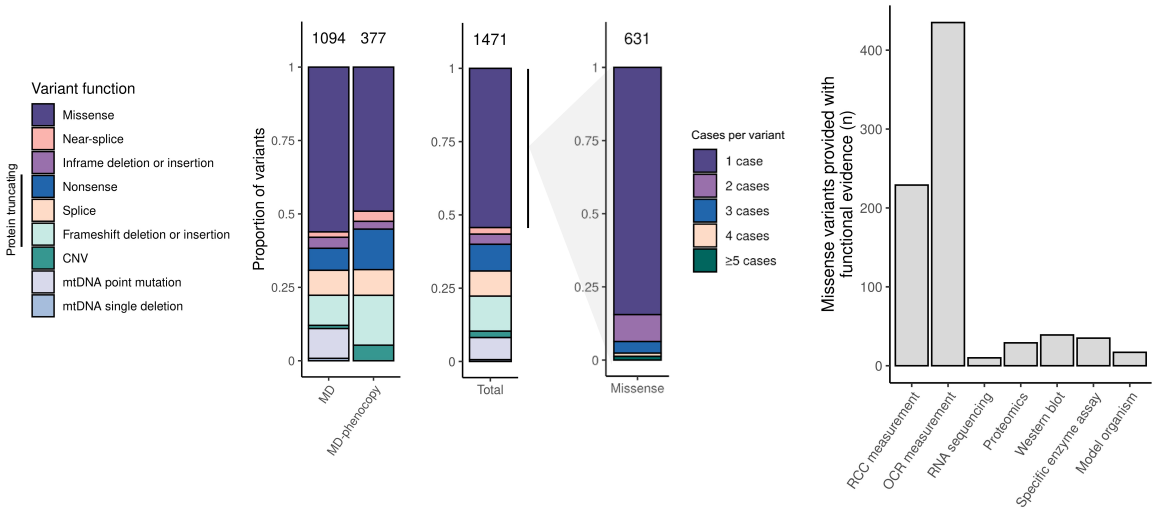


Figure 3.8: Variant function and validation.
Left, proportion of variants of each functional class stratified by genetic diagnosis. The majority of causative variants were missense and detected in single cases only necessitating functional validation for confirmation of pathogenicity. Right, functional validation assays utilised to assign pathogenicity to missense VUS. The total number of variants is given above each bar.

The gnomAD allele frequencies, CADD scores, and SIFT scores for all variants are displayed **Fig. 3.9**. Notably, two frequent near-splice variants with an allele frequency of 2.8% and 20.7%, respectively, were found as a rare combination in *cis* (only seen in this patient in our in-house WES database of >20,000 cases). RNA-seq analysis detected an aberrant splice event (exon skipping) in the corresponding exon. In combination with a second protein-truncating allele, these variants resulted in loss of the encoded protein documented by quantitative proteomics (as reported in [118] and described in detail in Chapter 3.4).

3.1.5 Association between phenotype and genotype

HPO encoded phenotypes were extracted for all 2,035 patients in this study. Phenotype analyses leveraged on the ontological tree of HPO terms to map all extracted HPO terms back to a common ancestor term. This facilitated the comparison of phenotypes between patients with different depths of clinical detail (see Chapter 2.2.6).

The majority of the 2,035 patients came to medical attention with neurological (77%), metabolic (76%), muscular (53%), digestive (29%), and/or cardiological (28%) symptoms. In the 800 solved MD patients, the most frequently reported clinical and imaging findings were neurodevelopmental abnormality (68%) and abnormality of MRI signal intensity in the basal ganglia and/or brain stem (52%). The most prevalent laboratory findings were increased serum lactate (63%) and abnormal activity of mitochondrial

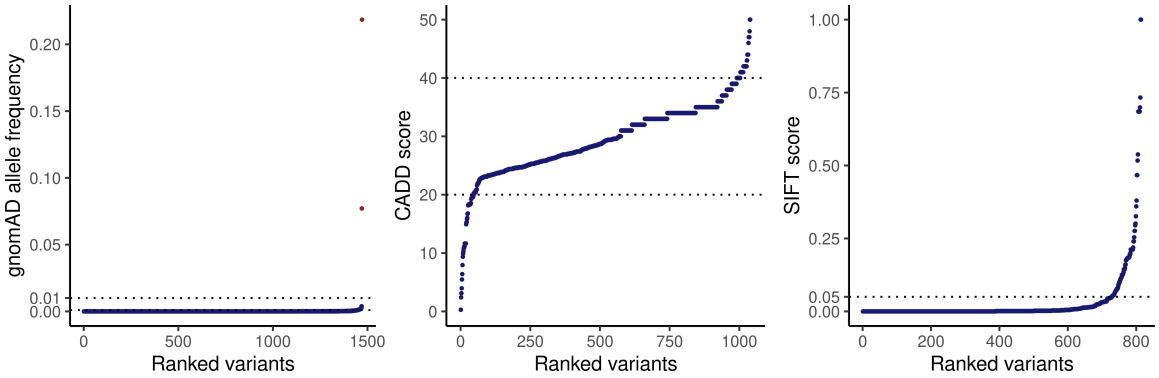


Figure 3.9: Variant allele frequency and *in silico* pathogenicity prediction. In rare disease and MD diagnostics, allele frequency thresholds of 1% and 0.1% are routinely applied, respectively, and variants with allele frequencies above these thresholds are considered benign (the two frequent near-splice variants, found as a rare combination in *cis*, are indicated in red). CADD scores ≥20 and ≥40 indicate the variant to be within the 1% and 0.1% most deleterious substitutions, respectively. SIFT scores <0.05 are predicted to be deleterious.

RCC enzymes (48%) (see **Fig. 3.10**). Across the entire cohort, 83 phenotypes were reported in at least 50 of the 2,035 patients. None of these 83 phenotypes were 100% specific for genetically-defined MD. However, 8/83 phenotypes were significantly enriched in the MD patients in comparison to the MD-phenocopy patients. This including phenotypes indicative of Leigh syndrome (encephalopathy, abnormality of the basal ganglia and/or brain stem, and increased serum lactate), mitochondrial RCCI defects, and cardiomyopathy (see **Fig. 3.10**). In combination, phenotypes indicative of Leigh syndrome were highly suggestive of MD (OR 4.2, 95% CI 2.4-7.6, p 1.7×10^{-9}, Fisher's exact test).

To objectively categorise the patients by clinical likelihood of MD, we applied the Morava MDC score [50] (see Appendices), demonstrating a continuum from possible (598/2,035, 29%, score 2-4) to probable (802/2,035, 39%, score 5-7) and definite (635/2,035, 31%, score 8-12) MD (see **Fig. 3.11**). Increase in the MDC score was reflected by increase in the overall diagnostic rate, increase in the proportion of MD gene defects, and decrease in the proportion of MD-phenocopy gene defects. Notably, a definite MDC classification (score 8-12) was reported in 50/293 MD-phenocopy patients, indicating these patients to be very difficult to distinguish from a true MD patient prior to genetic analysis. Given these data, no clear threshold of MDC score for the accurate diagnosis of a MD can be drawn in our cohort (see **Fig. 3.11**). The ability of the Morava MDC score [50] (AUC 0.68, 95% CI 0.65-0.72), the Witters MDC score [51] (AUC 0.64, 95% CI 0.60-0.67), and elevated serum lactate (AUC 0.68, 95% CI 0.64-0.7) in discerning between MD and MD-phenocopy in our cohort was limited. The Morava MDC score, incorporating tissue histopathology and biochemistry criteria,

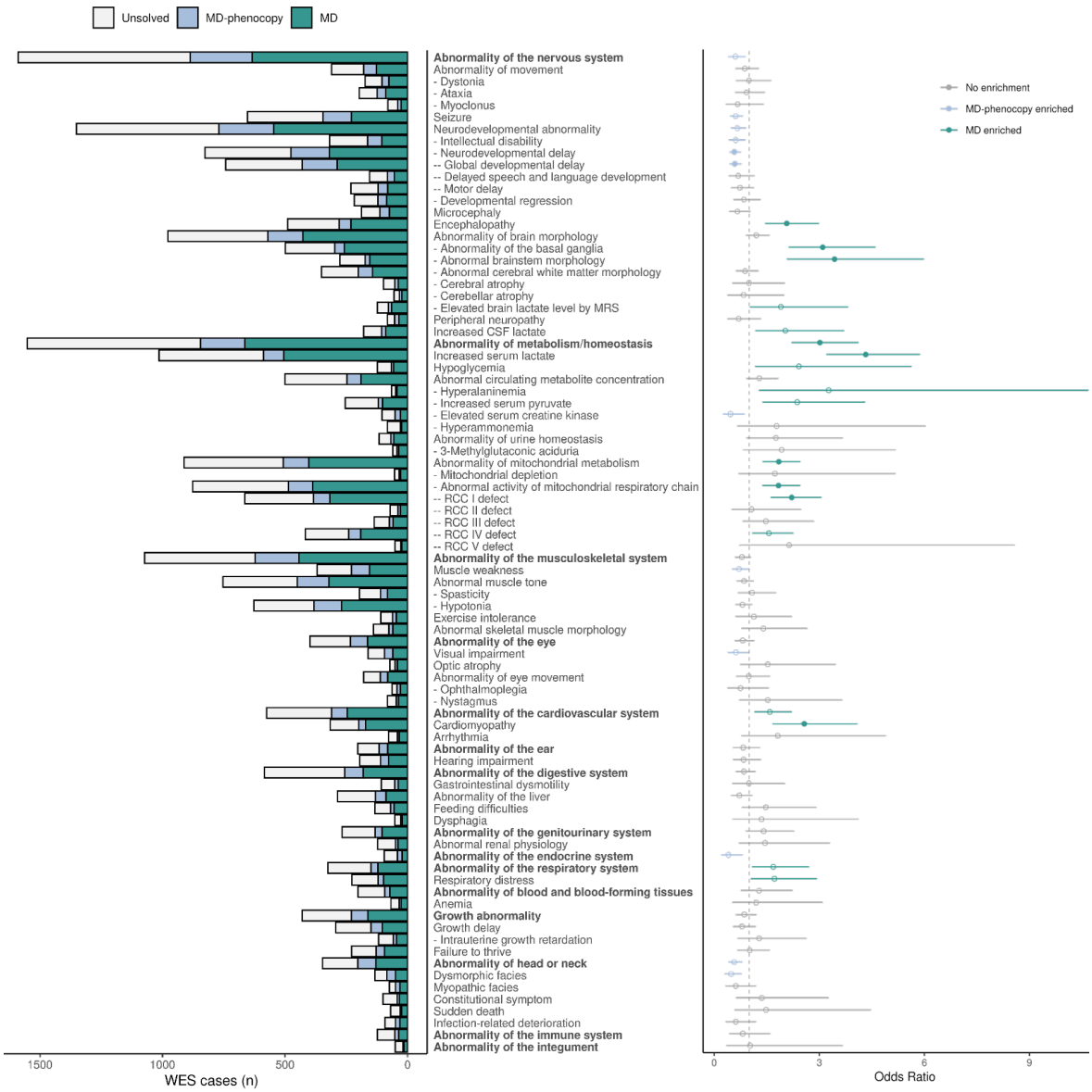


Figure 3.10: Clinical presentation and association with genetic diagnosis. Left, frequency of HPO phenotypes reported in ≥50/2,035 patients stratified by the underlying genetic diagnosis. System-level HPO terms are emboldened. Right, phenotype enrichment analysis comparing solved MD and MD-phenocopy patients. Nominally significant results are depicted in colour with unfilled shapes, phenotypes enriched with multi-testing corrected significance are depicted with filled shapes.

performed slightly better than the bedside Witters MDC score. However, these scores would be inadequate to confidently direct a targeted MD gene panel approach in our cohort without the risk of overlooking defects in disease genes beyond those captured by the panel (see **Fig. 3.11** and **Tab. 6**).

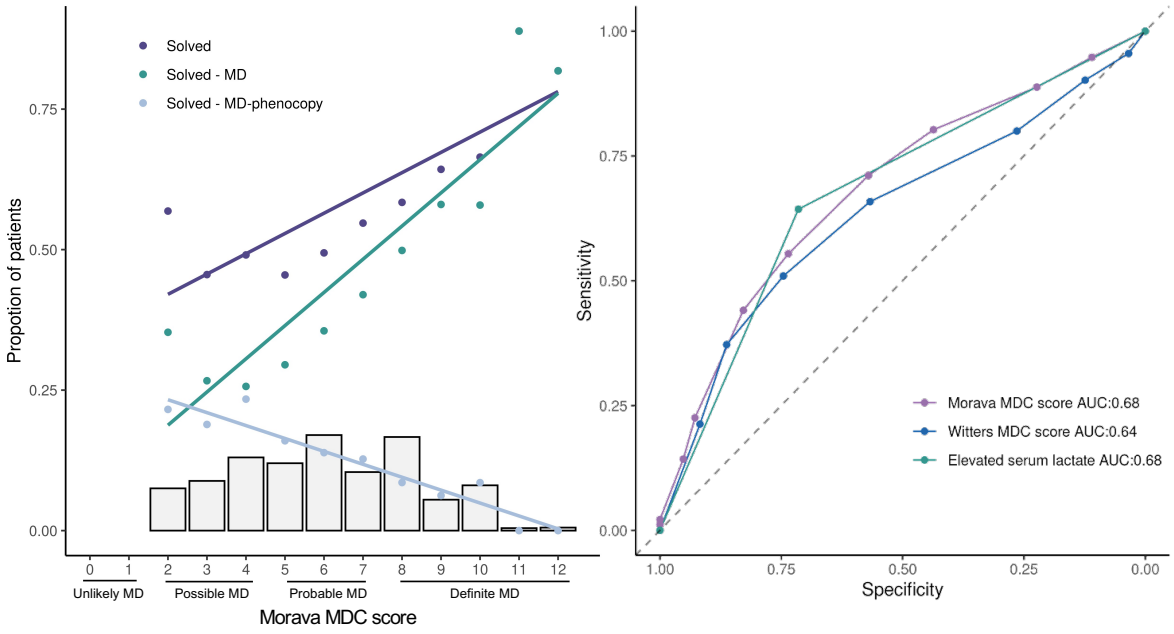


Figure 3.11: Mitochondrial disease prediction.
Left, distribution of Morava MDC score in the study population and the corresponding relationship to the proportion of solved cases, solved MD cases, and solved MD-phenocopy cases. Increase in the MDC score was reflected by increase in the overall diagnostic rate (R 0.83, p 3×10^{-3}, Spearman's rank correlation), increase in the proportion of MD gene defects (R 0.92, p 2.2×10^{-16}, Spearman's rank correlation), and decrease in the proportion of MD-phenocopy gene defects (R -0.96, p 2.4×10^{-6}, Spearman's rank correlation). The unexpectedly high diagnostic rate and proportion of solved MD at an MDC score of two (155 cases) reflects 143 legacy cases where only limited clinical information was available for analysis and 12 cases with single organ involvement (e.g., optic atrophy). Right, ROC curves evaluating the ability of the MDC scores and elevated serum lactate to detect MD in our cohort.

As described, 942 cases remained unsolved following WES analysis. In 209 of these unsolved cases (22%) WES analysis prioritised VUS in keeping with the expected mode of inheritance, of which 159 occurred in MD genes and 48 in candidate disease genes with mitochondrial localisation (according to Mitocarta3.0, see Chapter 2.2.7). A number of characteristics hint towards the pathogenicity of these VUS. First, most have at least a "probable" Morava MDC classification (score $\geq$5). Second, a number of phenotypes enriched in the solved MD patients in the study, including combinations of phenotypes indicative of Leigh syndrome, were reported in a significantly higher proportion of these cases than in the remaining unsolved cases (see **Fig. 3.12** and **Tab. 7**). However, current unavailability of fibroblast cell lines for these cases hindered

Predictive model	Model value	Sensitivity	Specificity
Morava MDC score	Score 1	1.00	0.00
	Score 2	0.95	0.11
	Score 3	0.89	0.22
	Score 4	0.80	0.44
	Score 5	0.71	0.57
	Score 6	0.55	0.74
	Score 7	0.44	0.83
	Score 8	0.23	0.93
	Score 9	0.14	0.95
	Score 10	0.00	1.00
	Score 11	0.01	1.00
	Score 12	0.02	1.00
Witters MDC score	Score 1	1.00	0.00
	Score 2	0.96	0.03
	Score 3	0.90	0.12
	Score 4	0.80	0.26
	Score 5	0.66	0.57
	Score 6	0.51	0.75
	Score 7	0.37	0.86
	Score 8	0.21	0.92
Elevated serum lactate		0.65	0.70

Table 6: Mitochondrial disease predictive model performance.
MDC score stratified by each individual score with corresponding sensitivity and specificity, in addition to elevated serum lactate calculated across the entire cohort of 2,035 patients. For elevated serum lactate, absence of a reported abnormality was considered as a normal result as negative findings were not reported in this study.

further functional studies to provide functional evidence of pathogenicity at this time, and the cases remain unsolved, highlighting the importance to acquire these bio-samples early in the diagnostic process.

Criteria	Odds Ratio (95% CI)	P value
Morava MDC score ≥5	1.53 (1.02-2.32)	0.032
Witters MDC score ≥5	1.26 (0.88-1.80)	0.219
Elevated serum lactate	1.54 (1.08-2.21)	0.014
MRI abnormality of the basal ganglia and/or brain stem	1.11 (0.72-1.68)	0.604
Leigh syndrome*	2.15 (1.33-3.43)	0.001
Cardiomyopathy	1.24 (0.72-2.06)	0.428
Mitochondrial RCC defect	1.63 (1.14-2.34)	0.006

Table 7: Enrichment analysis for clinical features indicative of mitochondrial disease.
Enrichment analysis for clinical features indicative of MD between unsolved cases with and without VUS in MD genes or candidate mitochondrially localised MD genes.

Akin to the MDC scores, the ability of HPO semantic similarity (SS) scoring to retrospectively predict the patient's genetic defect was found to be limited in pinpointing the correct genetic diagnosis (correctly predicted in just 4% of solved cases). Next, it was considered whether up to five genes could be sequenced by a targeted Sanger sequencing approach. However, in just 6% and 9% of patients the correct genetic diagnosis had the highest SS score with the patients reported HPO terms amongst all OMIM disease genes when considering the three and five highest SS scores,

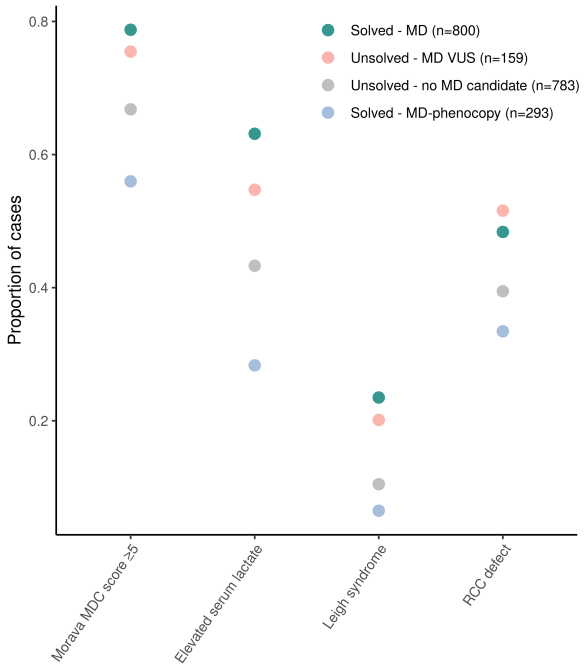


Figure 3.12: Comparison of key clinical features in solved and unsolved cases.
Proportion of patients meeting criteria found to be indicative of mitochondrial disease in the study population, stratified by underlying genetic diagnosis.

respectively (see **Fig. 3.13**). When considering the 10 and 100 highest SS scores, these figures increase to 14% and 50%, respectively.

These figures suggest that patient phenotype seems inadequate to direct a targeted candidate gene sequencing approach in the majority of suspected MD patients, and that diagnostic success depends on a selecting a exome-wide approach. Overall, more complex phenotypes (reflected by an increase in the number of HPO terms) correlated with improved SS scoring (see **Fig. 3.13**).

3.1.6 Development of the interactive web resource GENOMITexplorer

This study led to the creation of an open-access online resource "GENOMITexplorer" for the exploration of genotype-phenotype associations in MD. To enrich our data set of 2,035 patients from the WES study, additional HPO encoded phenotypes were extracted for 320 paediatric patients from mitochondrial disease registries (mitoNET and BESTA) and for 1,557 paediatric patients from the literature[14]. This resulted in a total of 3,912 patient HPO data sets, of which 2,970 have a genetic diagnosis. In

[14] Collection of MD cases from the literature and subsequent HPO encoding was undertaken with the support of Dr. Manting Xu and Dr. Rui Ban.

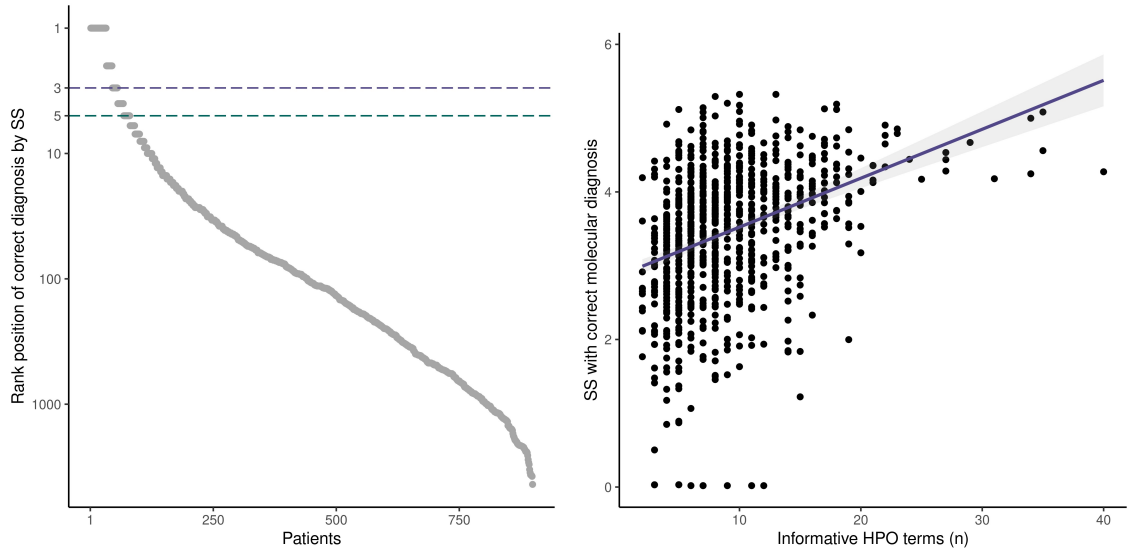


Figure 3.13: Ability of HPO phenotype to pinpoint the correct diagnosis.
Left, solved patients ordered by the rank of their respective disease-causing gene amongst all OMIM disease genes by HPO phenotype semantic similarity (SS). Right, correlation between the number of informative (non-redundant) HPO terms and the rank position of the correct genetic diagnosis amongst all OMIM disease genes by HPO semantic similarity (R 0.39, p 2.2×10^{-16}, Spearman's rank correlation).

GENOMITexplorer, HPO terms are freely browsable and downloadable on both the individual patient-level and on the gene-level (wes-d9159.web.app/, see **Fig. 3.14**).

In total, HPO terms associated with 414 different disease genes are included (217 MD, 207 MD-phenocopy disease genes), of which 114 were curated for five patients or more (100 MD, 17 MD-phenocopy disease genes). The resource reports 1,121 unique HPO terms (non-redundant) across 16 organ systems. A median of seven HPO terms are reported per patient (range 1-39) across a median of four systems (range 1-12). A median of 13 unique HPO terms are reported in total per gene (range 1-142) across a median of five systems (range 1-16).

Moreover, all P/LP identified in the WES study are listed on GENOMITexplorer with their genomic coordinates, functional consequence on the gene product, allele frequency in gnomAD, CADD score, SIFT score, ACMG classification, functional evidence of pathogenicity, and phenotype SS score with the disease-associated phenotype reported in the literature and can be associated, on the patient-level, with the resultant HPO phenotypes providing a high level of genotype-phenotype association resolution.

Gene level HPO terms

S.L.Stenton

Summary

All gene level HPO phenotypes from the WES study

in addition to paediatric molecular confirmed patients from the mitoNET and BESTA registries

and paediatric molecular confirmed patients curated from the literature

Total number of genes

```
## [1] 414
```

Number of genes with ≥5 cases

```
## [1] 114
```

Search:

HPO terms associated with each genetic defect

	Gene	HPO term	HPO ID	Association
	ACAD9	All	All	All
206	ACAD9	Increased serum lactate	HP:0002151	84/95 (88.4%)
207	ACAD9	Decreased activity of mitochondrial complex I	HP:0011923	77/95 (81.1%)
208	ACAD9	Abnormal heart morphology	HP:0001627	74/95 (77.9%)
209	ACAD9	Cardiomyopathy	HP:0001638	73/95 (76.8%)
210	ACAD9	Abnormality of the musculature	HP:0003011	57/95 (60%)
211	ACAD9	Pediatric onset	HP:0410280	57/95 (60%)
212	ACAD9	Exercise intolerance	HP:0003546	44/95 (46.3%)
213	ACAD9	Global developmental delay	HP:0001263	22/95 (23.2%)
214	ACAD9	Morphological central nervous system abnormality	HP:0002011	22/95 (23.2%)
215	ACAD9	Abnormal circulating amino acid concentration	HP:0003112	19/95 (20%)

Showing 1 to 10 of 98 entries (filtered from 9,823 total entries)

Previous 1 2 3 4 5 … 10 Next

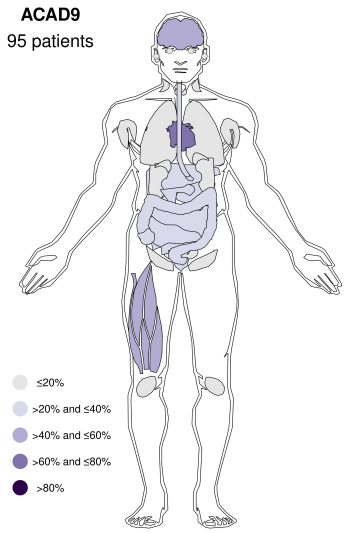


Figure 3.14: GENOMITexplorer interactive web resource.
Example page from GENOMITexplorer of the browsable table of gene level HPO phenotype associations with gene wise visual representation of organ system involvement for genes with ≥5 curated cases.

3.1.7 Summary of results

To summarise, in this study the diagnostic power of WES in >2,000 paediatric suspected mitochondrial disease patients was analysed, the study population was genetically and clinically characterised by integration of HPO encoded phenotype descriptors, and the value of functional data integration from patient-derived bio-material was explored. Overall, the study highlighted that interconnection of diagnostics and research at mitochondrial disease centres was the driving force for the high diagnostic yield of 54%, given the ability to undertaken functional validation studies to aid VUS interpretation, an overarching challenge in mitochondrial disease diagnostics.

3.2 *FDXR* variant validation expands the clinical spectrum of the disease

This study regards the validation of novel pathogenic variants in *FDXR*, an established mitochondrial disease gene primarily reported in association with sensorial neuropathies [124, 166, 167] and was published as "Expanding the clinical and genetic spectrum of FDXR deficiency by functional validation of variants of uncertain significance" by **Stenton S.L** et al. *Human Mutation* 2020 (DOI: 10.1002/humu.24160).[15]

3.2.1 Identification of *FDXR* variants

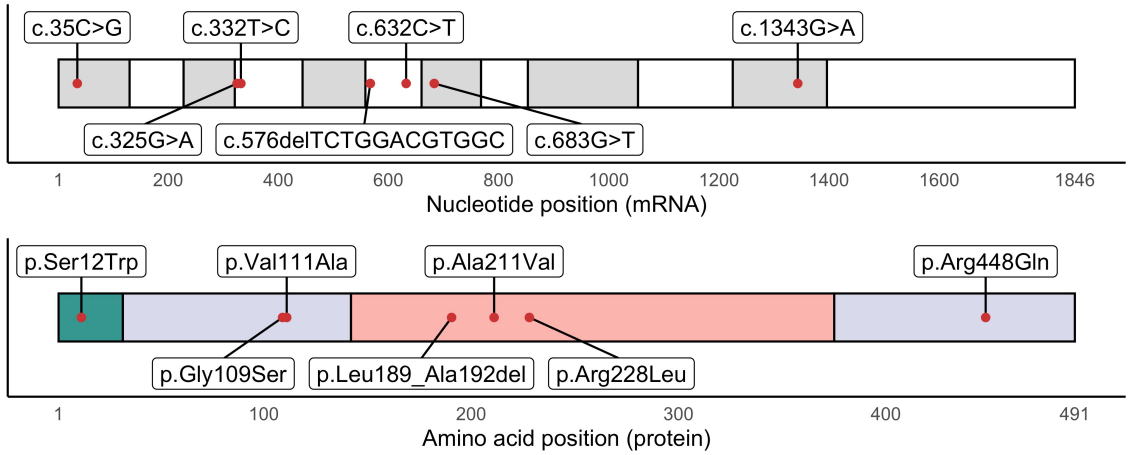
WES analysis of unsolved suspected mitochondrial disease patients at the Institute of Human Genetics (Munich, Germany) identified three unrelated patients with rare potentially biallelic variants in the known mitochondrial disease gene *FDXR*. *FDXR* encodes ferrodoxin reductase, an enzyme essential to the biosynthesis of iron-sulfur (Fe-S) clusters required by mitochondrial CI, II, and III, and thereby integral to electron transport in mitochondrial respiration [168, 169]. Submission of the gene to GeneMatcher (see Chapter 2.2.7) identified a further four unreported patients from three families. An overview of these patients is presented in **Tab.8**.

Amongst the seven patients, a total of nine unique variants were identified. Seven of these variants were novel and all were missense or in-frame indel in nature; thereby variants of uncertain significance according to the ACMG classification (US, class 3) [106]. The seven novel variants were rare with MAF ranging from 0 to 1.59×10^{-5}, and no homozygous carriers documented in the gnomAD database (see **Tab. 9**). All were positioned in residuals highly conserved across species, and were annotated with

[15] This project was led by myself and Dr. Holger Prokisch. All experiments and analyses presented in this chapter were personally performed unless otherwise indicated.

scores indicating a damaging nature by *in silico* variant pathogenicity prediction tools, CADD, SIFT, and PolyPhen (see **Tab. 9**). These variants are depicted on the mRNA and protein structure in **Fig. 3.15** along with their conservation across species.

In five families (F2-F6), parental DNA was available for sequencing. Segregation analysis confirmed biparental transmission of the variants with the exception of family 4 (F4). In this family, one variant was paternally inherited and one variant arose *de novo* in the patient. In family 1 (F1), parental DNA was unavailable, however, a substantial reduction of the FDXR protein in the patient-derived fibroblast cell line, investigated by immunoblotting, provided strong evidence for the compound heterozygous nature of the variants (see Chapter 3.2.3).




```
AA position:          12                  109 111              189

HUMAN       ASRCWRWWGWSAWPRTRLPPA   AHSGRCAFWGNVEVGRDVTVP   AVILGQGNVALDVARILLTPP
BOS         APRCWRWWPWSSWTRTRLPPS   ARSDRCAFYGNVEVGRDVTVQ   AVILGQGNVALDVARILLTPP
MOUSE       APRCWHWWRWSAWSGLRPSPS   ARSDRCAFQGNVVVGRDVSVP   AVILGQGNVALDVARILLTPP
DROSOPHILA  -----MGINCLNIFRGLHTSS   AEHPRLRYFGNISLGTDVSLR   VTIVGQGNVAVDVARMLLSPL
YEAST       --------------MSFVQIR   NQKHKFSFVGGITIGKEILLK   VGIIGNGNVAVDITRVLISNQ

AA position:        211                 228                  448

HUMAN       HLERTDITKAALGVLRQSRVK   SRVKTVWLVGRRGPLQVAFTI   YAAIQALLSSRGVRPVSFSDW
BOS         HLEKTDITEAALGALRQSRVK   SRVKTVWIVGRRGPLQVAFTI   SAFIKALLDSRGVWPVSFSDW
MOUSE       HLEKTDITEAALGALRQSRVK   SRVKTVWIVGRRGPLQVAFTI   YVAIQALLSNRGVRPVSFSDW
DROSOPHILA  ALKTTDTTEYALEALSCSQVE   SQVERVHLVGRRGPLQAAFTI   YDAD--------KRVVTWDGW
YEAST       IWENTDISSLALNLLRRAPVK   APVKDVKLIARRDFVHSKFTN   ALSLENSIDLS-IKHTTWKDW
```

Figure 3.15: Seven novel *FDXR* variants.
Upper, variant position on the *FDXR* mRNA-sequence (NM_024417) and FDXR protein amino acid (AA) sequence (NP_077728) coloured by functional domain (mitochondrial targeting sequencing, green; NAD(P)-binding domain, purple; FAD/NAD(P)-binding domain, red). Lower, AA conservation across five species. The AA positions of the pathogenic substitutions are marked in red. *Adapted from [125].*

Family	F1	F2	F3		F4	F5	F6
Origin	Germany	Germany	Germany		Poland	France	Spain
Individual	**P1**	**P2**	**P3.1**	**P3.2**	**P4**	**P5**	**P6**
Gender	F	M	F	M	M	F	F
Genotype	p.Ser12Trp; p.Arg228Leu	p.Val111Ala; p.189_192del	p.Gly109Ser; p.Arg448Gln	p.Gly109Ser; p.Arg448Gln	p.Arg228Leu; p.Cys353Tyr	p.Ala211Val; p.Ala211Val	p.Arg386Trp; p.Arg386Trp
Age at onset	Childhood	7 months	4 years	4 years	2 months	4 years	5 months
Age last Examined	Childhood (alive)	11 months (died)	8 years (alive)	6 years (alive)	17 months (died)	4 years (alive)	4 years (alive)
Hearing impairment	Yes	No	Yes	No	Yes	Yes	No
Optic atrophy	Yes	Yes	No	No	Yes	Yes	Yes
Muscular phenotypes	Myopathy Hypotonia Ptosis Ophthalmo-plegia	Myopathy					Hypotonia
Neurological phenotypes	Dystonia Ataxia Peripheral neuropathy	Global developmen-tal delay Seizures Nystagmus	Microcephaly Developmen-tal regression Movement disorder Ataxia	Microcephaly Movement disorder Ataxia	Microcephaly Developmen-tal regression Progressive encephalo-pathy Hypotonia	Developmen-tal regression Spasticity Ataxia	Global developmen-tal delay Developmen-tal regression Non-ambulatory
MRI changes	Basal ganglia Bilateral optic atrophy	Basal ganglia Bilateral optic atrophy	Normal	Not performed	Delayed myelination Cerebral atrophy	Cerebral atrophy Cerebellar atrophy Bilateral optic atrophy	Bilateral optic atrophy
Infection related deterioration	No	Yes	Yes	Yes	No	No	No
Additional phenotypes	Type I diabetes mellitus Psychiatric disorder	Anemia Raised CSF lactate Respiratory insufficiency			Cataract Respiratory distress Feeding difficulties		Retinal dystrophy Strabismus
Modified NPMDS score	14	27	11	3	27	13	19

Table 8: Clinical spectrum of disease in seven unreported *FDXR* patients.
Demographics and clinical presentation. *Adapted from [125].*

3.2.2 Genetic analysis

A comprehensive summary of all reported *FDXR* variants

The identification of seven novel variants expanded the mutational spectrum of *FDXR* defects to a total of 31 unique variants (see **Tab. 9**). The MAF was <0.1% for all, in line with the expected threshold for pathogenic variants in mitochondrial disease, and no homozygous carriers were documented in the gnomAD database. The variants are distributed throughout the gene in 11 of the 12 exons (see **Fig. 3.16**). The majority

Variant NM_024417	Polypeptide NP_077728	gnomAD MAF	CADD score	SIFT score	PolyPhen score	ACMG class	Study
Missense							
c.35C>G	p.Ser12Trp	0.00E+00	23.2	0.04	0.953	US	Stenton
c.151T>C	p.Phe51Leu	0.00E+00	25	0	0.973	US	Peng
c.221C>T	p.Pro74Leu	1.59E-05	25.5	0	1	US	Peng
c.325G>A	p.Gly109Ser	4.00E-06	28.5	0.01	1	US	Stenton
c.332T>C	p.Val111Ala	0.00E+00	26.8	0	0.999	US	Stenton
c.427A>T	p.Ile143Phe	0.00E+00	22.7	0	0.984	US	Peng
c.463C>T	p.Arg155Trp	1.42E-05	24.4	0	1	US	Slone
c.472G>A	p.Val158Met	7.97E-06	26.2	0	0.999	US	Peng
c.578G>A	p.Arg193His	1.61E-05	31	0	0.999	US	Slone
c.613A>G	p.Thr205Ala	0.00E+00	26.5	0	0.989	US	Peng
c.619A>T	p.Ile207Phe	0.00E+00	29.5	0	0.999	US	Peng
c.632C>T	p.Ala211Val	0.00E+00	29.1	0	0.867	US	Stenton
c.643C>G	p.Leu215Val	0.00E+00	31	0	0.999	LP	Paul
c.683G>T	p.Arg228Leu	1.59E-05	32	0	1	US	Stenton
c.724C>T	p.Arg242Trp	3.19E-05	32	0	1	US	Paul
c.916C>T	p.Arg306Cys	2.89E-05	24.7	0.02	0.959	US	Paul
c.979C>A	p.Arg327Ser	0.00E+00	25.3	0	0.946	US	Paul
c.1058G>A	p.Cys353Tyr	2.12E-05	27.2	0	1	US	Peng
c.1102G>A	p.Asp368Asn	7.96E-06	24.8	0	0.931	US	Peng
c.1156C>T	p.Arg386Trp	1.80E-04	32	0.01	0.999	US	Peng
c.1208C>T	p.Pro403Leu	1.21E-05	27.6	0	1	US	Peng
c.1309G>T	p.Gly437Cys	8.03E-06	25.9	0	1	LP	Peng
c.1309G>A	p.Gly437Ser	8.03E-06	25.5	0	0.998	US	Peng
c.1343G>A	p.Arg448Gln	8.12E-06	32	0.03	0.926	US	Stenton
c.1429G>A	p.Glu477Lys	0.00E+00	24.7	0	1	US	Paul
Nonsense							
c.820A>T	p.Lys274*	0.00E+00	44			P	Peng
c.925C>T	p.Arg309*	4.13E-05	41			P	Peng
c.1255C>T	p.Gln419*	0.00E+00	41			P	Paul
Indel							
c.576_587del	p.Leu189_Ala192del	4.02E-06				US	Stenton
c.929delG	p.Ser310fs	0.00E+00				US	Slone
Start loss							
c.1A>G	p.Met1?	0.00E+00	21.9			US	Slone

Table 9: Summary of 31 reported *FDXR* variants.
Variant function, frequency, predicted pathogenicity, and reporting study; Stenton et al., [125], Peng et al., [166], Slone et al., [167], Paul et al., [124]. US, uncertain significance; LP, likely pathogenic; P, pathogenic. *Adapted from [125].*

are missense variants (25, 81%), with a lesser number of nonsense (3, 10%), indel (2, 7%), and start-loss (1, 3%) variants. The absence of biallelic loss-of-function (protein-truncating) variants indicate the gene to be essential for life.

The CADD score was >20 (range 21.4-44) for all variants (see **Tab. 9**), indicating all variants to be within the 1% most deleterious substitutions, and for this threshold to be reasonable to assume variant pathogenicity in *FDXR*. To query this, we inspected the gnomAD database of presumably healthy individuals for homozygous variants with a CADD score >20. In total, 14 homozygous *FDXR* variants were reported, seven rare (allele frequency <0.1%) and seven more frequent (allele frequency ≥0.1%), of which

five have been annotated with a CADD score >20 and are not reported to cause an overt disease manifestation. We can thereby conclude that the use of *in-silico* pathogenicity prediction tools, such as CADD, provides an indication of the pathogenicity of variants in *FDXR*, though are insufficient to reach a definitive genetic diagnosis without further functional evidence. This was reflected by calculation of the ACMG criteria, as prior to any functional studies, 23 of the 25 missense variants were classified as variants of uncertain significance (US, class 3), confirming the need for a functional validation assay in designating pathogenicity to these variants and reaching a diagnosis for the patients.

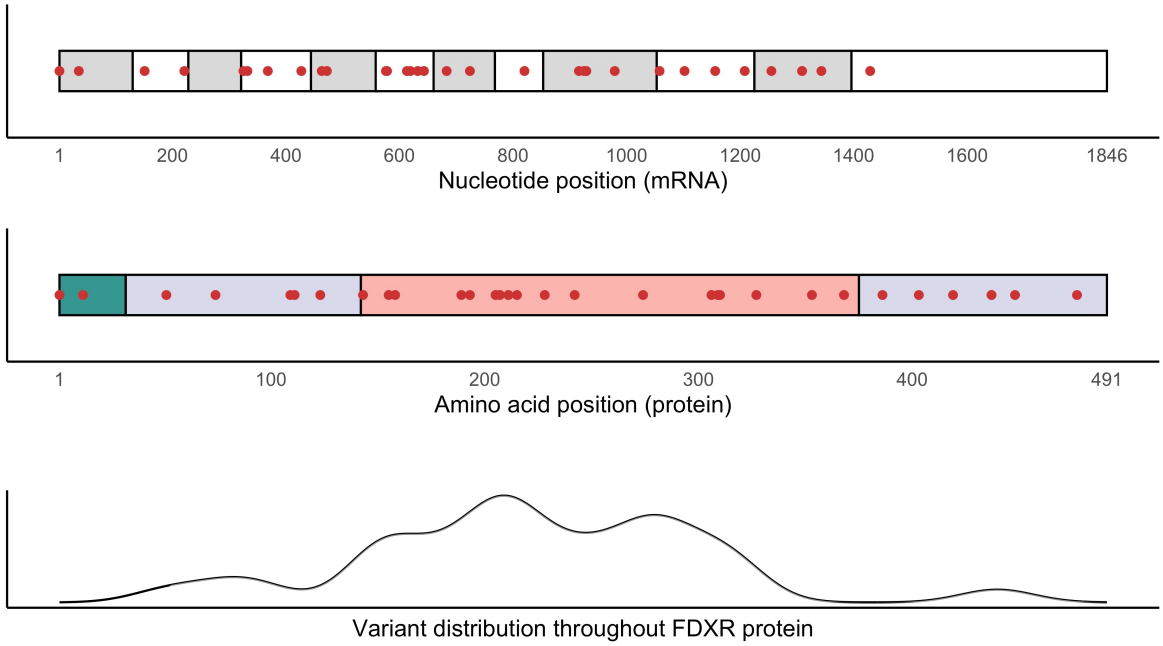


Figure 3.16: 31 novel and reported *FDXR* variants.
Upper, variant position on the *FDXR* mRNA (NM_024417) and FDXR protein (NP_077728) coloured by functional domain (mitochondrial targeting sequencing, green; NAD(P)-binding domain, purple; FAD/NAD(P)-binding domain, red). Lower, corresponding variant distribution on the protein AA sequence. *Adapted from [125].*

3.2.3 Functional analysis

Reduction of the FDXR protein on immunoblotting demonstrates variant pathogenicity

Patient-derived fibroblast cell lines were available for two patients, P1 with compound heterozygous p.Ser12Trp and p.Arg228Leu variants, and P2 with compound heterozygous p.Arg228Leu and p.Cys353Tyr variants. The pathogenic nature of the variants was confirmed by immunoblotting for the FDXR protein. This demonstrated a reduction

to 28% and 3% of residual FDXR protein in comparison to a control cell line, respectively (see **Fig. 3.17**).

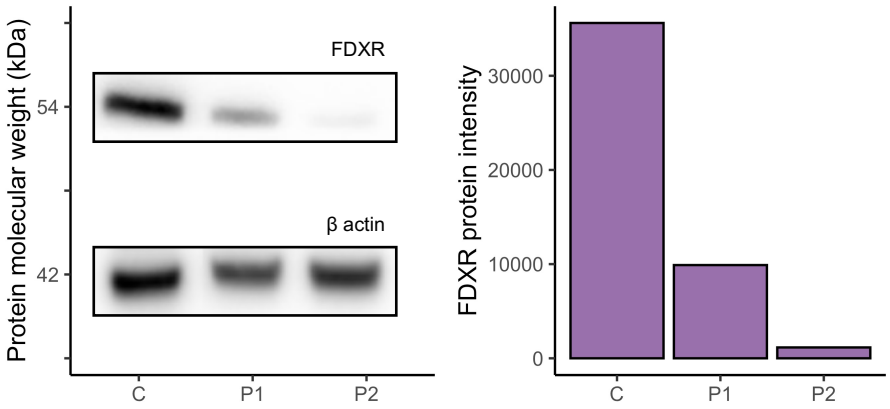


Figure 3.17: Immunoblotting for FDXR protein in patient-derived cell lines.
Left, immunoblotting in one control (C) and two patient-derived (P1 and P2) fibroblast cell lines using β actin as a loading control. Right, quantification of the FDXR protein band.

Impaired growth in a complemented *Arh1*-null yeast model demonstrates variant pathogenicity

Given the absence of patient-derived bio-material for the majority, we utilized an established *Arh1*-null yeast model [124] and complemented it with a plasmid expressing mutant FDXR protein forms to determine variant pathogenicity. The growth conditions, with predefined temperatures and substrates, were selected to progressively stress the cells by increase in temperature and transition from a fermentable substrate (YPD, with which cells can use glycolysis to generate ATP) to a non-fermentable substrate (YPG, with which cells require mitochondrial OXPHOS to generate energy) (see Chapter 2.2.4).[16] The selected conditions categorized the variants into 5-tiers of severity based on impaired growth with a (i) fermentable substrate at 30°C, (ii) fermentable substrate at 35°C, (iii) non-fermentable substrate at 30°C, (iv) non-fermentable substrate at 35°C, and (v) partially impaired growth with a non-fermentable substrate at 35°C. As a negative control, a common polymorphism (p.Arg123Gln) was selected. As positive controls, the empty plasmid and a known pathogenic *FDXR* variant p.Arg242Trp [124] were selected and confirmed to impair growth in the most favorable of the growth conditions (fermentable substrate at 30°C).

The growth assay confirmed disruption of FDXR function in association with all seven novel variants with variable degrees of severity as displayed in **Fig. 3.18**.

[16] Experiment performed with Robert Kopajtich and Lea Kulterer.

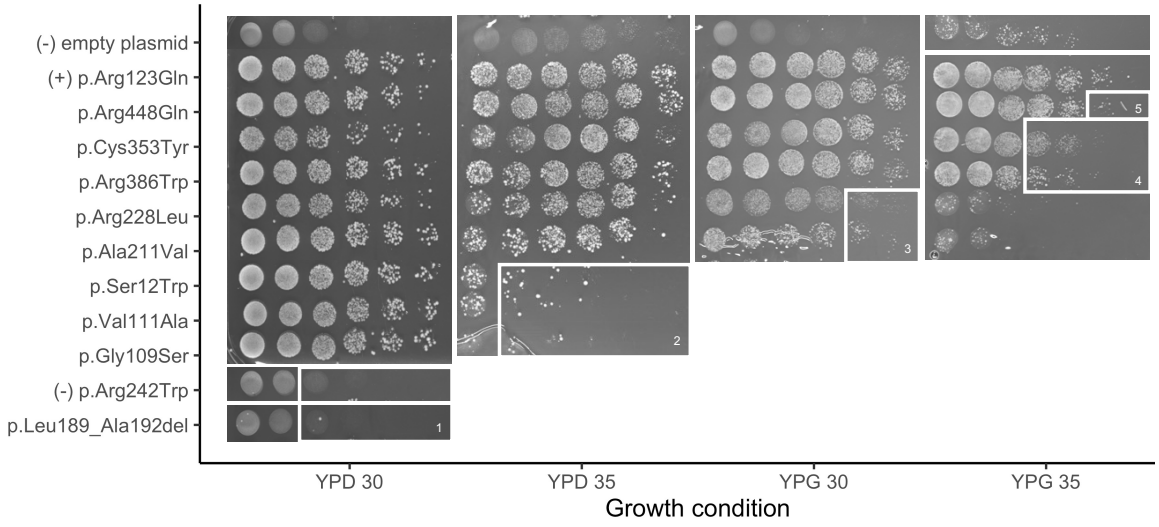


Figure 3.18: Growth assay utilising the complemented *Arh1*-null yeast model. Seven novel *FDXR* variants and two known pathogenic variants validated by impaired growth under four growth conditions. White boxes mark the severity of the growth defect. 1, most severe; 5, least severe; (-), negative control; (+), positive control. *Adapted from [125].*

3.2.4 Clinical analysis

Diverse phenotypes in the unreported patients expand the clinical spectrum of *FDXR* defects

The clinical spectrum of disease was diverse in the seven unreported patients (see **Tab. 8**). In part, their clinical presentation included the established *FDXR* associated phenotypes of optic atrophy in five patients (71%) and hearing impairment in four patients (57%) [124, 166, 167]. Additionally, a range of muscular, neurological, and ophthalmological phenotypes spanning myopathy, hypotonia, developmental regression, developmental delay, seizures, microcephaly, movement disorder, ataxia, spasticity, ptosis, ophthalmoplegia, and nystagmus was reported. Clinical characterization of the patients identified three phenotypes previously not reported in association with *FDXR* defects, inclusive of ptosis and type I diabetes mellitus, phenotypes known to be associated with other disorders of iron-sulfur protein metabolism such as Friedreich ataxia [170, 171]. Moreover, we confirmed the association of numerous phenotypes with *FDXR* defects previously reported in single cases only, inclusive of delayed myelination, dystonia, ophthalmoplegia, and cataracts (see **Tab. 10**).

Of our seven patients, two presented with disease more severe than currently reported. Patient P2 presented with a triad of clinical features indicating a diagnosis of Leigh syndrome [47]. Specifically, progressive neurological disease with motor and intellectual

developmental delay, signs and symptoms of basal ganglia disease, and raised lactate in the CSF, resulting in early demise at the age of 11 months. To our knowledge, this is the first patient to present with Leigh syndrome in association with biallelic *FDXR* variants. Patient P4 presented with infantile-onset progressive encephalopathy with delayed myelination, manifesting with microcephaly, global developmental delay, muscular weakness and hypotonia, severe visual and hearing impairment, and a overall failure to thrive resulting in early demise at the age of 17 months.

A comprehensive analysis of all reported *FDXR* patients reveals over 40 associated phenotypes

The identification of seven unreported patients expanded the clinical spectrum of *FDXR* defects to 43 unique phenotypes spanning eight organ systems, predominantly neurological, muscular, and ophthalmological (see **Tab. 10**).

Collating detailed clinical data from all 34 *FDXR* patients described to date, the most frequently reported phenotypes were found to be optic atrophy (29, 85%), visual impairment (23, 68%), neurodevelopmental delay (22, 65%, encompassing global, motor, and speech and language delay), developmental regression (16, 47%, of which 94% arose on the background of neurodevelopmental delay), hearing impairment (16, 47%), and hypotonia (15, 44%) (see **Tab. 10**).

The analysis of patient-derived muscle tissue from 12 patients demonstrated a mitochondrial RCC defect in six (50%). Mitochondrial CI was the most commonly defective (5/6, 83%), followed by CIII and IV (4/6, 67% and 3/6, 50%, respectively).[17] Given the role of FDXR in the biosynthesis of iron-sulfur (Fe-S) clusters, and the presence of such clusters in mitochondrial CI, II, and III, the limited number of defects in patient-derived muscle tissues is remarkable.

Calculation of a modified NPMDS score exposes a wide spectrum of clinical severity

The Newcastle Paediatric Mitochondrial Disease Scale (NPMDS) [52] was modified to provide an objective measure of disease severity comparable across all *FDXR* patients of different paediatric age groups (see Appendices). The median modified NPMDS score was 14 (range 2-27), representing mild to severe multi-systemic disease. The most severely affected patients, with an NPMDS score of 27, were those reported for the first time in the study with infantile-onset Leigh syndrome (P2) and familial encephalopathy with delayed myelination (P4) (see **Fig. 3.19**). Amongst the least

[17] Measurement performed at respective diagnostic centres.

Phenotype	Stenton et al., 2020	Reported	Total frequency
General			
Infection-related deterioration	2	10	35%
Failure to thrive	2	6	24%
Feeding difficulties	1	7	24%
Neuroradiological			
Cerebral atrophy	2	5	21%
Basal ganglia involvement	2	4	18%
Cerebellar atrophy	1	2	9%
Delayed myelination	1	1	6%
Leukoencephalopathy	1	1	6%
Neurological			
Neurodevelopmental delay	4	19	68%
Developmental regression	3 (2 with delay)	13 (13 with delay)	47% (94% with delay)
Global developmental delay	3	13	44%
Hypotonia	3	12	44%
Microcephaly	3	7	29%
Ataxia	4	6	29%
Spasticity	1	9	29%
Movement disorder	2	6	24%
Encephalopathy	1	5	18%
Polyneuropathy	1	4	15%
Abnormal EEG	0	5	15%
Motor developmental delay	1	3	12%
Non-ambulatory	1	3	12%
Psychiatric disorder	2	2	12%
Seizures	1	3	12%
Speech and language delay	0	3	9%
Dystonia	1	1	6%
Metabolic			
RCC abnormality	1	5	18%
Mitochondrial morphology	0	4	12%
Elevated CSF lactate	1	0	3%
MRS lactate	0	1	3%
Muscular			
Muscle weakness	3	6	27%
Myopathy	3	2	15%
Ophthalmoplegia	1	1	6%
Ptosis	1	0	3%
Ophthalmological			
Optic atrophy	5	24	85%
Visual impairment	5	18	68%
Nystagmus	1	7	24%
Retinal dystrophy	1	7	24%
Strabismus	1	4	15%
Cataract	1	1	6%
Other system involvement			
Hearing impairment	4	12	47%
Ventilation difficulties	2	5	21%
Anaemia	1	1	6%
Type I diabetes mellitus	1	0	3%
Total patients	7	27	34

Table 10: Complete clinical spectrum of disease in 34 reported *FDXR* patients.
Frequency of reported FDXR-associated phenotypes. *Adapted from [125].*

severely affected patients, with an NPMDS score ≤ 3, was a patient from the study (P3.2) with a mild movement disorder.

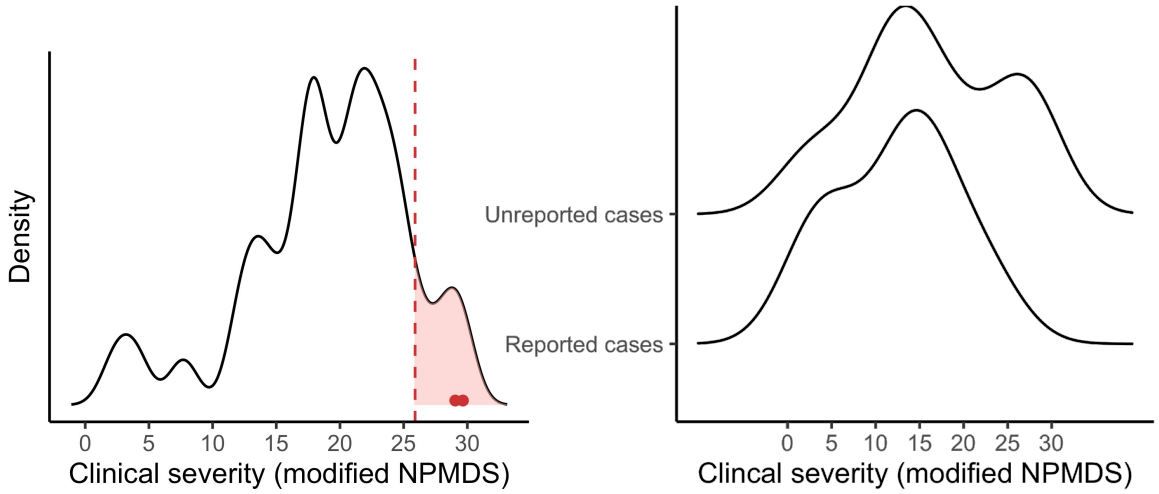


Figure 3.19: Severity of clinical phenotype in 34 FDXR patients.
Left, distribution of disease severity according to the modified NPMDS score in all 34 described FDXR patients, highlighting the 90th percentile (red). P4 and P6 are indicated as red points. Right, distribution of disease severity according to the modified NPMDS stratified by unreported and reported, demonstrating a second peak of more severely affected patients reported in this study.

3.2.5 Genotype-phenotype association analysis

Clinical severity can be determined by the underlying variant in a limited number of patients

Across the 34 reported patients, four unique *FDXR* variants were inherited in a homozygous manner, of which three arose in more than one patient. These specific circumstances allowed the analysis of phenotype severity associated with homozygous p.Arg306Cys (four patients), p.Asp368Asn (two patients), and p.Arg386Trp (eight patients) variants. For each of these variant, homozygous carriers demonstrated close similarity in modified NPMDS score (see **Fig. 3.20**). The p.Arg306Cys variant resulted in a mild phenotype (median modified NPMDS score 3, range 3-6), the p.Asp368Asn variant in a severe phenotype (median modified NPMDS score 21.5, range 21-22), and the p.Arg386Trp variant in a moderate phenotype with greater variability (median modified NPMDS score 14, range 8-19).

In the most common circumstance of compound heterozygous combinations of pathogenic variants, interpretation of severity is more challenging as it is dependent on the effect of two different alleles. However, in one patient reported by [124] to be heterozygous for the mild p.Arg306Cys variant in combination with a loss-of-function allele (p.Gln419*) a modified NPMDS score of two was calculated, indicating this

variant to be decisive in conferring a milder severity of disease, as displayed in **Fig. 3.20**.

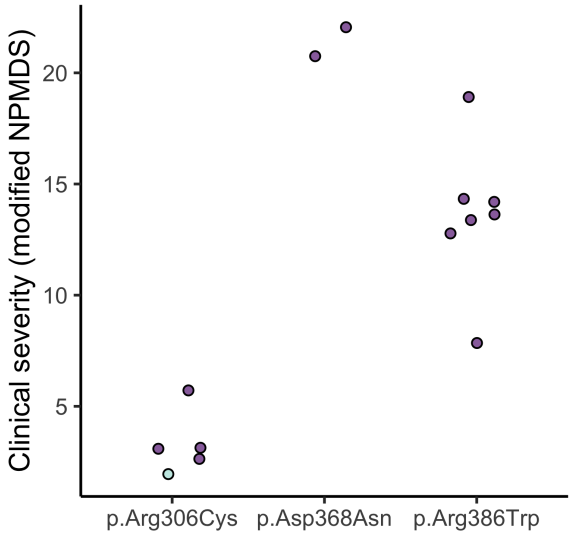


Figure 3.20: Severity of clinical phenotype in association with specific *FDXR* variants. Clinical severity score as calculated by the modified NPMDS in three groups of patients stratified by pathogenic variant (homozygous, purple; compound heterozygous, green). *Adapted from [125].*

Clinical severity cannot be determined by the degree of defective growth in the *Arh1*-null yeast model

In total, 10 unique pathogenic *FDXR* variants were validated in the *Arh1*-null yeast model. Nine of 10 were identified in our patients. In the vast majority of patients (5/7, 71%), these variants were inherited in a compound heterozygous manner. Similarly to the aforementioned analysis, compound heterozygosity presents a challenge in conducting a correlation analysis between clinical severity and yeast model growth impairment severity as each of the two variants may influence disease severity differently. We therefore addressed the question with three analyses, correlating the modified NPMDS score with the yeast model severity of (i) the least severe variant, (ii) the most severe variant, and (iii) the mean severity of the variants, according to the growth defect. All patients reported to carry one of the 10 variants in homozygosity in the literature (11 patients) were also included in the analysis, resulting in a total of 18 analysed patients.

The correlation analyses did not demonstrate significant positive correlation between clinical severity and variant severity as per the growth defect in the *Arh1*-null yeast model (Spearman's rank correlation coefficient, 0.07, 0.21, and 0.24, p=0.77, 0.41, and 0.33, respectively) (see **Fig. 3.21**). Though useful in validating variant pathogenicity,

this assay was therefore not sensitive in determining the resultant clinical severity of disease.

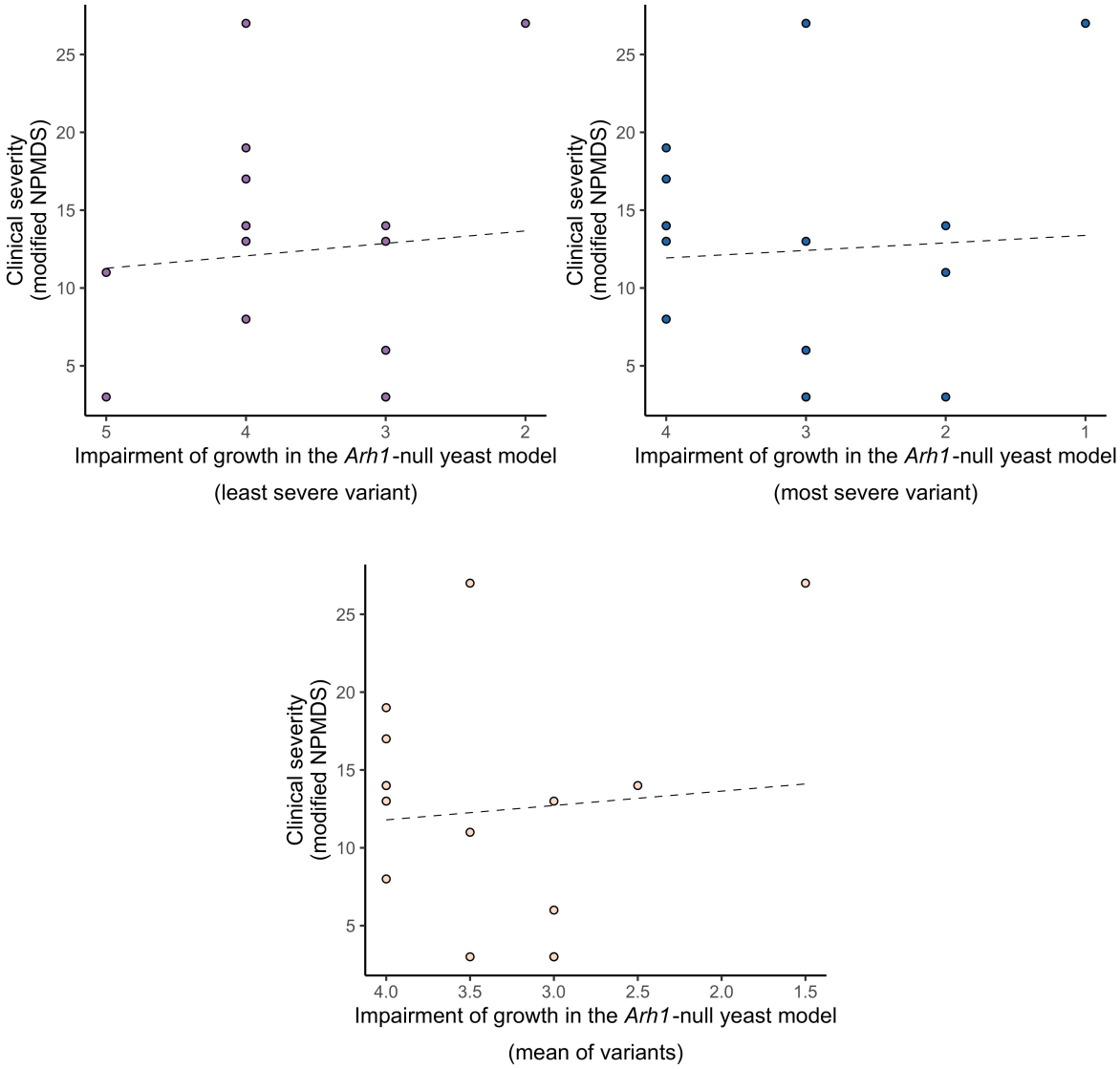


Figure 3.21: Correlation between clinical severity and growth impairment. Correlation of the modified NPMDS score with the severity of growth impairment (1, most severe; 5, least severe, reflecting the five different growth conditions) in 18 patients with biallelic variants tested in the *Arh1*-null yeast model. *Adapted from [125].*

3.2.6 Summary of results

To summarise, in this study the functional validation of seven novel disease-causing variants in the known disease gene *FDXR* demonstrated the use of an easily and

rapidly manipulated model organism to facilitate functional validation of VUS in the absence of patient-derived bio-material (e.g., patient-derived fibroblast cell lines, only available for 2/7 patients). This was a single gene focused approach necessitating variants to be prioritised prior to a targeted functional study. These variants were missense and in-frame indel in nature, challenging variant classes in which to predict functional consequence based on genetic data alone, and typically resulting in VUS designation by the ACMG criteria prior to functional studies. The study also provided an illustrative example of the variability in clinical spectrum and severity associated to a single mitochondrial disease gene and demonstrated how the phenotypic spectrum of a disease can be expanded by the continual identification and report of patients with a shared genetic defect.

3.3 *DNAJC30* variants cause impaired complex I repair and recessive LHON

This study regards the discovery of a novel mitochondrial disease gene responsible for a recessive form of LHON. The mechanism underlying the disease is found to be an impairment of mitochondrial CI repair due to pathogenic variants in the gene *DNAJC30*. The study was published as "Impaired complex I repair causes recessive Leber's hereditary optic neuropathy" by **Stenton S.L** et al., *Journal of Clinical Investigation* 2021 (DOI: 10.1172/JCI138267) [123].[18]

3.3.1 Identification of *DNAJC30* variants

WES analysis of suspected mitochondrial disease patients at the Institute of Human Genetics (Munich, Germany) identified a homozygous missense variant in *DNAJC30* shared by four patients across three families from the German Network for Mitochondrial Disorders (mitoNET).[19] Three patients were male with adult-onset LHON, a hereditary form of optic atrophy presenting with subacute, simultaneous or sequential, bilateral painless loss of central vision due to selective degeneration of the retinal ganglion cells (RGCs) and their axons, and one patient was female with childhood-onset Leigh syndrome, two distinct phenotypes that have been infrequently reported to result from a common genetic cause [46, 45]. Given that *DNAJC30* encodes a mitochondrially localizing protein and has previously been reported in association with impaired mitochondrial function [172], it was a promising candidate for further exploration.

[18] This project was led by myself, Dr. Ilka Wittig, and Dr. Holger Prokisch. All experiments and analyses presented in this chapter were personally performed unless otherwise indicated.

[19] WES variant prioritisation performed by Dr. Reka Kovacs-Nagy.

The identification of three LHON patients prompted screening of further unsolved LHON patients, gathered by international collaboration partners, primarily the Research Centre for Medical Genetics (Moscow, Russia) and the European Network for Mitochondrial Disorders (GENOMIT).[20] This screening process identified a further 29 patients, resulting in a collection of 33 patients across 29 families (see **Tab. 11**). In all patients, either complete mtDNA sequencing or mtDNA screening for pathogenic mtDNA variants was negative.

We therefore stratified LHON into LHON due to autosomal recessive variants in *DNAJC30* termed "arLHON" as compared to LHON due to maternally-inherited mtDNA variants, which we termed "mtLHON".

	Country of origin	Gender	Age of onset (years)	Age of last follow-up (years)	Clinical phenotype	*DNAJC30* variant	Haplogroup
P1	Germany	Female	2	21	LS	p.Tyr51Cys	K1a3a1
P2	Luxemburg	Male	19	25	LHON	p.Tyr51Cys	H4a1a1a
P3-1	Poland	Male	29	32	LHON	p.Tyr51Cys	J1c+16261
P3-2	Poland	Male	17	30	LHON	p.Tyr51Cys	J1c+16261
P4	Romania	Male	16	18	LHON	p.Tyr51Cys	H
P5	Russia	Male	20	-	LHON	p.Tyr51Cys	H2a2b1
P6	Russia	Male	14	16	LHON	p.Tyr51Cys	J1c3f
P7	Ukraine	Male	17	20	LHON	p.Tyr51Cys	H56
P8	Russia	Male	38	39	LHON	p.Tyr51Cys	T2c1d1a
P9	Russia	Male	13	24	LHON	p.Tyr51Cys	U5b3
P10	Russia	Male	13	15	LHON	p.Tyr51Cys	J1c8a
P11	Russia	Male	24	30	LHON	p.Tyr51Cys	U4a
P12	Russia	Male	15	23	LHON	p.Tyr51Cys	H15a1
P13	Russia	Male	19	25	LHON	p.Tyr51Cys	H15a1
P14	Russia	Male	28	30	LHON	p.Tyr51Cys	H4a1
P15-1	Russia	Male	17	28	LHON	p.Tyr51Cys	H2a
P15-2	Russia	Female	25	25	LHON	p.Tyr51Cys	H2a
P16	Russia	Female	40	49	LHON	p.Tyr51Cys	H6a1a
P17	Russia	Male	15	29	LHON	p.Tyr51Cys	U2e2a1a
P18-1	Tunisia	Male	16	25	LHON	p.Pro78Ser	H7c
P18-2	Tunisia	Male	Unclear	48	LHON	p.Pro78Ser	H7c
P19	Canada	Male	15	17	LHON	p.Tyr51Cys	-
P20-1	Turkey	Male	19	20	LHON	p.Leu101Gln	G2a2a
P20-2	Turkey	Male	12	31	LHON	p.Leu101Gln	G2a2a
P21	Russia	Male	21	22	LHON	p.Tyr51Cys	HV9b
P22	Russia	Male	22	23	LHON	p.Tyr51Cys	I1b
P23	Russia	Male	19	19	LHON	p.Tyr51Cys	U4a
P24	Russia	Male	16	17	LHON	p.Tyr51Cys	-
P25	Russia	Male	16	22	LHON	p.Tyr51Cys	-
P26	Russia	Male	40	40	LHON	p.Tyr51Cys	J1c2f
P27	Russia	Male	12	19	LHON	p.Tyr51Cys	-
P28	Ukraine	Male	12	12	LHON	p.Tyr51Cys	-
P29	Ukraine	Male	15	15	LHON	p.Tyr51Cys	-

Table 11: Summary of *DNAJC30* patients.
Patient demographics, clinical presentation, and genotype. All *DNAJC30* variants presented are homozygous. *Adapted from [123].*

[20] Sanger screening of unsolved LHON patients performed at multiple centres.

3.3.2 Genetic analysis

Three *DNAJC30* variants identified across 33 patients from 29 families

In 29 of the 33 patients, across 27 families, a homozygous NM_032317.2 c.152A>G (7:73,097 602, NP_115693.2 p.Tyr51Cys) missense variant was identified. This variant has a MAF of 0.125% (351:281,136 alleles) with no homozygous carriers in the gnomAD database (see Chapter 2.2.7). This allele frequency is greater than expected for a mitochondrial disease (defined as ≤0.1%), initially arguing for the variant to be benign (as per ACMG criterion BS1). However, given the overwhelming number of patients with LHON in association with this variant, we endeavored to validate its pathogenicity. In the remaining four of the 33 patients from two families, a homozygous c.232C>T (p.Pro78Ser) and a homozygous c.302T>A (p.Leu101Gln) missense variant were detected, respectively. Each of these variants is absent in the gnomAD database, indicating a MAF $<3.5 \times 10^{-6}$ given that the gnomAD database contains exome and whole genome sequences from 125,748 and 15,708 unrelated individuals, respectively (see **Tab. 11**).

A *DNAJC30* founder variant accounts for the majority of diagnoses and the frequent occurrence of arLHON in the founder population

The frequent occurrence of the p.Tyr51Cys variant prompted consideration of founder status. The majority of the 29 patients with this variant (26, 90%) originated from Russia, Poland, Romania, and Ukraine, indicating an Eastern European origin. Assuming founder status, the variant is estimated to have arisen 85 generations ago (95% confidence interval 43-168 generations). This figure was calculated by the genetic length of ancestral haplotypes shared between patients with available WES data as per [173] and is depicted in **Fig. 3.22**. Reflecting this finding, *DNAJC30* was confirmed to account for 21% (18/86) of all LHON cases at the Research Centre for Medical Genetics in Moscow. The remaining 79% (68/86) of cases were due to a pathogenic mtDNA variant. This argues for a lesser contribution of pathogenic mtDNA variants to LHON in the founder population, which is currently more generally reported to be 95% in LHON patients [41].

A higher than expected allele frequency and male predominance revealed gender-dependent incomplete penetrance

The MAF of the p.Tyr51Cys variant (0.125%) was higher than expected for a mitochondrial disease, defined as <0.1%, leading to uncertainty of variant pathogenicity. However, rather than considering the variant to be benign, we sought after and found

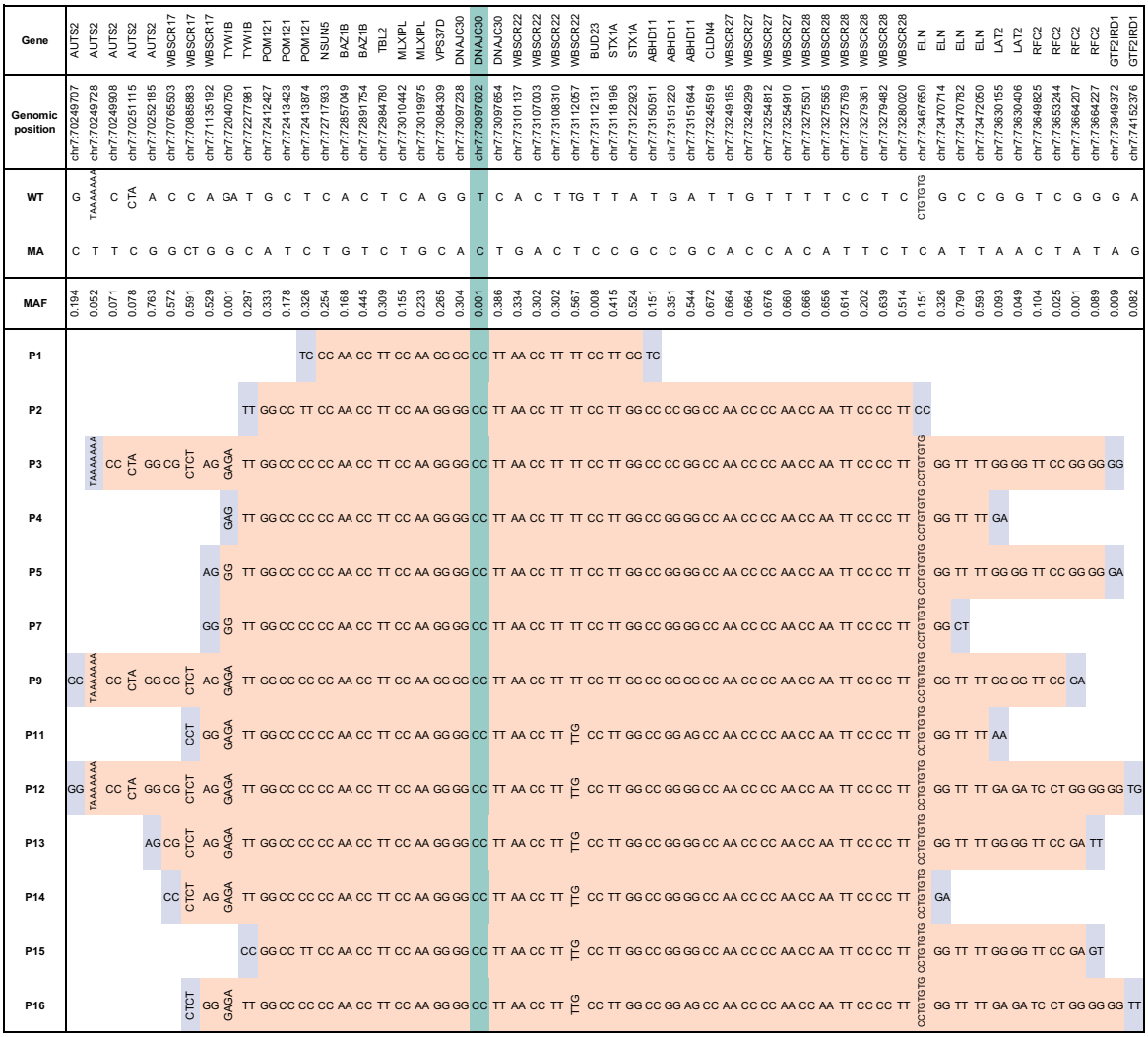


Figure 3.22: Genetic length of ancestral haplotypes shared between patients. Depiction of ancestoral haplogroups shared between patients (orange) in the region of the p.Tyr51Cys *DNAJC30* variant (green) for patients analysed by WES. WT, wild type allele; MA, mutant allele; MAF, minor allele frequency. *Adapted from [123].*

three lines of evidence indicated gender-dependent incomplete penetrance as an explanation:

1. Sequencing of *DNAJC30* in 19 unaffected siblings identified five asymptomatic homozygous carriers of the p.Tyr51Cys variant, resulting in an approximate penetrance of 97% in males (30/31) and 43% in females (3/7) and a significant male predominance of 10:1 ($p<0.001$, Fisher's exact test) (see **Fig. 3.23**).

2. The allele frequency of the variant in male and female individuals in the gnomAD database (see Chapter 2.2.7) is equal, arising in 184/152,888 alleles in males

(0.1%) and 167/128,248 alleles in females (0.1%), and would be expected to results in an equal number of affected male and female carriers if fully penetrant.

3. Screening of WES data from 1,036 patients with suspected inherited disease at the Research Centre for Medical Genetics (Moscow, Russia) identified 10 heterozygous carriers, indicating a MAF of 0.5% in the founder population, and predicting 2.3 per 100,000 homozygous carriers which would by far exceed the expected number of LHON patients if fully penetrant.

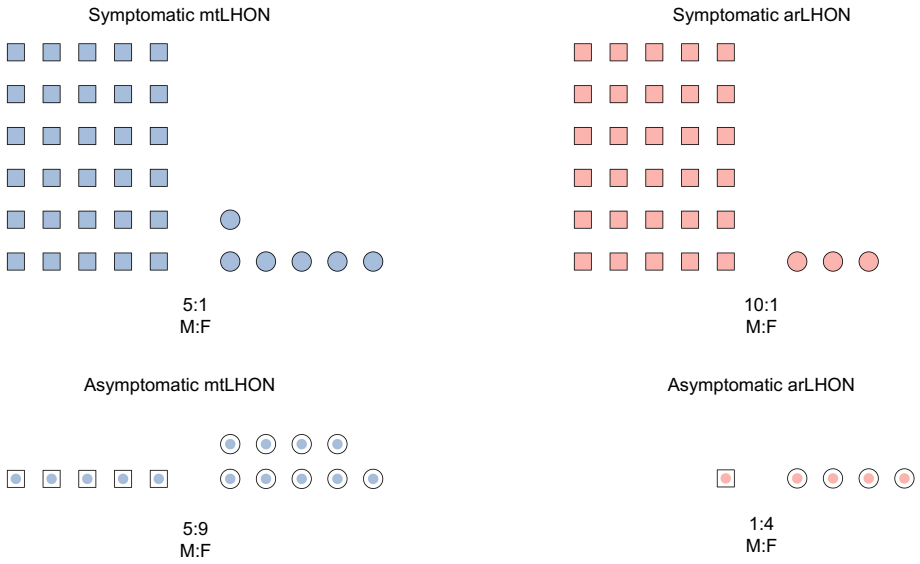


Figure 3.23: Gender-dependent incomplete penetrance in mtLHON and arLHON. mtLHON compared to arLHON with regards to gender-dependent incomplete penetrance. Squares represent males, circles represent females, completely filled shapes represent symptomatic carriers, partially filled shapes represent asymptomatic carriers. *Adapted from [123].*

Though the penetrance calculated from our pedigrees is expected to be an overestimate, due to the limited number of available siblings to screen and the inability to identify pedigrees where all carriers are asymptomatic in the diagnostic setting, we calculated the estimated prevalence of LHON caused by *DNAJC30* variants by applying the principal of the Hardy-Weinberg equilibrium ($p^2 + 2pq + q^2 = 1$, where p^2 = dominant homozygous frequency, $2pq$ = heterozygous frequency, and q^2 = recessive homozygous frequency) [174] to the gnomAD reported MAF of the variant (0.125%) with adjustment for our observed incomplete penetrance estimates (97% in males, 43% in females). This resulted in a prevalence estimate for symptomatic arLHON of 1.1 per million individuals, a figure falling below the estimated prevalence of mtDNA negative LHON patients [41], assuming up to 95% to be accounted for by pathogenic mtDNA variants

(calculated as 1.6 per million, 95% CI 1.2-2 per million), thereby supporting our assumption.

To investigate the underlying cause of the unequal penetrance between genders, we measured DNAJC30 expression on the RNA and protein level in 79 (42 male, 37 female) and 105 (60 male, 45 female) control fibroblast cell lines, respectively. This analysis did not find a significant difference in RNA expression (p=0.66) or protein expression (p=0.78) to account for the gender-dependent variable penetrance.

We also explored whether any mtDNA haplogroup, a combination of population-specific polymorphisms, influenced penetrance given that haplogroup J is reported to confer higher penetrance of mtDNA variants in mtLHON [175]. Five of 27 investigated arLHON patients (19%) were reported to have a haplogroup J genetic background. This is over the expected 9% for the European population [176], potentially indicating a role in increasing the penetrance of the p.Tyr51Cys variant. A further epidemiological study would be required to confirm this association.

DNAJC30 variants do not modify the penetrance of pathogenic variants in the mtDNA in mtLHON

To determine whether coding variation in *DNAJC30* plays a modifying role in the penetrance of pathogenic mtDNA variants responsible for mtLHON, we screened *DNAJC30* in 64 symptomatic and 53 asymptomatic homoplasmic pathogenic mtDNA variant carriers. In total, only five frequent variants were detected. The analysis did not reveal enrichment for common or rare *DNAJC30* variants in symptomatic carriers to provide such an explanation (see **Tab. 12**).

DNAJC30 variant	Symptomatic carriers (alleles)	Asymptomatic carriers (alleles)	Odds ratio (95% CI)	P value (adjusted)
c.34C>G (p.Leu12Val)	6/128 (5%)	4/106 (4%)	1.25 (0.29-6.21)	1 (1)
c.41C>T (p.Pro14Leu)	4/128 (3%)	5/106 (5%)	0.65 (0.13-3.12)	0.74 (1)
c.100G>A (p.Gly34Arg)	37/128 (29%)	42/106 (40%)	0.62 (0.35-1.11)	0.10 (0.40)
c.516C>T (p.Tyr172Tyr)	37/128 (29%)	44/106 (41%)	0.57 (0.32-1.02)	0.05 (0.25)
c.672T>C (p.Phe224Phe)	6/128 (5%)	10/106 (9%)	0.47 (0.14-1.50)	0.20 (0.60)

Table 12: *DNAJC30* screening in mtLHON patients.
Five frequent *DNAJC30* variants detected by screening symptomatic and asymptomatic pathogenic mtDNA variant carriers with corresponding frequency and enrichment. CI, confidence interval.

3.3.3 Clinical analysis

arLHON presents as a clinical phenocopy of mtLHON

The arLHON patients were clinically assessed by ophthalmologists and neurologists from national centres of expertise for LHON. The clinicians reported inability to distinguish arLHON from mtLHON patients on the basis of the ophthalmological findings (see **Tab. 13**). The pathognomonic triad of ophthalmological features hallmarking LHON [177] were documented in all patients, as follows:

1. Circumpapillary telangiectatic microangiopathy.
2. Vessel tortuosity of the central retinal vessels without leakage on fluorescein angiography.
3. Subacute phase swelling (pseudoedema) of the retinal nerve fiber layer (RNFL).

In each of the 32 LHON patients, the subacute phase of the disease was followed by an atrophic chronic phase with generalized thinning of the RNFL due to retinal ganglion cell (RGC) and axonal degeneration. No macular or peripheral retinal abnormalities were identified. MRI of the brain was reported normal in 19 of 22 investigated arLHON patients. In the three patients with abnormal findings, T2 signal hyperintensities were reported and were limited to the visual pathway (proximal optic nerve, the optic chiasm, and the optic tract).

The mean age of onset in arLHON was 19.9 years (± 7.9 years s.d.) (see **Fig. 3.24**). The median time from involvement of the first eye to involvement of the second eye was one week (range 0-2 years). 14 patients (47%) demonstrated bilateral involvement at onset. The median time from onset to nadir, the point of lowest visual acuity, was eight weeks (range 0-2 years). Clinically relevant recovery of visual impairment from nadir (CRR), defined as improvement in LogMAR (Logarithm of the Minimum Angle of Resolution) visual acuity of $\geq$0.2 [66], was observed in 42 eyes (68%, in 22 patients). In eight eyes (13%, six patients) visual acuity recovery was complete.

Idebenone therapy was received by 18 patients based on their clinical presentation of LHON. Idebenone is a potent antioxidant and electron donor approved for the treatment of LHON by EMA. It bypasses mitochondrial CI to restore downstream mitochondrial electron flow and respiration [66]. Of the 36 treated and 26 untreated eyes, CRR was reported in 29 (81%) and 13 (50%), respectively (see **Fig. 3.24**). The mean time from onset to first CRR was 13.0 months (± s.d. 10.4 months), and 25.8 months (± s.d. 30.3 months), in the treated and untreated eyes, respectively.

	Time to involvement of second eye (weeks)	Time to nadir (weeks, OD/OS)	Visual acuity at nadir (logMAR, OD/OS)*	Visual acuity at last visit (logMAR, OD/OS)*	CRR of VA (OD/OS)	Complete recovery of VA (OD/OS)	Visual field defect	Idebenone treatment
P2	12	40/32	1.60/1.40	1.68/1.68	N/N	N/N	CS	Y
P3-1	12	2/8	1.40/2.00	1.40/2.00	N/N	N/N	CS	Y
P3-2	Bilateral onset	13/13	2.00/2.00	0.30/0.40	Y/Y	N/N	CS	N
P4	1	52/52	1.68/1.68	1.10/1.10	Y/Y	N/N	CS	Y
P5	Bilateral onset	-	2.00/2.00	1.30/0.50	Y/Y	N/N	-	Y
P6	4	8/4	2.00/1.00	1.00/0.00	Y/Y	N/Y	-	Y
P7	4	8/8	2.30/2.30	1.00/0.50	Y/Y	N/N	-	Y
P8	Bilateral onset	12/12	1.00/1.00	0.10/0.10	Y/Y	N/N	CS	Y
P9	Bilateral onset	12/12	1.40/1.50	0.54/0.00	Y/Y	N/Y	CS	N
P10	8	16/16	1.10/1.10	0.00/0.00	Y/Y	Y/Y	CS	Y
P11	Bilateral onset	2/2	2.00/2.00	1.92/1.51	Y/Y	N/N	CS	Y
P12	Bilateral onset	8/8	1.40/1.40	0.30/0.90	Y/Y	N/N	CS	Y
P13	16	4/4	1.40/1.40	0.70/0.70	Y/Y	N/N	CS	Y
P14	Bilateral onset	16/16	1.40/1.40	0.00/0.10	Y/Y	Y/N	CS	Y
P15-1	Bilateral onset	8/8	1.90/1.20	0.18/0.18	Y/Y	N/N	CS	Y
P15-2	Bilateral onset	-	1.51/1.40	1.51/1.40	Y/Y	N/N	CS	N
P16	2	4/4	1.68/1.68	1.30/1.60	Y/N	N/N	CS	N
P17	Bilateral onset	4/4	1.10/1.10	1.31/1.11	N/N	N/N	CS	Y
P18-1	Bilateral onset	0/0	1.68/1.30	0.10/0.10	Y/Y	N/N	CS	N
P18-2	-	-	-	-	N/N	N/N	CS/PS	N
P19	24	4/4	0.60/1.30	0.70/1.30	N/N	N/N	CS	N
P20-1	Bilateral onset	52/52	1.00/1.00	1.00/1.00	N/N	N/N	CS	N
P20-2	-	-	-	-	Y/Y	N/N	-	N
P21	24	12/9	1.70/1.70	0.87/0.28	Y/Y	N/N	CS	Y
P22	1	12/11	2.00/1.40	0.59/0.41	Y/Y	N/N	CS	Y
P23	1	14/15	1.20/1.35	0.98/1.03	Y/Y	N/N	CS	Y
P24	24	24/1	-	0.27/0.02	-	-	CS	N
P25	Bilateral onset	14/14	2.00/2.00	0.00/0.00	Y/Y	Y/Y	CS	N
P26	Bilateral onset	-	-	1.00/1.00	N/N	N/N	CS	N
P27	48	-	-/1.11	0.18/0.00	Y/Y	N/Y	CS	N
P28	12	-	0.40/0.40	0.10/0.00	Y/Y	N/N	CS	Y
P29	Bilateral onset	-	-	-	N/N	N/N	CS	N

Table 13: Ophthalmological features of arLHON patients.
Off-chart visual acuity defined as LogMAR 1.68 and counting fingers/hand motion/light perception as LogMAR 2.0/2.3/2.6. OD, right eye; OS, left eye; CS, central scotoma; PS, peripheral scotoma. *Adapted from [123].*

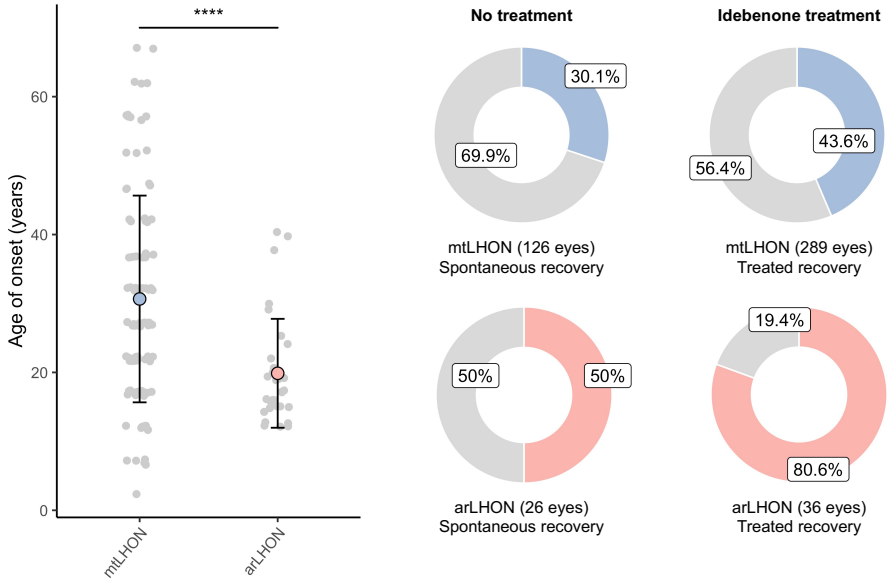


Figure 3.24: Significant clinical differences between mtLHON and arLHON. Left, age of onset in mtLHON and arLHON. Data are depicted as the mean ± s.d. Right, spontaneous and idebenone treated recovery in mtLHON and arLHON. *Adapted from [123].*

Though clinically challenging to distinguish, the statistical analyses revealed two significant differences between arLHON and mtLHON patients, as depicted in **Fig. 3.24**:

1. The age of onset in the arLHON patient was significantly earlier and more condensed (mean 19.9 ± s.d. 7.9 years) than reported for mtLHON (mean 30.7 ± s.d. 15.0 years, $p<0.001$) [178].

2. Clinically relevant recovery of visual acuity from nadir in response to idebenone therapy was significantly higher in the arLHON patients (29/36 treated eyes, 81%) than reported for mtLHON (126/289 treated eyes, 44%) ($p<0.001$, Fisher's exact test), in contrast to untreated eyes where a subtle non-significance difference was reported (13/26 untreated eyes, 50% in arLHON, 44/146 untreated eyes, 30% in mtLHON, $p=0.22$, Fisher's exact test).

DNAJC30 variants can present with LHON or Leigh syndrome

Notably, one patient with the homozygous p.Tyr51Cys missense variant presented with the typical constellation of clinical features suggestive of Leigh syndrome in the absence of optic involvement. This female patient presented at two years of age with spasticity, dysarthria, disturbance of gait, a moderate lactate peak on magnetic resonance spectroscopy (MRS), and bilateral necrosis of the putamen with lesions in the pedunculi cerebelli on brain MRI.

3.3.4 Variant consequence on protein abundance

All three *DNAJC30* variants occurred in a highly conserved area of the protein, the J domain (see **Fig. 3.25**). This domain belongs to a family of chaperone proteins and is key to their functional interactions [179].

The residuals affected by the p.Tyr51Cys and p.Leu101Gln variants are positioned in close proximity on the 3-dimensional protein structure. In association with these variants, quantitative proteomics on patient-derived fibroblast cell lines demonstrated protein degradation (2% and 7% of protein remaining for the p.Tyr51Cys and p.Leu-101Gln variants, respectively) (see **Fig. 3.26**). These amino acids are therefore likely to be fundamental to the structural integrity of the protein, and despite their missense character, result in loss-of-function. The p.Pro78Ser variant occurs in a crucial functional region of the J domain, the His, Pro, and Asp (HPD) tripeptide [179] (see **Fig. 3.25**), and is thereby predicted to disrupt function. Quantitative proteomics demonstrated normal protein abundance in association with the p.Pro78Ser variant (see **Fig. 3.26**).

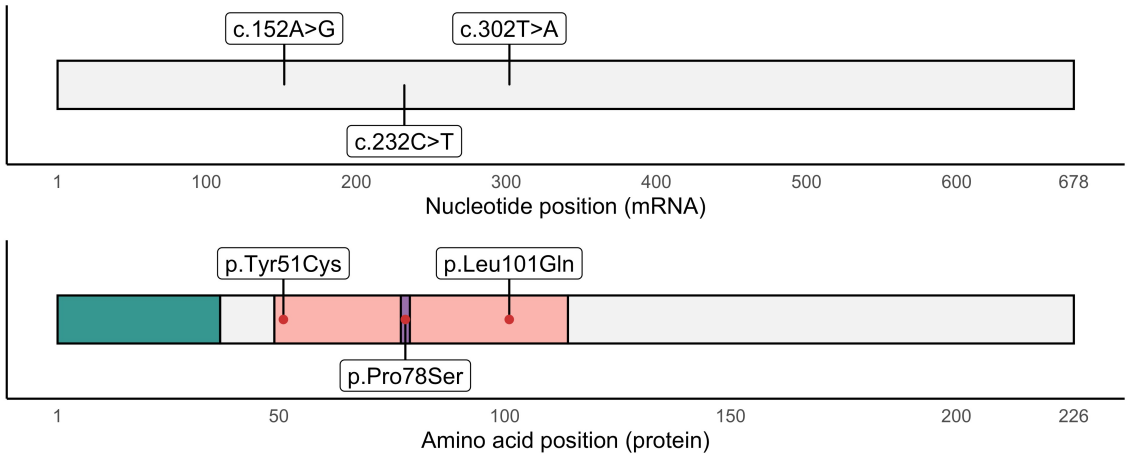
3.3.5 Measurement of mitochondrial CI function

Patient-derived skeletal muscle biopsies were available for three patients (P1, P2, and P7) and demonstrated an isolated mitochondrial CI defect (see **Fig. 3.27**).[21] A mitochondrial CI defect was further evidenced in patient-derived fibroblast cell lines (available for seven patients), spanning all three *DNAJC30* variants, demonstrating a consistent mitochondrial CI dependent respiration defect (mean 69% of control, s.d. 16%, $p<0.001$ Student's T-test) (see **Fig. 3.28**).

The mitochondrial CI dependent respiration defect was rescued by re-expression of naïve-*DNAJC30* in one patient-derived fibroblast cell line (P3-1) by lentiviral transduction (see Chapter 2.2.5). The CI dependent respiration defect was recapitulated in the *DNAJC30*-KO HEK cell line (mean 66% of control, s.d. 18%, $p<0.0001$ Student's T-test). In three mtDNA LHON patient-derived fibroblast cell lines a comparable magnitude of CI dependent respiration defect was measured as for arLHON (mean 59% of control, s.d. 20%, $p<0.001$ Student's T-test) (see **Fig. 3.29**).

Given the CI defect measured in muscle and fibroblasts, we sought to investigate the abundance and assembly of CI to determine whether DNAJC30 could be a structural subunit or assembly factor of CI.

[21] Measurement performed at respective diagnostic centres.




```
AA position:     51                      78                    101

HUMAN            ALYDLLGVPSTATQAQIKAAYYRQCFLYHPDRNSGSAEAAERFTRISQAYVVLGSATLRRKYDRGL
MOUSE            ALYELLGVPSTATQAQIKAAYYRQSFLYHPDRNPGSAEAAERFTRVSEAYLVLGSTILRRKYDRGL
LIZARD           GLYEILGVSQAATQAQIKTAYYKQSFLYHPDRNAGSEEAARHFTRVNEAYLVLGSVGLRKKYDRGL
CHICKEN          TLYEVLGVPPTATPAQIKAAYYQQSFRFHPDRNAGSAAAAERFATVSEAYRVLGSAALRRKYDLGA
FROG             AYYDILGVTGNATQTQIKTAYYKQSFRYHPDRNSGSDEATDRFGEISEAYSVLGSVSLRKKYDRGI
FISH             AYYDILEVSQSATHAQIKTAYYKQSFIYHPDKNAGSEDATHRFSQINEAYSVLGNKELRRKYDRGI
SHARK            AYYDLLQISPTATRSQIKTAYYKQSFIFHPDKNAGSEQAVLKFTQISEAYSVLSNGILRRKYDRGI
LANCELET         HYYDVLGVTPIATQTQIKSAYYKKSMQFHPDRNKT-AFAADRFTEITEAYEVLGHVDLRRKYDQGV
SEA HARE         HYYDVLELSPKATQNQIKSAYYKKSKLHHPDVSQRPES-HKLFTEINEAYEVLGNLRKRRMYDQGV
DROSOPHILA       NYYDALGIGKKCTQNEIKAAYYKLSMLYHPDRNQGSDSAAKKFRDINQAYEVLGNYRLRRLYDKGI
WATER FLEA       NHYKVLGISSKATQAEIKASFYKLSKLFHPDVSDQSEETAVKFRQITAAYEILGNLKLRKMYDRGL
TRICHOPLAX       NYYYVLGISPSATQAEIKASYYRLCKKYHPDVIGSSLVDHNKFTEITEAYSVLASVELRRKYDRSM
```

Figure 3.25: Identified pathogenic variants in *DNAJC30*.
Upper, three identified variants presented on the *DNAJC30* mRNA (NM_032317.2, 678 nt) and DNAJC30 protein (NP_115693.2, 226 aa). The mitochondrial targeting sequence (green) DNAJ domain (pink), and HPD tripeptide (purple) are highlighted on the protein. Lower, amino acid conservation across 12 species. The AA positions of thepathogenic substitutions are marked in red. *Adapted from [123].*

3.3.6 Measurement of mitochondrial CI assembly and abundance

We measured CI abundance in three patient-derived and 105 control fibroblast cell lines by quantitative proteomics. The abundance of mitochondrially localised proteins was compared between *DNAJC30* defect patients and controls by differential expression analysis (DEA) followed by gene set enrichment analysis (GSEA) for 149 mitochondrial pathways (MitoPathways3.0) (see Chapter 2.2.2). The analysis demonstrated a 12% increase in CI subunits in the patient cell lines (adjusted p=0.04, according to mean log2 fold change). There was no significant signal for other OXPHOS complexes to indicate a general increase in mitochondrial biogenesis. RNA-seq of three patient-derived fibroblast and 79 control fibroblast cell lines demonstrated no corresponding increase in CI subunit expression. The increase in CI subunits was recapitulated in the *DNAJC30*-KO HEK cell line (p=0.0002) by a quantitative proteomic analysis of DDM

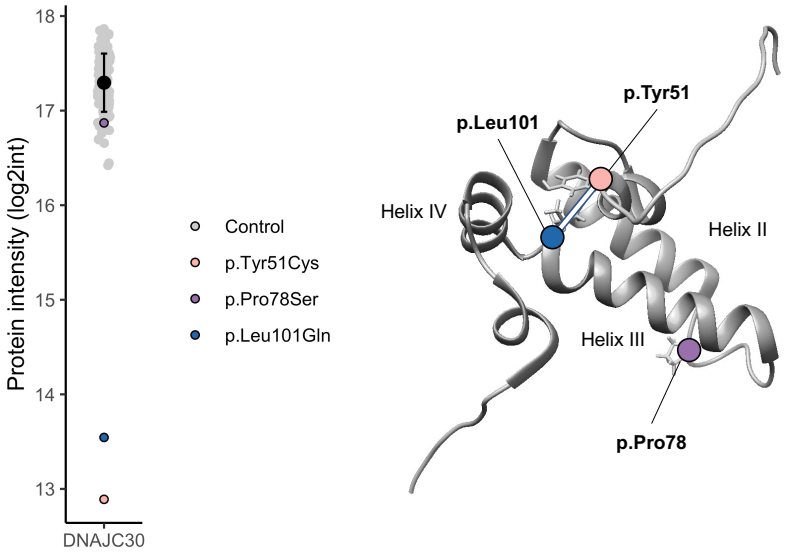


Figure 3.26: Variant consequence on the DNAJC30 protein.
Left, DNAJC30 protein abundance measured by quantitative proteomics in patient-derived fibroblast cell lines spanning all three reported *DNAJC30* variants in comparison to 105 control fibroblast cell lines depicted with the mean ± s.d. of controls. Right, DNAJC30 3-dimensional protein structure highlighting the affected residuals. *Adapted from [123].*

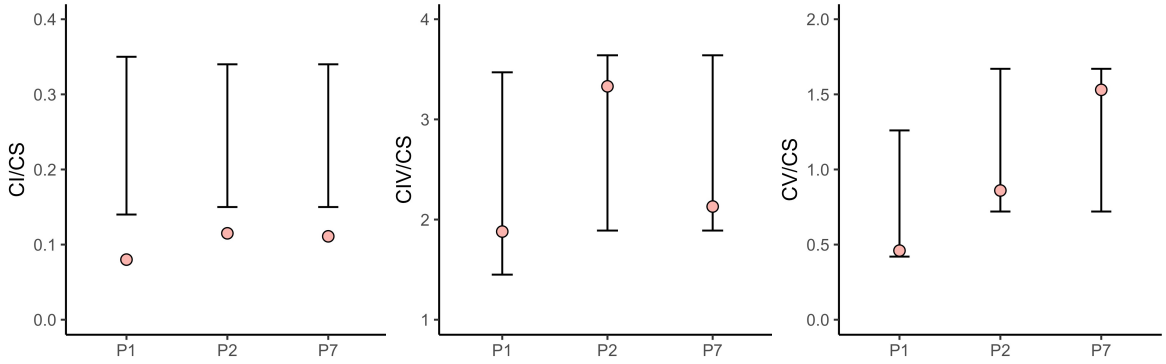


Figure 3.27: Mitochondrial RCC defects in patient-derived muscle biopsies.
Mitochondrial CI (left), CIV (middle), and CV (right) measurement in patient-derived muscle biopsies. Measurements are normalised to the citrate synthase (CS). The laboratory reference range is depicted by the bars. *Adapted from [123].*

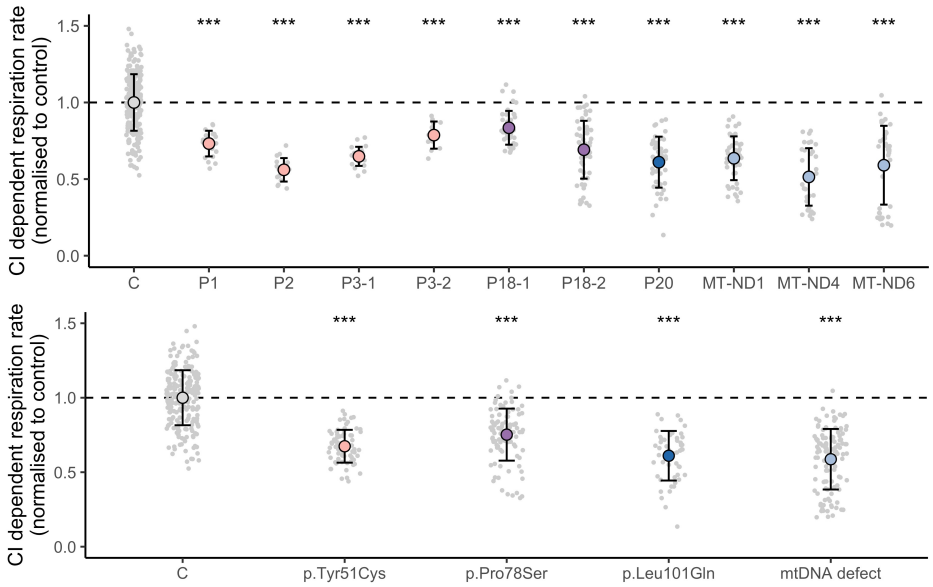


Figure 3.28: CI dependent respiration rate in patient-derived fibroblast cell lines. Upper, CI dependent respiration rate in seven individual *DNAJC30* defect patient-derived fibroblast cell lines and three mtLHON patient-derived fibroblast cell lines. Lower, CI dependent respiration rate grouped by genetic defect. Data are depicted as the mean ± s.d. C; control. *Adapted from [123].*

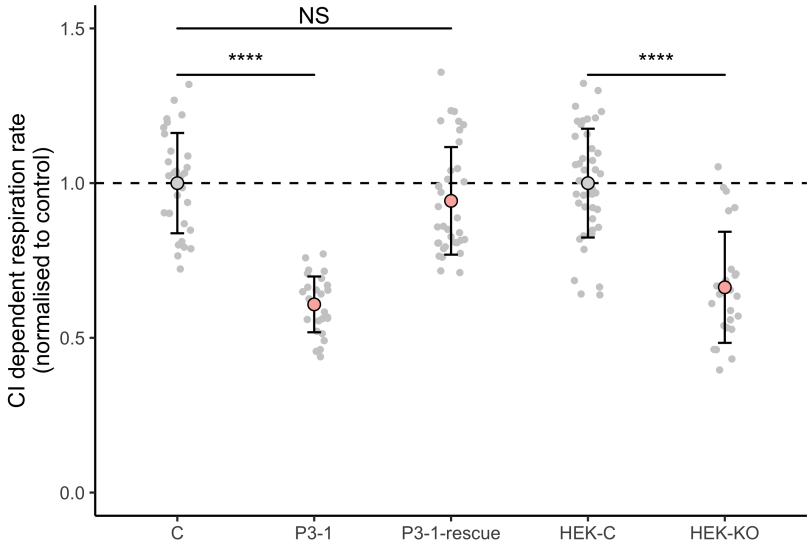


Figure 3.29: CI dependent respiration rate in rescued and *DNAJC30*-KO cell line. CI dependent respiration rate in a patient-derived fibroblast cell line ± rescue by re-expression of naïve-*DNAJC30* and in the control (HEK-C) and *DNAJC30*-KO (HEK-KO) HEK cell lines. Data are depicted as the mean ± s.d. *Adapted from [123].*

solubilised RCC separated by BNE. The abundance of the remaining mitochondrial RCC was not significantly increased (CIII p=0.38, CIV p=0.50). We investigated CI assembly in two patient-derived fibroblast cell lines and the *DNAJC30*-KO HEK cell line by BNE followed by complexome profiling. In all, CI was demonstrated to be correctly assembled with no assembly intermediates, ruling DNAJC30 out as a CI assembly factor (see **Fig. 3.30**).[22]

In the complexome profiling, DNAJC30 was detected at the size of the CI containing supercomplex where it was present in substochiometric quantities with CI (1:200 DNAJC30:CI) indicating a transient interaction with the CI containing supercomplex (**Tab. 14**). This was independently confirmed in the Bioplex Interactome database of protein-protein interactions [180, 181] (see Chapter 2.2.7), where DNAJC30 is demonstrated to interact with five CI subunits (see **Fig. 3.31**).

	IBAQ (control)
CI subunits (mean)	137659268
DNAJC30	687930
Ratio	200

Table 14: IBAQ values for CI and DNAJC30 in a control fibroblast cell line.
Adapted from [123].

These data thereby excluded the involvement of DNAJC30 in the structure or assembly of CI, but indicate an interaction with CI and a potential issue with degradation of CI subunits given the discrepancy in RNA and protein expression of DNAJC30.

3.3.7 Measurement of mitochondrial CI subunit exchange

CI N-module subunits have higher rates of turnover compared to other CI subunits in controls

We measured the turnover of single proteins within mitochondrial protein complexes in fibroblast and HEK cell lines by combining pSILAC with mass spectrometry of assembled RCC separated by BNE (see Chapter 2.2.2).[23]

In the control cell lines (seven fibroblast, one HEK), the rate of individual CI subunit turnover was found to be different across CI subunits. These differential rates of subunit turnover by 12 h subdivided the CI subunits into three categories (according to turnover rate in the control fibroblast cell lines):

[22] The BNE and complexome experiments were performed by Dr. Ilka Wittig and analysed by myself
[23] Experiment performed by Dr. Ilka Wittig and analysed by myself.

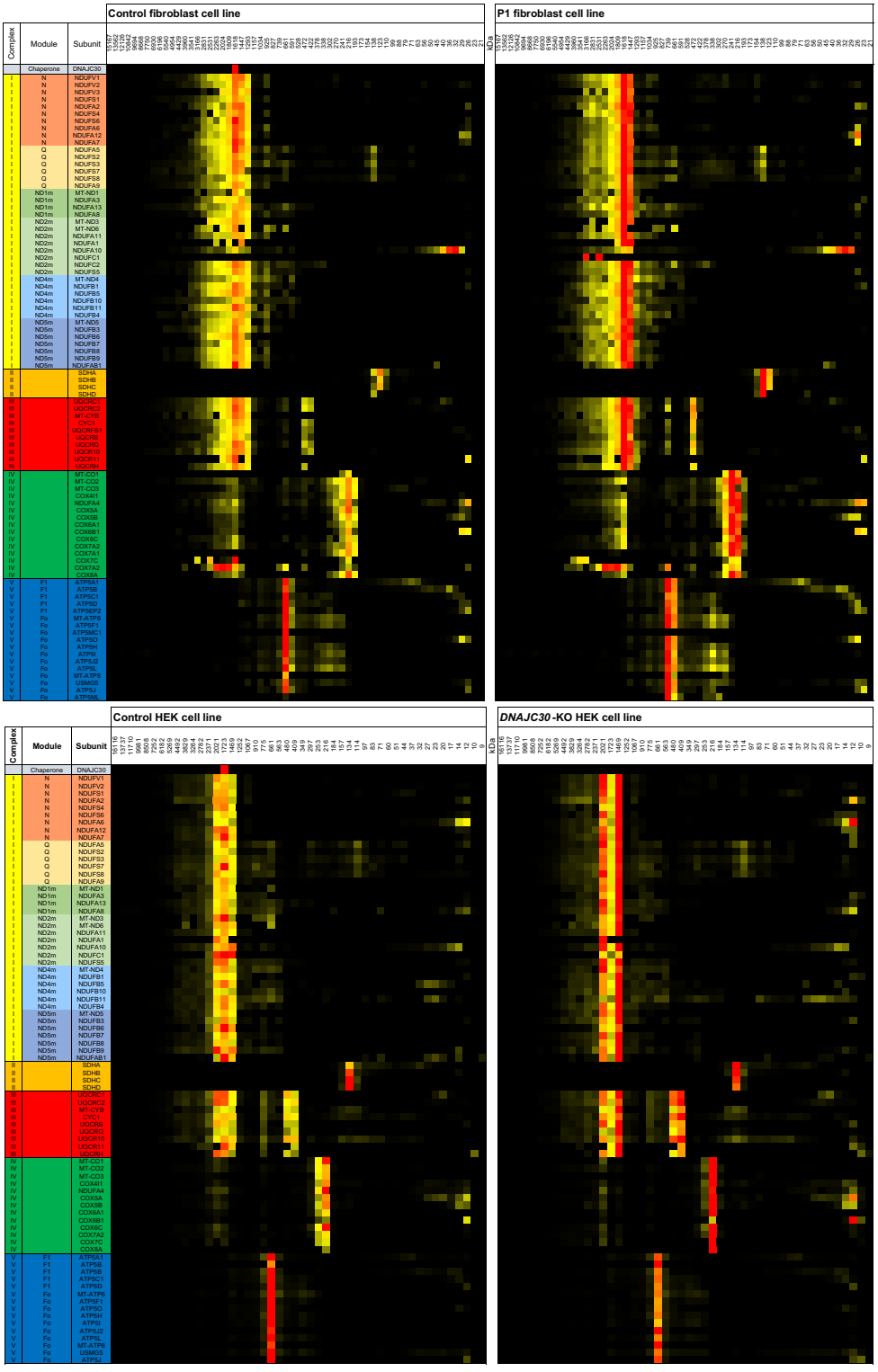


Figure 3.30: Complexome profiling of the mitochondrial OXPHOS complexes. Upper, control and patient-derived fibroblast cell line. Lower, control and *DNAJC30*-KO HEK cell line. DNAJC30 is depicted at the top of the profile. Color represents protein abundance, progressing from yellow to red with increasing protein abundance. *Adapted from [123].*

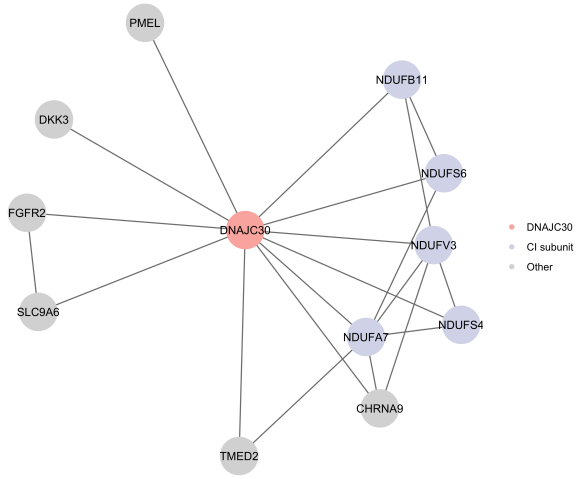


Figure 3.31: DNAJC30 interaction partners.
DNAJC30 and its interaction partners according to the Bioplex database [180, 181].

1. CI^{HIGH} accounting for five subunits of the N-module (mean turnover >25%)
2. CI^{MOD} accounting the remainder of the CI N-module (mean turnover 10-25%)
3. CI^{LOW} accounting for the remainder of CI subunits (mean turnover <10%)

The CI^{HIGH} subunits (NDUFV3, NDUFS4, NDUFS6, NDUFA6, and NDUFA7) demonstrated a mean turnover of 34% by 12 h (± 11% s.d.). These subunits have been identified as late participants in CI assembly according [182], indicating limited stability of their bond within the complex and ready exchangeability. Four out of the five subunits are direct interaction partners of DNAJC30 [180, 181] (see **Fig. 3.31**). The CI^{MOD} subunits (NDUFV1, NDUFV2, NDUFS1, NDUFA2, and NDUFA12) demonstrated a mean turnover of 18% by 12 h (± 6% s.d.). The CI^{LOW} subunits (Q-, ND1-, ND2-, ND4-, and ND5-module subunits) demonstrated a mean turnover of 9% by 12 h (± 6% s.d.).

These findings were recapitulated in the control HEK cell line, where CI^{HIGH}, CI^{MOD}, and CI^{LOW} subunits were demonstrated to have different turnover rates of 49%, 37%, and 16% by 12 h, respectively. Collectively these data indicate a role for DNAJC30 in CI subunit exchange.

A summary of the turnover of all detected OXPHOS subunits in the fibroblast and HEK cell lines at 0, 6, 8, 10, and 12 h is depicted in **Fig. 3.32**.

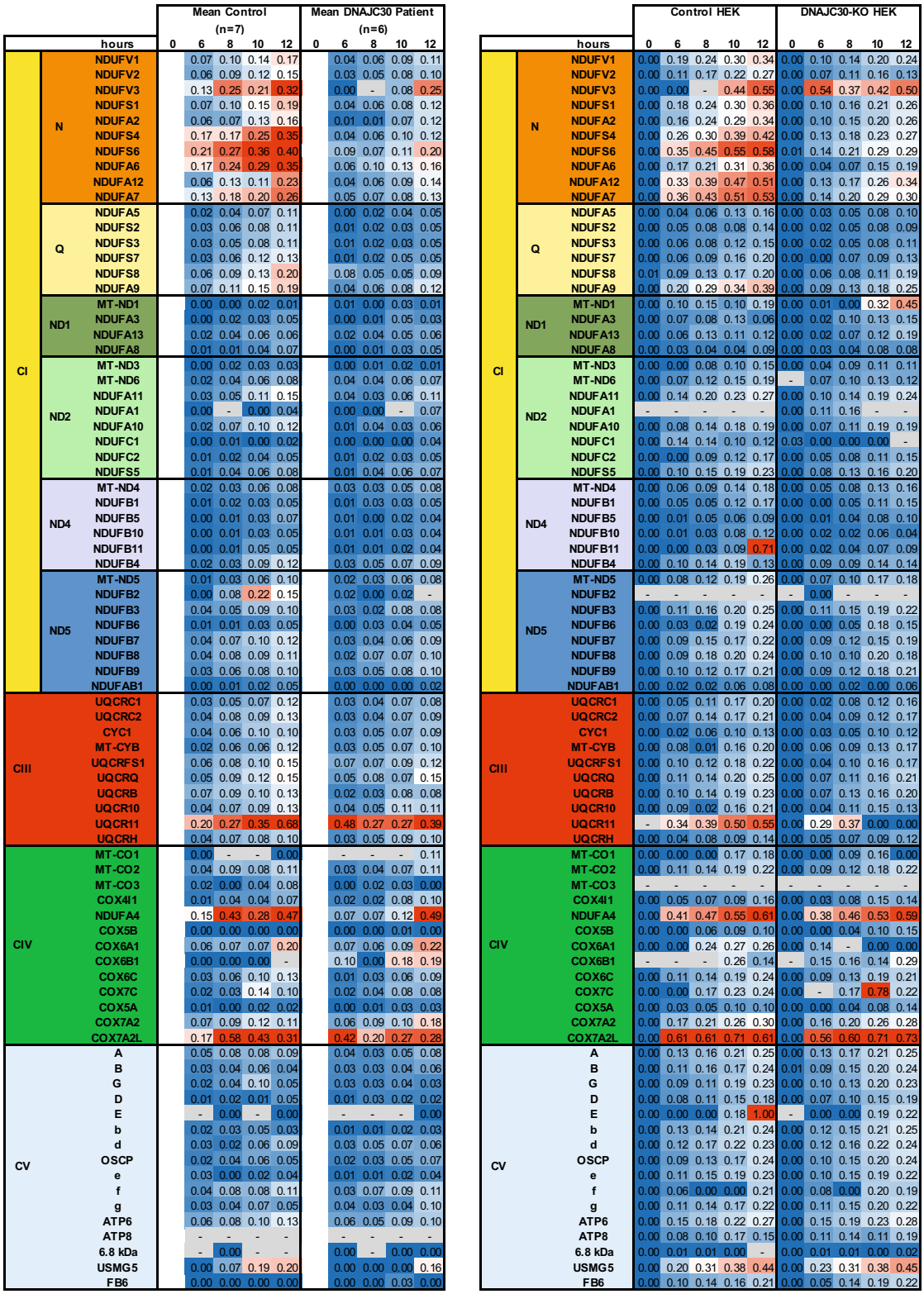


Figure 3.32: Subunit turnover in assembled OXPHOS complexes.
Individual subunit turnover of OXPHOS complexes in fibroblast and HEK cell lines. Color represents turnover rate progressing from blue (0) to white (0.15) to red (0.3) with increasing turnover rate. *Adapted from [123].*

CI N-module subunits have impaired turnover in *DNAJC30* defect

Measurement of individual OXPHOS subunit turnover in six patient-derived fibroblast cell lines, spanning all three *DNAJC30* variants, demonstrated a significant specific decrease in the turnover of the N-module subunits in assembled CI (see **Fig. 3.32**).

The mean turnover rates of the CI^{HIGH} and CI^{MOD} subunits were significantly lower than control (17% $p<0.0001$, and 13% $p<0.001$ Student's T-test, respectively). These findings were recapitulated in the *DNAJC30*-KO HEK cell line (31% $p=0.013$, and 25% $p=0.002$, in CI^{HIGH} and CI^{MOD} subunits, respectively) (see **Fig. 3.33**). The turnover defect was specific to arLHON as measurement in three mtLHON patient-derived fibroblast cell lines demonstrated normal CI subunit turnover. Moreover, the turnover defect did not demonstrate differences between gender to provide an explanation for the gender-dependent penetrance associated with the disease.

To validate the specificity of the defect for CI, we compared the turnover of 1,236 mitochondrial proteins, participating in 145 mitochondrial protein complexes, detected in the approach in the control and *DNAJC30*-KO HEK cell line. In this analysis, mitochondrial CI, and specifically the CI N-module were the only (sub)complexes with a significant differential turnover due to the knock-out of *DNAJC30* (mean delta turnover, 14% ± 8% s.d., adjusted $p=0.04$, Student's T-test).

Among the top six differentially expressed mitochondrial proteins of similar phylogenetic profile to DNAJC30, according to ProtPhylo (see Chapter 2.2.7) were CLPX and CLPB, components of the mitochondrial protein degradation machinery (forming part of the protein complex Clp protease, CLPXP), and the mitochondrial HSP70 (HSPA9) [183] (see **Tab. 15**).

Protein	Hamming distance (HD)
DNAJC4	323
DNAJC11	350
DNAJA3	356
HSPA9	363
CLPX	387
CLPB	389

Table 15: Mitochondrial proteins of similar phylogenetic profile to DNAJC30.
Similarity between the phylogenetic profile of DNAJC30 and other mitochondrial proteins quantified by the Hamming distance (HD). The top six results are presented. *Adapted from [123].*

3.3.8 Proposed role of DNAJC30 in mitochondrial CI repair

The following conclusions of the study provide lines of evidence to support the role of DNAJC30 in the repair of assembled CI (see **Fig. 3.34**):

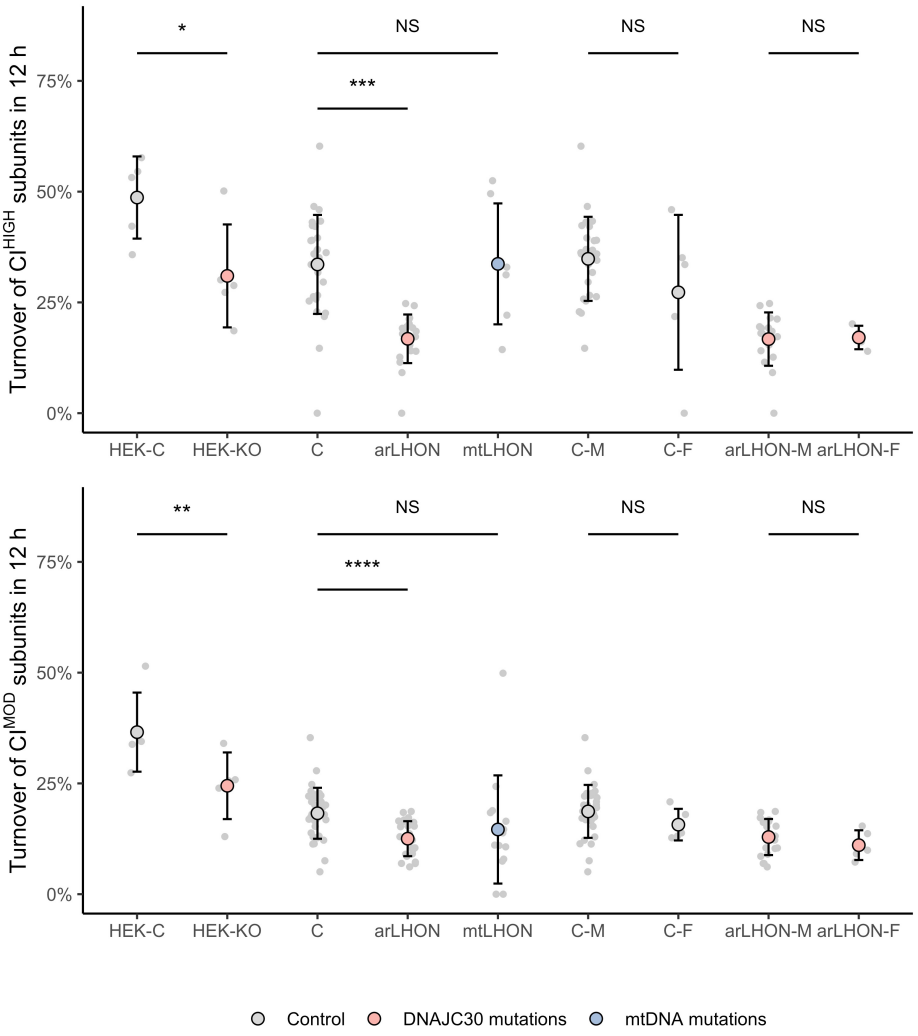


Figure 3.33: CI^{HIGH} and CI^{MOD} subunit turnover.
Turnover measurement of CI^{HIGH} and CI^{MOD} subunits by 12 h in seven control, six arLHON, and three mtLHON patient-derived fibroblast cell lines, and one control and one *DNAJC30*-KO HEK cell lines. Data are depicted as the mean ± s.d. C, control; M, male; F, female. *Adapted from [123].*

1. DNAJC30 is a mitochondrially localized chaperone protein specifically interacting with CI subunits.

2. These CI subunits are exposed to damaging reactive oxygen species and require repair, they are also late in the sequence of CI assembly, and thereby readily exchangeable resulting in high rates of turnover.

3. Defects in DNAJC30 impaired turnover of these CI subunits resulting in an accumulation of functionally defective CI.

4. DNAJC30 is likely to be functionally linked to the mitochondrial protein degradation machinery including CLPXP, defects of which also result in accumulation of functionally defective CI [10, 184].

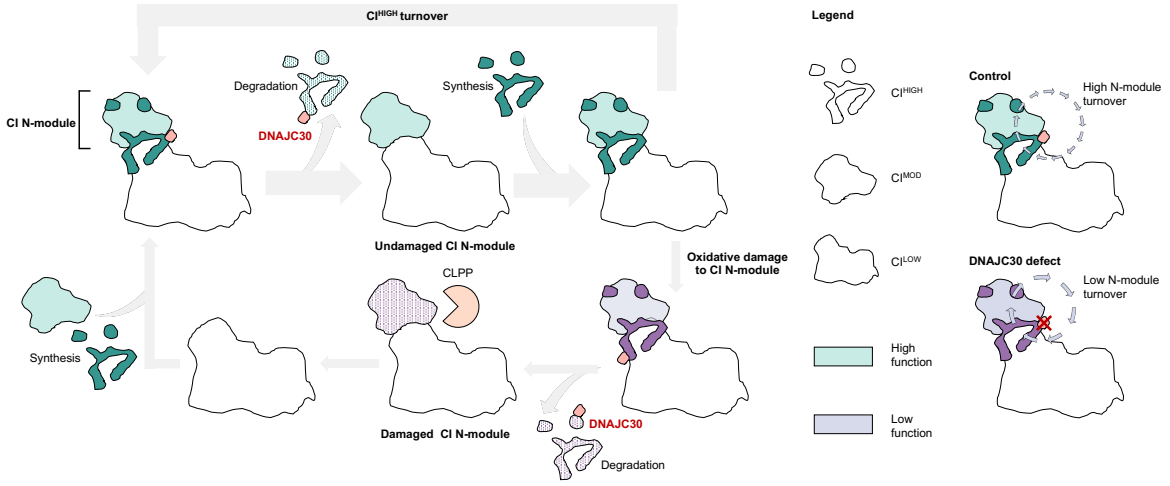


Figure 3.34: Schematic of the proposed role for DNAJC30 in CI repair.
Left, schematic of DNAJC30 functioning to facilitate the removal and replacement of specific CI subunits in a CI repair mechanism. Right, schematic of the effect of mutated *DNAJC30* on CI repair. *Adapted from [123].*

Under normal physiological conditions we thereby propose DNAJC30 to interact with the CI^{HIGH} subunits to facilitate their disassembly and subsequent degradation. In the setting of highly functional CI, these proteins may be newly synthesized and replaced without degradation of further CI subunits. While, in the case of oxidative damage to the CI N-module, leading to lowly functional CI, upon disassembly of the CI^{HIGH} subunits by DNAJC30 the protease CLPXP may access and remove the damaged CI^{MOD} subunits [10, 184]. Along with the CI^{HIGH} subunits, these subunits may then be subsequently resynthesised and replaced, negating the need for complete degradation and synthesis of CI at high energetic cost.

3.3.9 Summary of results

To summarise, in this study, a novel disease gene discovery resolved the enigma of LHON in the absence of pathogenic mtDNA variants, by the report of biallelic variants in a nuclear encoded gene, *DNAJC30*. The finding argued for "LHON" as a clinical description independent of transmission, and for the sub classification of this clear clinical syndrome into "mtLHON" and "arLHON" based on genetic cause, with important implications for genetic counselling and recurrence risk estimates. It was a single gene focused approach necessitating variants to be prioritised prior to functional study. The

availability of patient-derived muscle biopsies and fibroblast cell lines in addition to the generation of a *DNAJC30*-KO cellular model led to the discovery of a repair mechanism in mitochondrial CI. *DNAJC30* defects resulted in impaired mitochondrial CI function due to impairment in the turnover of specific subunits of the CI N-module, a module of CI exposed to higher risk of oxidative damage, thereby requiring maintenance.

3.4 Integrative analysis of multi-omic data solves inconclusive cases

This study regards the diagnosis of inconclusive WES cases by the integration of genetic, phenotypic, RNA-seq, and proteomic data. This section provides an overview of a study of 145 individuals and focus is placed on the diagnosis of two patients where the integrated omics analysis reprioritised candidate diagnoses previously discarded in the WES data analysis as they were determined not to be disease-causing due to failure to meet the standard criteria for prioritisation, namely to be rare (ACMG criterion BS1 and BA1, see Chapter 3.4.5) and to segregate with the disease in the family (ACMG criterion BS2 and BS4, see Chapter 3.4.6).

The corresponding manuscript was released as a preprint entitled "Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders" by Kopajtich R*, Smirnov D*, **Stenton S.L*** et al., *MedRxiv* 2021 (DOI: 10.1101/2021.03.09.21253187) [118] and at the time of writing this thesis was under review for publication in *Nature Communications*.[24]

3.4.1 Study population

Fibroblast cell lines from 145 individuals (121 unsolved, 22 solved positive controls, and two healthy controls) were utilised for RNA-seq and TMT-labelled quantitative proteomics. The 22 solved patients carried known protein destabilising pathogenic variants, previously confirmed by immunoblot analysis, and were included in the study to successfully validate the proteomic approach in detecting variant consequence on the protein level. The 121 unsolved patients had previously been investigated by WES/WGS under the clinical suspicion of mitochondrial disease and were deemed inconclusive. 21 of the 121 unsolved patients had VUS prioritised in the WES/WGS analysis requiring further functional evidence to determine pathogenicity. In the re-

[24] This project was led by Robert Kopajtich, Dimitrii Smirnov, myself, and Dr. Holger Prokisch. I personally contributed by patient HPO data curation and integration to advance the interpretation of candidate variants, ACMG variant classification, and data interpretation.

maining 100 patients, no variant(s) could be prioritised by the genetic analysis following standard analysis procedure for further exploration, as depicted in **Fig. 1.8**.

3.4.2 Disease gene coverage of RNA-seq and quantitative proteomics

Using established laboratory procedures (see Chapter 2.2.1 and Chapter 2.2.2) our RNA-seq and quantitative proteomics protocols detected a median of 10,425 transcripts and 7,686 proteins per sample. This provided coverage sufficient for the quantification of a median of 91% (353) and 80% (310) of mitochondrial disease gene products, and 59% (2,535) and 51% (2,159) of all Mendelian disease gene products per sample in RNA-seq and proteomics, respectively.

3.4.3 RNA and protein outlier calling

RNA and protein outlier calling utilising OUTRIDER and PROTRIDER for RNA-seq and quantitative proteomic data, respectively. A median of two aberrantly expressed transcripts and six aberrantly expressed proteins were identified per sample after multiple-testing correction. These expression outliers were stratified into three classes: (i) RNA-only outliers, (ii) protein-only outliers, (ii) and RNA-and-protein outliers.

The expression outliers were visualised by their respective z scores (the number of standard deviations below or above the population mean), as exemplified for two patients in **Fig. 3.35** and **Fig. 3.37** and were further annotated with:

1. The number of rare variants in the corresponding gene in the WES/WGS data, as either no rare variant, one rare variant, or potentially biallelic variants ($\geq$2 variants). In this analysis, rare was defined as an allele frequency $\leq$1% in keeping with standard practice for Mendelian disease [185], as opposed to the more stringent threshold of $\leq$0.1% routinely applied in suspected mitochondrial disease to increase variant capture.

2. The HPO semantic similarity scoring to the reported disease-associated phenotype for the encoding gene (where applicable, not calculated for candidate novel disease genes).

Given that we expected impaired function to result in disease, we focused on the under-expression outliers of all outlier class (median four per sample), which accounted in total for two-thirds of all outliers. We further prioritised these outliers by those with rare variants in-keeping with the mode of inheritance of variants in the gene and with a HPO phenotype semantic similarity score of $\geq$2 with the respective encoding disease gene. This prioritisation strategy resulted in a median of one RNA and/or protein outlier of interest per sample, a manageable number for manual curation.

3.4.4 Diagnostic rate

Overall, our integrative workflow led to the genetic diagnosis of 26 of the 121 unsolved cases (22%). This included 14 patients with WES/WGS prioritised VUS (confirmed by nominally significant protein under-expression) and a further 12 cases solved by the integration of genomics, phenomics, RNA-seq, and proteomics where the disease-causing variants had not previously been prioritised. Of these 12 cases, all were validated by aberrantly low protein expression and eight with simultaneous aberrantly low RNA expression. Moreover, seven cases received an additional layer of functional evidence by the under-expression of their corresponding protein complex, as exemplified in Chapter 3.4.6. All patient analyses were made freely browsable on our web interface "omicsDiagnostics", developed by Dimitrii Smirnov (see Chapter 2.2.7).

3.4.5 *DARS2* defect due to a rare combination of frequent variants

The first patient, a male, presented in childhood with muscular hypotonia, cardiomyopathy, neurodevelopmental abnormality, abnormalities in the cerebral white matter on MRI, elevated serum lactate, and a combined mitochondrial CI and IV defect on muscle biopsy.

Upon inspection of the integrative multi-omic analysis, *DARS2*, a mitochondrial tRNA synthetase and known mitochondrial disease gene, was detected as an RNA-and-protein under-expression outlier with a semantic similarity score of 3.8 with the reported disease-associated phenotype (see **Fig. 3.35**). This highlighted *DARS2* as a promising diagnosis for further consideration. However, defects in *DARS2* are inherited in an autosomal recessive manner, and only one single rare variant could be identified in the WES data. The identified variant was a splice variant, c.492+2T>C (NM_018122.5), reported in ClinVar as pathogenic and with an MAF of 0.038% in gnomAD (no homozygous carriers). Due to lack of a potentially biallelic second rare variant it was, however, discarded during the prioritisation of variants in the WES analysis.

Given the significant under-expression of DARS2 on the RNA and protein level, we manually inspected the WES data in the IGV (see Chapter 2.2.7) in search of a more atypical genetic cause for the disease. In the genetic data, we identified two frequent near-splice variants in *cis* on the second allele of *DARS2*, c.228-12C>G and c.228-20T>C, the latter was homozygous in this case. The c.228-12C>G variant was predicted benign by the Ensembl Variant Effect Predictor (VEP) (see Chapter 2.2.7), and is reported in 7,849/281,358 alleles (984 homozygous carriers) in the gnomAD database, resulting in an allele frequency of 2.8%. The second c.228-20T>C variant, also predicted benign by the VEP, is reported in 57,295/214,832 alleles (2,277 homo-

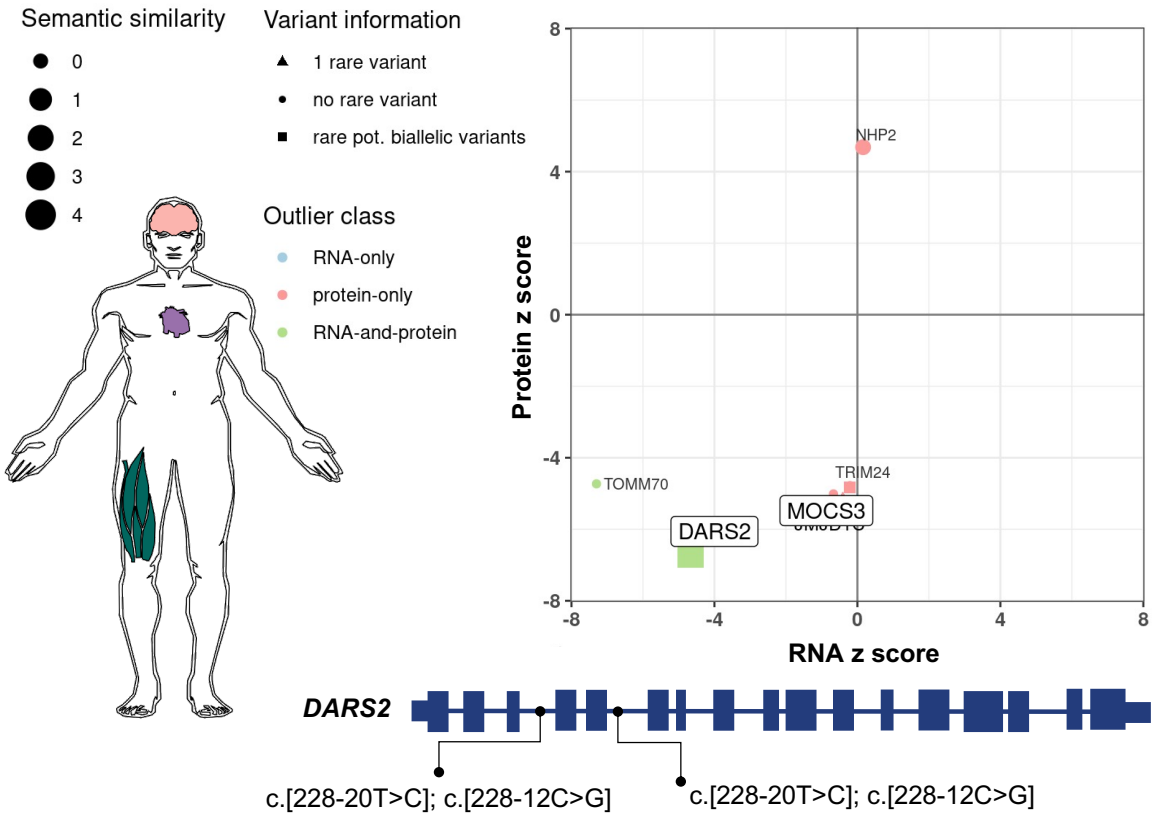


Figure 3.35: Multi-omics integration in a patient with *DARS2* defect.
Integration of genotype, phenotype, RNA-seq, and proteomic data for the diagnosis of a patient with *DARS2* defect. The patient phenotype is depicted by system involvement. *Adapted from [118].*

zygous carriers) resulting in an allele frequency of 20.7%. Individually, the allele frequencies of these variants are much too high to be considered disease-causing (ACMG criterion BS1 and BA1). However, this patient was discovered to be the the only WES case amongst our in-house database of >20,000 WES data sets to carry this mono-allelic combination of frequent variants, deeming the combination of c.228-12C>G and c.228-20T>C rare (allele frequency <0.0025%).

Inspection of the Sashimi plot of the RNA-seq data in the IGV provided further evidence for pathogenicity. The variant combination led to an aberrant splice event with skipping of exon 3 in 10% of reads (see **Fig. 3.36**). By comparison, RNA-seq data from patients homozygous for each of these frequent variants demonstrated exon skipping in just 1% and 2% of reads, respectively. The consequence of the pathogenic c.492+2T>C variant was demonstrated by skipping of exon 5 in just under 70% of reads. Collectively, these aberrant splice events resulted in significant RNA under-expression (z score -4.66), likely due to NMD of the abnormally spliced transcript, and

leading to a subsequent protein under-expression outlier (z score -6.7) leaving 71% of the transcript and 37% of the protein (see **Fig. 3.35**).

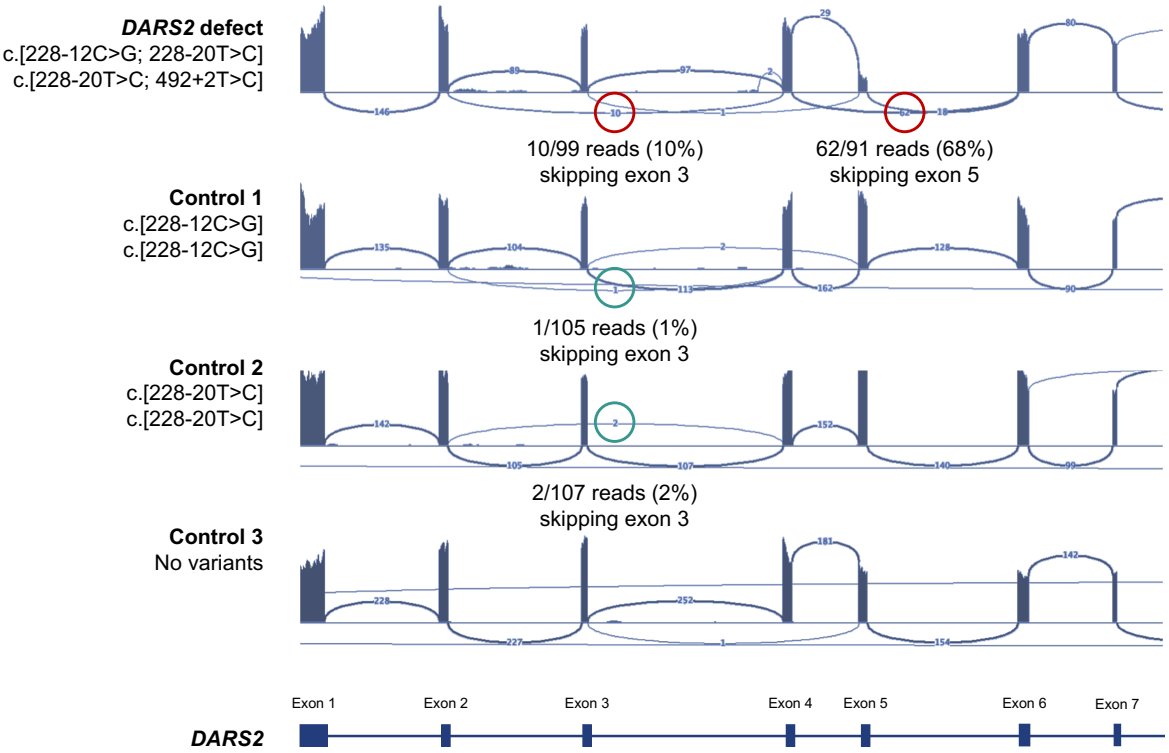


Figure 3.36: Splice defects in *DARS2* depicted by Sashimi plot.
Sashimi plot to depict the number of reads per sample following each splice pattern for the *DARS2* defect patient in comparison to two controls homozygous for each of the two frequent variants and a control with no splice disrupting variation in *DARS2*.

3.4.6 *NDUFB11* defect due to an incompletely penetrant variant

This male patient presented small for gestational age in the neonatal period with feeding difficulties, muscular hypotonia, hypertrophic cardiomyopathy, global developmental delay, anaemia, recurrent hypoglycaemia, elevated serum lactate, elevated serum alanine, and elevated lactate on brain magnetic resonance spectroscopy (MRS).

Upon inspection of the integrative multi-omic analysis, *NDUFB10* and *NDUFB11*, mitochondrial CI subunits and known mitochondrial disease genes, were detected as protein-only expression outliers (see **Fig. 3.37**). Defects in *NDUFB10* are inherited in an autosomal recessive manner, however, in this patient only two predicted benign intronic variants in *NDUFB10* could be found. In contrast, defects in *NDUFB11* are inherited in an X-linked recessive manner, and in keeping with this, the patient carried a novel hemizygous missense variant in *NDUFB11*, c.440T>C, p.Met147Thr

(NM_019056). The variant is absent in gnomAD and was annotated with *in silico* computational predictions suggestive of pathogenicity (SIFT 0.02, CADD 22.2).

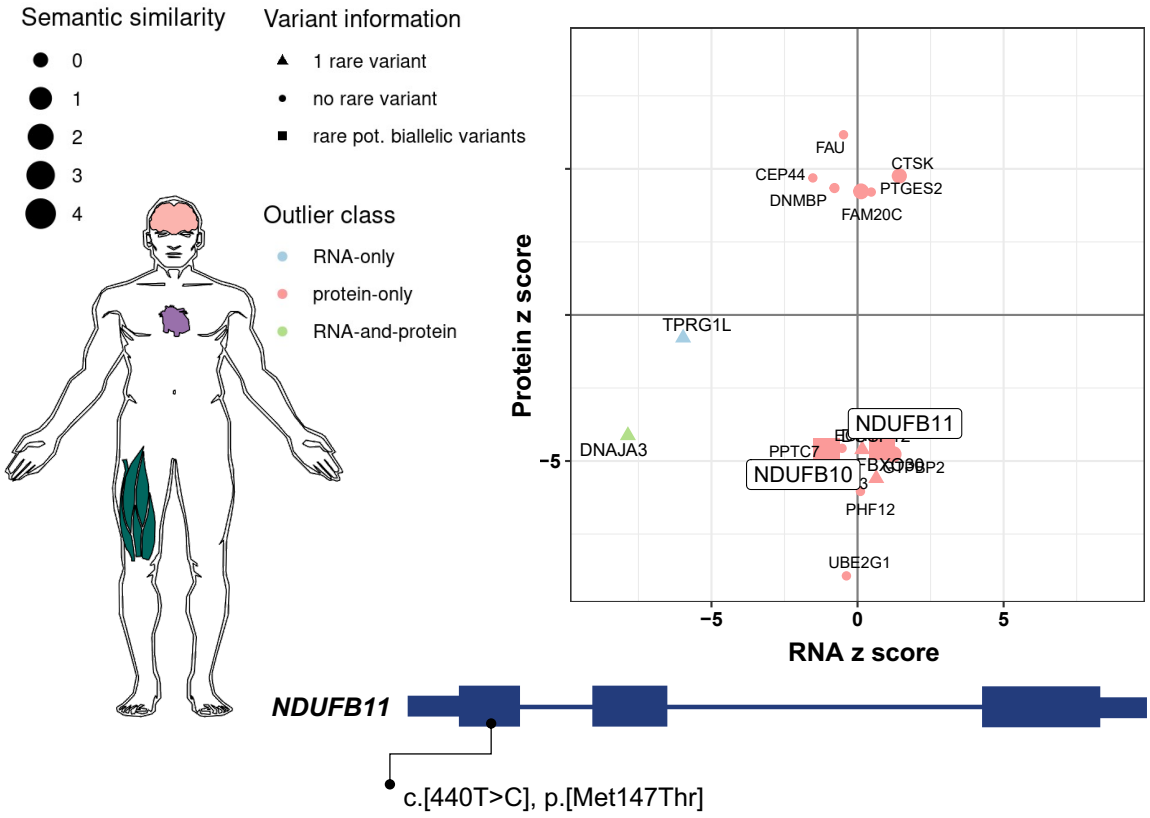


Figure 3.37: Multi-omics integration in a patient with *NDUFB11* defect.
Integration of genotype, phenotype, RNA-seq, and proteomic data for the diagnosis of a patient with *NDUFB11* defect. The patient phenotype is depicted by system involvement.

The c.440T>C variant was discarded in the WES analyse due to the variant being confirmed to also be carried by the unaffected maternal grandfather (ACMG criterion BS2 and BS4). It was determined to be benign for this reason. However, by exploration of the case in a integrated multi-omic approach, multiple layers of evidence support the pathogenicity of this novel variant:

1. The patient's phenotype has high semantic similarity (score 4.0) with the established disease-associated phenotype, due to a match of intrauterine growth retardation, hypertrophic cardiomyopathy, and elevated serum lactate .

2. The variant resulted in significant aberrant protein underexpression (z score -4.1, 46% of protein remaining).

3. Reduction in the NDUFB11 protein led to the collective pathologically low abundance of mitochondrial CI subunits (63% remaining), due to destabilisation of the complex. This reduction was comparable with other patients with defects in CI subunits (second patient with *NDUFB11* defect, 83% of CI subunits remaining) and defects in CI assembly factors (two patients with *TIMMDC1* defect, 70% and 76% of CI subunits remaining, one patient with *FOXRED1* defect, 73% of CI subunits remaining).

4. No rare variants were identified within any other CI subunit encoding gene, with the exception of two predicted benign intronic variants in *NDUFB10*, to provide an explanation for these findings.

The reduction in CI subunits was most pronounced in the CI ND4-module (44% remaining, lowest in dataset), of which NDUFB11 is a subunit. This was in-keeping with a second confirmed NDUFB11 defect patient where 55% of the ND4-module remained (second lowest in dataset) (see **Fig. 3.38**). These findings effectively exemplify the power of proteomics, not only in detecting variant consequence on the encoded protein, but on the entire corresponding protein complex. In this manner, protein outliers without rare variants in the encoding gene (here the CI subunits, including NDUFB10) may be explained indirectly as a consequence of protein complex instability due to a defect in one of the interaction partners, as has previously reported in single cases in the literature [117, 140, 141, 142].

The multiple layers of evidence for pathogenicity from the proteomic and HPO data deemed the variant likely pathogenic (LP, class 4) according to the ACMG criteria, and in combination with detection of the variant in the unaffected maternal grandfather argued for incomplete penetrance. Fibroblasts were not available from further family members for analysis. This challenging diagnosis, made possible by genome-wide quantitative proteomics, was comparable to the result discussed in Chapter 3.3, where quantitative proteomics was able to provide evidence for the loss-of-function character of an incompletely penetrant missense variant in the novel mitochondrial disease gene *DNAJC30*.

3.4.7 Summary of results

To summarise, in this study, multiple levels of omics data were integrated to validate prioritised VUS from WES/WGS in 14/121 unsolved cases and to simultaneously discover and validate the genetic cause of disease in a further 12 cases, resulting in an overall diagnostic success rate of 21% (26/121 WES/WGS unsolved cases). The comprehensive approach facilitated the diagnosis of patients by the validation of VUS previously discarded in the standard approach to WES/WGS analysis due to features

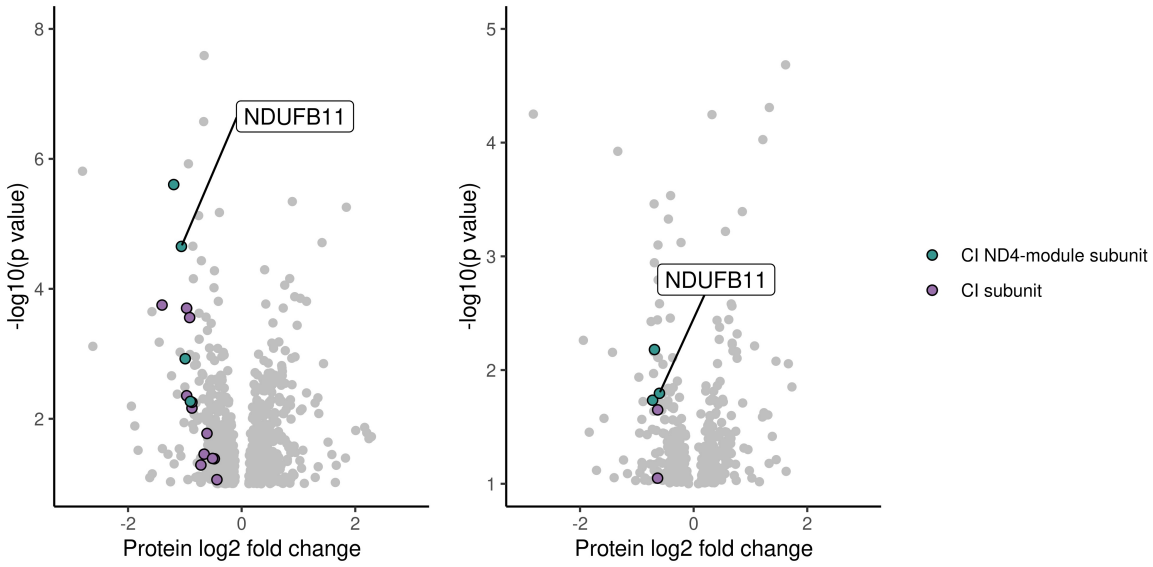


Figure 3.38: Protein expression depicted by volcano plot.
Left, all detected expressed proteins in the discussed patient (4/6 ND4-module subunits and 24/45 CI subunits detected). Right, all detected expressed proteins in the second patient with *NDUFB11* defect (4/6 ND4-module subunits and 31/45 CI subunits detected).

indicating them to be benign, such as higher than expected allele frequency of the splice variants in *DARS2*, and failure of the missense variant in *NDUFB11* to segregate with the disease in the family, thereby allowing more challenging diagnoses to be made.

4 | Discussion

Human genetics endeavours to understand the molecular genetic cause of a patient's disease. Advancements in sequencing technology and bioinformatic infrastructure in the field over the last decade have greatly increased capacity to do so, with movement from clinical descriptions of disease and observation of inheritance patterns, to the reliable identification of disease-causing variation, at single base resolution, within a search space of over 20,000 genes. Amongst rare genetic diseases, mitochondrial diseases are a prime example of the diagnostic challenges faced by human geneticists, due to their vast genetic underpinning and broad spectrum of clinical manifestations. Despite these hurdles, the motivation to provide a genetic diagnosis prevails, given its value in genetic counselling and accurate recurrence risk estimation. Moreover, knowledge of the genetic defect can direct implementation of preventative measures and anticipatory care, provide a clearer outlook on disease course and prognosis, and allow the selection of individually targeted treatments to ameliorate disease. In these regards, WES has been instrumental, accelerating Mendelian disease gene discovery to over 300 novel disease genes per year [85], and resulting in 4,032 disease genes in total (as reported by OMIM), of which mitochondrial diseases account for a substantial proportion (∼10%).

In the course of this Ph.D, the diagnostic successes and challenges of WES in reaching a genetic diagnosis for clinically suspected mitochondrial disease were comprehensively analysed in over 2,000 paediatric patients in study 1 of this thesis and discussed in Chapter 4.1. Beyond identifying known pathogenic variants in known disease genes, WES was successfully utilised to identify novel disease-causing variants in known mitochondrial disease genes, and novel disease-causing variants in novel disease genes not previously associated with disease, as exemplified in study 2 and 3 of this thesis and discussed in Chapter 4.2 and Chapter 4.3, respectively. In the setting of inconclusive WES, particularly where no potentially pathogenic variants could be prioritised for further exploration, study 4 leveraged systematic integration of WES, HPO phenotype descriptors, RNA-sequencing, and quantitative proteomics to identify candidates and simultaneously provide functional evidence for pathogenicity to establish a diagnosis, discussed in Chapter 4.4.

4.1 Diagnostic utility of WES in suspected mitochondrial disease

Over 2,000 (2,035) paediatric patients were analysed by WES under the clinical suspicion of mitochondrial disease, constituting the largest study of its kind to date, previously only analysed in cohorts of up to 142 patients [81, 87, 88, 89, 90, 91, 92, 93, 94]. Such a large collection of mitochondrial disease patients was only made possible by inclusion of patients in mitochondrial disease networks, and by collaboration across international diagnostic centres. In this section of the discussion, mitochondrial disease will be abbreviated to MD.

A genetic diagnosis was made in just over 50% (1,093/2,035) and 63 novel disease gene discoveries were made in the cohort over the 10 year period of the study (as demonstrated by the discovery of *DNAJC30* in this thesis, discussed in Chapter 4.3). The high diagnostic rate was achieved by close interconnection between research and routine diagnostic approaches, and places suspected MD amongst the highest yielding metabolic and neurological primary indications for WES (on average 20-35%) [86] (see **Fig. 4.1**). Without functional follow-up studies, the diagnostic rate would have been in the same range as published for other diseases (35%) [86]. The estimation of the true diagnostic rate is further complicated by different diagnostic workflows in the contributing centres, varying from case to case and ranging from WES as the first tier approach in Munich and Beijing, to single gene studies followed by targeted a gene panel in other centres (e.g., Newcastle, Paris, and Milan). However, for a number of reasons, the reported diagnostic rate is still conservative. First, by nature the study focused on more difficult cases referred to specialist centres, and does not intend to provide an epidemiological perspective on the diagnostic rate. Across these centres, we diagnose at least 15% of patients without WES, mostly by mtDNA sequencing (see **Fig. 3.2**), a

calculation based on more than 3,000 cases among the contributing centres. Thereby, had WES been applied as a first step in the diagnostic pathway, the yield would have been considerably higher. This is reflected by the most recent 250 patients analysed at the Institute of Human Genetics, Munich, Germany, where WES is the first genetic test in the majority and the diagnostic rate was >60%. Second, in a further 8% of patients (160/2,035) VUS in established MD genes or candidate MD genes (with mitochondrial localisation of the encoded protein) were identified with marked phenotypic similarity to solved MD cases (see **Fig. 3.12**). These VUS have high likelihood for pathogenic designation upon future functional studies. Despite these limitations, overall, given the large number of patients and involved centres, the study provides a reliable estimate of the power of WES in MD.

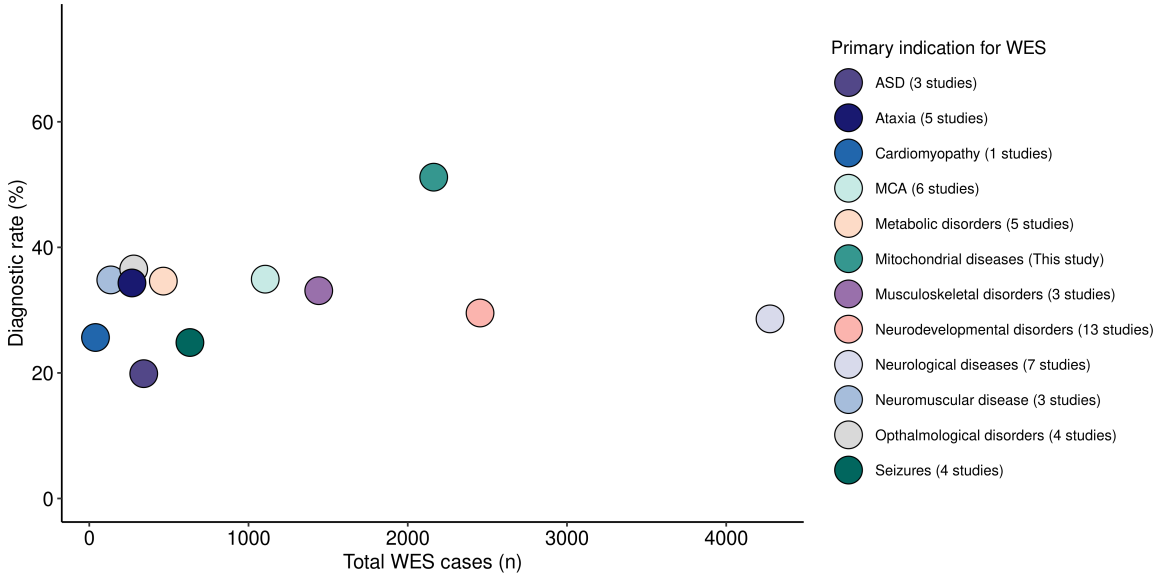


Figure 4.1: Diagnostic rate of WES by primary indication.
Collective diagnostic rate of WES by primary metabolic or neurological indication according to the current literature as compared to the study presented here of >2,000 suspected mitochondrial disease patients. ASD, autism spectrum disorder; MCA, multiple congenital abnormality.

Functional studies had high value in variant interpretation, necessitated by the preponderance of missense variants with no prior functional annotation unique to a single patient (>80%). Without functional validation, the diagnostic rate would have capped at approximately 35% (see **Fig. 3.2**), strongly arguing for proactive bio-material sampling in patients, (e.g., a minimally invasive skin biopsy to establish a fibroblast cell line) and parallel analysis of WES and functional data. Fibroblasts are an excellent resource for the study of MD. They can be used for mitochondrial RCC measurement with results comparable to invasive muscle biopsy [60, 57], for detection of impaired mitochondrial respiration by measurement of oxygen consumption rate, and for rescue

complementation (as demonstrated in the validation of *DNAJC30*, see Chapter 4.3). Moreover, fibroblasts open the option for multi-omic studies, pioneered in MD, such as RNA-seq and proteomics [117, 118, 132] (as discussed Chapter 4.4). These assays pave the way for integrated high-throughput functional readouts, allowing interpretation of the vast number of VUS detected by WES sequencing.

WES reanalysis is another valuable approach to increase diagnostic rate, and has been demonstrated in the literature to provide a genetic diagnosis to up to 10-30% of unsolved cases just 1-3 years after initial analysis [110, 111, 112, 113, 114, 115, 116]. In this study, on average a 2.6 year interval (± 1.7 year s.d.) was needed to enable a genetic diagnosis by reanalysis of the WES data, due to the continual discovery of novel disease genes (>300 per year in OMIM and 212 MD genes during the course of this study) (see **Fig. 1.4**), expansion of genotype-phenotype associations, and the description of different modes of inheritance for established disease genes. Of the cases solved by reanalysis, 55 genetic diagnoses were made in MD genes and 36 in association with MD-phenocopy genes. A prime example was the diagnosis of eight patients with pathogenic *de novo* dominant variants in *MORC2*, by reanalysis following a report expanding the phenotypic spectrum to include neurodevelopmental and metabolic abnormalities, including a Leigh syndrome-like phenotype [186].

Reaching a genetic diagnosis is important for family counselling, prevention, treatment, and clinical trial development. Disease modifying treatments, such as a ketogenic diet in *PDHA1* defects, are reported for 28 of the identified MD gene defects, across 174 patients in the study [64, 65]. In addition, 44 identified MD-phenocopy disease genes across 62 patients, have reported treatments (including dietary modifications in *GCDH, FBP1*, and *PHKA2* defects, specific enzyme replacement in *GLA* and *TPP1* defects, and copper histidinate injections for Menkes disease due to *ATP7A* defects) [64]. In total, these 236 patients where treatments may be of benefit (236/1,093 solved cases, 22%) endorse a genome-wide approach over selecting a targeted panel in a heterogeneous cohort to identify and optimise care for unexpected diseases without delay. This figure of more than 20% with defect driven modifying treatment is widely underestimated for suspected MD and needed a large study, such as that presented here, to be demonstrated.

Despite the challenges to identify MD clinically, over 70% of the solved cases were genetically confirmed as such, with defects equally distributed across all classes of mitochondrial function. This figure is comparable to previous larger studies (of >100 cases) of WES in suspected MD cohorts, where between 50-85% of solved cases were due to defects in MD genes [89, 90, 91, 92]. Though a preponderance of nuclear DNA variants (>85%) and autosomal recessive inheritance (>75%) in the solved MD

patients was anticipated for a paediatric population [39], the figures are inflated by ascertainment bias given mtDNA screening in some centres. This is also exemplified by the small number of mtDNA mutations responsible for LHON and MELAS (11 patients) in the study. *PDHA1*, *ECHS1*, and *ACAD9* were the most frequent nuclear-encoded MD gene hits, in-keeping with large collections of patients in the literature [187, 188, 189] and the report of founder mutations in *ECHS1* [190, 191]. Nevertheless, certain mtDNA-encoded genes with wide clinical variability, such as *MT-ATP6*, were frequents hit in line with mutation reports in collections of >200 patients including those with non-typical presentations [34, 35, 36]. This reinforces the value of calling mtDNA variants from WES data as per [31] in a non-primary affected tissue in children. Importantly, the low frequency of certain individual nuclear-encoded genetic diagnoses in the study, may be distorted by prior targeted sequencing of candidate disease genes, such as *POLG* (seven patients) and *SURF1* (10 patients), two of the more frequently reported nuclear-encoded MDs in the literature [192, 39, 42]. In contrast to MD, 6-fold more MD-phenocopy gene defects in patients with autosomal dominant and *de novo* dominant inheritance of disease were reported, highlighting the importance of trio-based WES for efficient interpretation, especially in a clinically heterogeneous cohort with significant overlap with neurodevelopmental delay, where these modes of inheritance are commonplace [193].

A dual diagnosis was identified in less than 0.5% (7/2,035), a figure lower than the expected 1-5% reported for patients with different rare Mendelian diseases in the literature [194, 195]. Dual diagnoses can lead to a blended phenotype, the risk of overlooking a second diagnosis upon identification of the first, and consequently, the risk of incorrectly expanding a genotype-phenotype association. An example in the study is of a patient with a dual diagnosis in *ACADS* and *HIVEP2*. The patient's clinical presentation included phenotypes both mutual (global developmental delay and muscle hypotonia) and exclusive (dicarboxylic aciduria to *ACADS*, and cerebral atrophy to *HIVEP2*) to the two diseases. We can, however, exclude a overlooked second diagnosis based on reported ClinVar pathogenic variants in further cases, given a systematic screen of all cases for such variants.

From the clinical perspective, the digitilisation of the phenotype data as HPO terms facilitated ease in comparison between patients with different levels of clinical detail, and focused on the comparison of the genetically diagnosed MD and MD-phenocopy patients. These analyses revealed encephalopathy, cardiomyopathy, MRI signal intensity changes in the basal ganglia and/or brain stem, elevated serum lactate, and mitochondrial RCC defects to be the most enriched HPO terms in solved MD patients (see **Fig. 3.10**). Using the digitalised phenotype information, it could be shown that based on the MDC score all patients were possible (598, 29%, score 2-4), probable (802,

39%, score 5-7), or definite (635, 31%, score 8-12) to have a MD. However, overall, given extensive phenotypic overlap with other Mendelian diseases, MDC scores were unable to reliably stratify the MD and MD-phenocopy patients within the selected cohort, and HPO phenotype was only able to prioritise the correct genetic diagnosis less than 5% of solved patients, a figure considerably lower than reported for general paediatric genetic studies [196]. This phenotypic heterogeneity not only arises within the spectrum of mitochondrial disease genes, but also between variants within a single gene defect [188], indicating the need to study MDs on the variant level. A starting point is GENOMITexplorer, providing patient-level resolution of genotype-phenotype associations so far not available in established databases such as OMIM and Orphanet. In terms of biomarkers, elevated serum lactate was found to have a low sensitivity and specificity within the study, of 65% and 70%, respectively (AUC 0.68) (see **Tab. 6**). Measurement and interpretation of serum lactate is challenged by the risk of missing a transient elevation in MD, and reporting a non-specific elevation in an acutely unwell child, a struggling child, or due to tourniquet use. Similarly, the low specificity of RCC defects for MD likely arose as a consequence of secondary mitochondrial dysfunction associated with pathology [197]. However, these measures do not reflect the sensitivity and specificity expected in clinical practice, as in this study MD patients were not compared to a general pediatric disease population. Within more general cohorts, serum lactate is reported to have a slightly higher AUC of 0.71 [56]. For these reasons, MD is most efficiently and reliably diagnosed by WES early in the diagnostic pathway, in order to direct the need for specific clinical investigations and to prevent unnecessary interventions in children.

A genetic diagnosis was unable to be provided to just under half of the investigated patients. The possible reasons for this are manifold. First, due to the 15% of cases with WES prioritised VUS which may be designated with pathogenic status upon future functional validation studies. Second, due to pitfalls in variant prioritisation, such as *de novo* variants given the minority of trio-based WES in the study, and variant interpretation (e.g., hypomorphic and synonymous variants causing abnormal splicing, and deep intronic and regulatory variants), a limitation beginning to be addressed by multi-omic integration [118] (discussed in Chapter 4.4). Third, due to certain genetic variants being intractable to WES given insufficient depth of coverage, locus-specific features, (e.g., GC-rich regions and homopolymeric repeats), sequencing biases, and genomic alterations (e.g., large deletions, insertions, chromosomal rearrangements, short tandem repeats), leading to pathogenic variants in known disease genes evading detection. The inability to delineate some of these alterations may be overcome by WGS, specifically long-read sequencing. This is supported by the multi-omics study presented in this thesis (discussed in Chapter 4.4), in which non-coding variants were found to be disease-causative in a fraction of at least 10% of the unsolved WES cases.

Finally, due to individuals with suspected MD carrying pathogenic variants in as yet unidentified disease-relevant genes, the discovery of which continues year-on-year (see **Fig. 1.4**). These disease gene discoveries, as exemplified by the 63 novel disease genes discovered in this cohort, are accelerated by collation of patient datasets to increase the signal for detection and to facilitate challenging diagnoses to be made, such as variants with incomplete penetrance as described in *DNAJC30*, a novel recessive MD manifesting as LHON and responding to idebenone therapy (see Chapter 4.3). This underpinned the motivation to form an international MD network and encourages patient referral to centres of expertise in which research projects flank state of the art routine diagnostics.

Overall, the study demonstrated the power and efficiency of WES in diagnosing patients with suspected MD, illuminating the substantial fraction benefiting from early diagnosis and targeted treatment, and exemplifying the high impact selecting WES early in the diagnostic pathway may have on health economics.

4.2 Novel variant discovery and validation in a known disease gene

With the rare exception of founder pathogenic variants (as described in *DNAJC30*, see Chapter 4.3) and recurrent *de novo* variants, such as arising in *PDHA1* [193], variants detected in known mitochondrial disease genes are often seen for the first time in a single patient only, and are therefore absent in curated disease variant databases. This was exemplified for >80% of variants in the WES study (discussed in Chapter 4.1). When these variants are not confidently predicted to result in protein truncation (loss-of-function), further functional validation is needed to assign pathogenicity.

This thesis describes a study validating the pathogenicity of novel missense and in-frame indel variants in the known mitochondrial disease gene *FDXR*, responsible for iron-sulfur (Fe-S) cluster biosynthesis. FDXR deficiency is an AR mitochondrial disease with an estimated lifetime risk of 0.11 per 100,000 (95% CI, 0.08-0.14). It is thereby amongst the 50 mitochondrial diseases of highest prevalence, as estimated by rare variation in population databases [42]. In reflection of this, the study collated the genetic and phenotypic data of over 30 patients with *FDXR* defects, of which seven were reported for the first time.

FDXR is an essential gene, defined as being required for the survival of an organism or a cell. In keeping with this, to date, no patients are reported with biallelic protein-truncating variants (predicted to lead to loss-of-function), a genotype expected to be incompatible with life. Instead, the majority of defects are caused by missense variation, with variable effect on the encoded protein. This results in phenotypes

spanning relatively benign single system disorders to severe multi-systemic infantile-onset disease with early lethal course. In the seven newly reported cases, the frequent *FDXR*-associated phenotypes (e.g., optic atrophy, acoustic neuropathy, and developmental delays with regression) in addition to a number of novel and more severe phenotype associations were observed. These novel phenotypes included ptosis and type I diabetes mellitus, reported in single cases, alike to other disorders of Fe-S protein metabolism, such as Friedreich ataxia, with which FDXR deficiency has considerable phenotypic overlap (e.g., neuropathy, ataxia, and nystagmus). The severe phenotypes in the study included the first patient to be reported with *FDXR* defects meeting the diagnostic criteria for Leigh syndrome [47], manifesting with a triad of developmental regression, bilateral basal ganglia MRI signal intensity changes, and elevated CSF lactate, in combination with the commonly-reported *FDXR* phenotype, optic atrophy. Further supporting evidence for the association of *FDXR* defects with delayed myelination, dystonia, ophthalmoplegia, and cataracts, phenotypes previously only reported in single cases [124, 166, 167], was provided. The report of additional patients with these phenotypes increased the strength of the evidence linking them to their shared genetic diagnosis, indicating that they are indeed due to *FDXR* defects and are less likely to be sporadic occurrence or due to a second (dual) genetic diagnoses resulting in a blended phenotype. Disease-associated phenotype expansions, such as described here for *FDXR*, are becoming increasingly commonplace in human genetics with the uptake of unbiased WES/WGS, given the ability to identify patients presenting a-typically for their genetic diagnosis (e.g., most recently for *MT-ATP6* [34, 35, 36]). Moreover, these findings increasingly question the value of prior genotype-phenotype associations in predicting the genetic cause of disease, given that the phenotypes associated with individual mitochondrial disease genes is in continual flux. This was effectively exemplified in the WES study presented in this thesis (see Chapter 4.1), by HPO phenotype being unable to retrospectively prioritise the correct genetic diagnosis in over 95% of solved cases, and more generally by a study of paediatric patients referred for WGS in the intensive care setting, where the phenotypic description of the child was a poor predictor of the causative gene identified in 90% of cases [196].

A small number of variants in *FDXR* demonstrated consistent clinical severity across patients. This was exemplified by the homozygous p.Arg306Cys variant, with four patients invariably expressing a milder phenotype (median NPMDS score 3), and the homozygous p.Asp368Asn variant, with two patients expressing a severe phenotype (median NPMDS score 21.5). The recognition of such patterns, also recently reported for variants in the mitochondrial disease genes *HIBCH* and *ECHS1* [188], are valuable for genetic counselling. However, they require collection and report of multiple patients harboring the same variant(s) which is challenged by both the limited number of shared variants between mitochondrial disease patients and the vast number of variants

inherited in a compound heterozygous manner, whereby the consequence of two alleles of potentially different effect needs to be taken into consideration. Development of GENOMITexplorer (discussed in Chapter 4.1) takes steps towards providing this data to the mitochondrial disease community, by providing patient-level HPO terms and their corresponding causative variant(s) for over 1,000 genetically diagnosed patients.

To date, there is no specific biomarker for *FDXR* defects, and despite Fe-S cluster inclusion in mitochondrial RCCI, II, and III, no consistent defect in OXPHOS activity was measured on the muscle biopsies of 12 patients (akin to other disorders of Fe-S cluster biosynthesis [198, 199]). Moreover, alike to the mitochondrial diseases detected in the WES study presented in this thesis, the vast majority of disease-causing variants were missense (81%, 25/31), and prior to essential functional validation were classified as variants of uncertain significance by the ACMG criteria (US, class 3). As discussed in Chapter 4.1, patient-derived bio-material is a valuable resource for the study of mitochondrial diseases. However, fibroblast cell lines were only available for two of the seven patients in this study. This necessitated the use of a model organism, of which yeast models (*S. cerevisiae*) offer ease in genetic manipulation and rapid growth. Approximately 40% of human genes whose mutations lead to disease have an ortholog in yeast [200], and *FDXR* is amongst the ∼200 mitochondrial disease genes with a yeast ortholog (i.e., *Arh1*). In this study, the *Arh1*-null yeast model was utilised to successfully validate seven novel and three previously reported pathogenic *FDXR* variants. The pathogenicity of these variants was confirmed by stepwise changes to the growth conditions, sequentially stressing the *Arh1*-null yeast cells expressing mutant FDXR proteins. The fact that all published variants have functional evidence demonstrates the power and ease of a yeast system for validation experiments, also demonstrated for variants in *POLG*, *SDHAF1*, and *MTO1* [201, 202, 203], amongst many others (reviewed in [204]). Moreover, the yeast model may be utilised in modeling mitochondrial disease mechanisms and discovering therapies [205].

Although this assay successfully determined the pathogenicity of VUS, it does have limitations. First, expression of human FDXR is able to rescue the *Arh1*-null yeast model, however, this is not the case for all human genes and their respective ortholog in yeast. Thereby, when this is not the case, mutagenesis of the yeast ortholog gene itself is the only option, and issues arise when patient variants are not conserved as far down as to yeast. Second, the pathogenicity of a variant becomes questionable if growth impairment only manifests in the most unfavorable of growth conditions, in which differences may even be observed between wild-type and common polymorphism. To address this limitation, the study presented here included a common polymorphism as a negative control, selecting p.Arg123Gln (allele frequency 20.1% in gnomAD) and only designating variants as pathogenic when the associated growth defect was more severe

than that associated with the polymorphism. Third, there was no correlation between the severity of the growth defect in the yeast model and the severity of the patient's clinical manifestation, indicating that the model does not have value is phenotype severity prediction. These data should not therefore be relied upon to direct genetic counselling.

4.3 Novel disease gene discovery and validation

With the advent of NGS in 2010, the pace of novel mitochondrial disease gene discovery quickened, with approximately 20-25 disease genes discovered per year (see **Fig. 1.4**). In recent years however, the rate of discovery appears to be plateauing, presumably as the "low hanging fruit" has been picked. The detection of more challenging diagnoses is aided by the collection of patient data sets by international networks, as exemplified and discussed in Chapter 4.1. Such close collaboration increases the signal for novel disease gene identification, simply by increasing the chance to identify a gene defect in multiple case when the disease is extremely rare, as is true for a number of mitochondrial diseases where there are less than 10 cases published. In this thesis, the discovery of a novel disease gene harbouring variants of higher than expected allele frequency and incomplete penetrance further exemplifies this point.

In this study, a recessive phenocopy of the most frequent maternally-inherited mitochondrial disease, LHON, due to mutations in the nuclear DNA encoded gene *DNAJC30* was identified. This discovery attributed a novel disease gene and an alternative mode of inheritance to the historic disease. In total, 33 patients were identified with biallelic variants in *DNAJC30*, of which 29 shared a homozygous missense founder variant (p.Tyr51Cys). Notably, given the unexpectedly high allele frequency of this variant for a mitochondrial disease (0.125% in gnomAD), the variant came under considerable scrutiny, and greatly contributed to the eventual rejection of the study by a high impact journal, as a reviewer questioned the potential of the variant to be pathogenic. However, in association with this variant, incomplete penetrance and male predominance were observed, phenomena seldom occurring in recessive disease, yet know to be associated with maternally-inherited LHON (mtLHON) [43]. The evidence supporting these observations derived firstly from the relatively high allele frequency (0.125% in gnomAD) of the founder variant and the equal distribution of the variant allele amongst genders, which should otherwise result in a more frequent disease of equal gender distribution in the setting of full penetrance. In an effort to identify potential modifiers of penetrance, rare variants in genes encoding mitochondrial CI subunits and assembly factors (given the CI defect underpinning the disease), and rare mtDNA variants were searched for. However, we did not find potentially modifying variants in these genomic regions (CI subunits, assembly factors, and the mtDNA) to

explain an increase in penetrance in these affected individuals. Moreover, measuring the abundance of DNAJC30 on the RNA and protein level in a substantial number of control fibroblast cell lines did not find a significant gender-difference in expression to explain the marked male predominance of the disease. Overall, the data indicated that the factor(s) driving incomplete penetrance and male predominance is likely to originate downstream of the primary genetic event, perhaps secondary to subtle anatomical, hormonal, or otherwise physiological discrepancies between genders, or due to as yet unidentified genetic or epigenetic factor(s). Therefore, akin to mtLHON, the factor(s) influencing gender-dependent penetrance remains elusive at this time.

The penetrance estimate provided by the study should be cautiously interpreted. First, due to reports of LHON manifesting up to the eighth decade of life [206], giving rise to the possibility that the asymptomatic carriers in the study may in the future manifest disease. However, given that the vast majority of the arLHON patients manifested the disease in the second to third decade of life, and that the current age of almost all of the asymptomatic homozygous variant carriers is within the range of 30-50 years, it can be assumed they will continue to be asymptomatic. Second, due to the relatively limited number of large pedigrees in which to accurately calculate the penetrance. Third, due to the inability to identify pedigrees with only asymptomatic carriers, as in the absence of affected siblings these carriers would not come to medical attention to be sequenced. A future epidemiological study is needed to extensively explore this matter and to provide accurate values for penetrance.

In keeping with the discussion in Chapter 4.1, this study provides a further example of the difficultly to predict phenotype from genotype in mitochondrial disease given the identification of one patient with Leigh syndrome in association with the founder variant, a markedly different phenotype to LHON. The occurance of LHON and Leigh syndrome in association with a shared genetic cause has also occasionally been reported for specific pathogenic mtDNA mutations affecting subunits of CI [46, 45]. This was a further notable similarity between arLHON and mtLHON. DNAJC30 defects thereby provide a further example of Leigh syndrome within a LHON/MELAS/Leigh syndrome spectrum. However, it cannot be absolutely excluded that the Leigh phenotype was driven by a yet unknown second genetic disease in this patient.

Prior to this study, LHON was considered to be an exclusively maternally-inherited disease. Infrequently, "LHON-like" visual loss has been described in association with nuclear DNA abnormalities, such as in patients with Charcot-Marie-Tooth disease, in optic atrophy associated with dominant variants in *OPA1* [207], and in a single report of recessive "LHON-like" optic neuropathy due to mutations in *NDUFS2* [208]. In contrast to these reports, the study presented here argues for the frequent occurrence

of a recessive from of LHON, leading to the suggested stratification of LHON into mtLHON and arLHON. This was effectively exemplified at the Research Centre for Medical Genetics in Moscow, where the p.Tyr51Cys *DNAJC30* founder mutation accounted for over 20% of genetically diagnosed LHON patients. For this reason, sequencing of the one exon gene *DNAJC30* in parallel with the complete mtDNA sequence is recommended to reduce the diagnostic gap in LHON, especially in the founder population. To date, the only mitochondrial disease with an EMA licensed treatment, idebenone, is LHON. In keeping with mtLHON, the arLHON patients demonstrated clear improvement in their visual acuity in association with idebenone treatment. Thereby, the identification of *DNAJC30* as a novel LHON-associated gene not only had implication for genetic counselling and diagnostic approach, but had immediate implication for treatment.

In arLHON, the mitochondrial CI defect characteristic of the disease was discovered to result from the impaired exchange of specific CI N-module subunits exposed to higher risk of oxidative damage. The exchange of these subunits maintains high CI functionality, and can occur at a lower energetic cost to complete *de novo* replacement of the complex [10, 184, 209, 210]. The level of mitochondrial CI impairment measured by oxygen consumption rate in patient-derived fibroblast cells lines was subtle. This subtlety may reflect the role of DNAJC30 in maintenance rather than in the structure or assembly of CI; akin to a factor optimizing CI to ensure full bioenergetic function. The finding is remarkable in a broader context than this LHON focused study, as it not only provided a novel function to an orphan protein, but demonstrated for the first time a disease based on the repair of a protein complex. A study by Tebbenkamp et al., [172] also demonstrated DNAJC30 to play a role in mitochondrial and neuronal function and morphology by interaction with the ATP-synthase machinery (mitochondrial CV), thereby facilitating ATP synthesis. In complement to the study presented here of defective *DNAJC30* in patient-derived fibroblast cell lines and in the *DNAJC30*-KO HEK cellular model, Tebbenkamp et al., utilized a *Dnajc30*-KO mouse model to demonstrate increased abundance of CI containing supercomplexes in association with defective mitochondrial respiration rate in cultured primary neurons. Given the results of this study, the impact of the pathogenic *DNAJC30* variants identified in the study on CV was considered. Analysing CV activity on muscle biopsy, CV subunit abundance by quantitative proteomics, and CV subunit turnover, failed to detect any difference to control. From the clinical perspective, Tebbenkamp proposed *DNAJC30* to contribute to the pathogenesis of Williams-Beuren syndrome, a 7q11.23 hemi-deletion of 26-28 genes inclusive of *DNAJC30*, which akin to LHON includes diminished neuronal function. However, homozygous and heterozygous *DNAJC30* variant carriers in the study did not display any clinical features of Williams-Beuren syndrome. Moreover, gnomAD reports 25 heterozygous carriers of protein-truncating

(loss-of-function) variants in addition to 351 heterozygous carriers of the p.Tyr51Cys founder variant (confirmed in the study to result in near-complete loss of the protein), comparable to a hemizygous deletion of the gene, in a reportedly healthy population. These lines of evidence argue that though defective *DNAJC30* may enhance the phenotype of the haploinsufficiency of the genes within the Williams-Beuren syndrome locus and contribute to the disease pathophysiology, *DNAJC30* defects alone are unable to recapitulate all elements of the syndrome.

Mitochondrial CI defects are the most frequent biochemical feature of mitochondrial disease [211, 212] and are more broadly implicated in the pathogenesis of common diseases (e.g., cancer, diabetes, parkinsonism) and in aging [213, 214, 215]. Thereby, the implication of *DNAJC30* in the function of mitochondrial CI in this study, highlights DNAJC30 as a target within the CI repair pathway with potential therapeutic implications far beyond the optic neuropathies.

4.4 Integration of multi-omic approaches to detect and validate diagnoses

The study of >2,000 suspected mitochondrial disease patients investigated by WES demonstrated 46% of WES analyses to be inconclusive (see Chapter 3.1). These inconclusive results were partially due to VUS in know disease genes or potential novel disease genes requiring further functional evidence of pathogenicity (14%). The majority, however, were due to the inability to prioritise potentially pathogenic variants for further exploration (32%). These cases remained unsolved with no lead to follow.

By the study of inconclusive cases in a systematic integrative approach utilising genetic (WES/WGS), phenotypic (HPO descriptors), RNA-seq, and quantitative proteomic data, the ability to provide a diagnosis to a further ∼20% (26/121) of patients with suspected mitochondrial disease was demonstrated. Moreover, the capacity of these approaches to act together as a diagnostic discovery tool was exemplified by the 12/26 solved patients where no variant could be prioritised from the genetic data.

RNA-seq and proteomics provided functional evidence of pathogenicity and fulfilled criterion PM3 of the ACMG classification (see **Tab. 2**), a criterion of high weighing in determining overall pathogenicity. Proteomics was of especially high value in the validation of protein destabilising missense variants, a variant class with a functional consequence eluding detection by RNA-seq, and only detectable as protein underexpression outliers. Moreover, proteomics provided an additional layer of evidence to RNA expression outliers (typically due to frameshift, splice, and nonsense variants),

by confirming a corresponding loss of the protein. Additionally, the digitilisation of the phenotype data as HPO terms allowed for automated analysis of phenotype congruity with the established phenotype-genotype association in the literature, and provided further supportive evidence fulfilling criterion PP4 of the ACMG classification. The layering of these multiple forms of evidence of variant pathogenicity increased diagnostic confidence and provided the basis for family counselling and prenatal diagnostics, an area in which uncertainty is not acceptable.

In Chapter 3.4, two cases were presented in detail with diagnoses made in *DARS2* and *NDUFB11*. These diagnoses relied on high semantic similarity scoring in combination with functional evidence to reprioritise and simultaneously validate variants discarded in the WES analysis, due to features indicating the variants to be benign in nature. In *DARS2*, the allele frequency of two near-splice variants was far greater than expected for the disease (ACMG criterion BS1), and in *NDUFB11* lack of segregation of a missense variant with the disease in the patient's family (ACMG criterion BS2 and BS4) led to their discard, respectively. These cases exemplify that, though stringent filtering is important to reduce the burden of variant interpretation and to confidently identify a manageable number of potentially pathogenic variants to consider in more detail, a number of pathogenic variants will not conform to these criteria and will be overlooked. However, such variants can be successfully recovered by layering multiple levels of omics data.

The *DARS2* and *NDUFB11* cases exemplify a number of additional important points with regard to the scope of variant consequence detectable by multi-omics. First, the *NDUFB11* defect demonstrated the power of proteomics not only to detected under-expression of the protein encoded by the gene containing disease-causing variants, but also in the interaction partners of the respective protein complex, here mitochondrial CI, and more specifically the CI ND4-module to which NDUFB11 belongs. Given that approximately 20% of mitochondrial, and 45% of OMIM disease genes encode proteins that form part of a protein complex, according to the >2,350 protein complexes listed in CORUM (see Chapter 2.2.7), the downstream functional consequence on the protein complex is theoretically readily detectable. In keeping with this, pathogenic consequence on the protein complex level in a total of 11/26 solved cases (42%) was observed. Second, the *DARS2* defect demonstrated the added value of RNA-seq in the characterisation of aberrant splice patterns, a level of detail inaccessible to proteomics. This was vital in designating pathogenicity to a rare combination in *cis* of two frequent near-splice variants in *DARS2*, otherwise evading prioritisation due to their individual high allele frequency. Given the observed aberrant splice event, the defect could be pinpointed to a discrete region within a single intron of the gene (as depicted in **Fig. 3.36**) and the WES data could be manually inspected to identify the

causative variants. Thus, though RNA-seq does not allow the functional interpretation of missense variants, arguably a significant limitation in comparison to proteomics given the vast number of missense variants involved in Mendelian disease, it retains high value in providing detail on pathogenic RNA consequence, and in assigning potential pathogenic variants to reduced protein level in the absence of causative missense variation. Moreover, RNA-seq had deeper coverage of expressed genes than proteomics, capturing approximately 35% more gene products (median 10,425 RNA transcripts in comparison to 7,686 proteins in proteomics), and is thereby useful for the detection of pathogenic aberrations arising as of consequence of variants in genes encoding lowly expressed proteins.

The value of RNA-seq in diagnostics is reflected in multiple studies to date, complementing WES with RNA-seq for variant interpretation in unsolved WES/WGS cases. These studies identified the genetic diagnosis in up to 35% of cases across diverse indications [129, 130, 131], and 10-15% of suspected mitochondrial disease cases [132, 117]. However, as demonstrated in this study, RNA-seq contributed to the diagnosis of only half of the solved cases. Thereby, in this cohort, complementing WES with RNA-seq only would have capped the diagnostic rate at ∼10%. These figures argue for the diagnostic application of RNA-seq and proteomics in unison to gain the most comprehensive picture. With regards to proteomics, studies have previously utilised proteomics to elucidate underlying disease mechanism [98, 141, 142], to provide cumulative functional evidence to RNA seq candidates [117], and to resolve a small number of cases with suspected monogenic disease by guiding targeted genetic testing [216]. However, the utility of the systematic application and integration of quantitative proteomics with WES/WGS data into a diagnostic pipeline in a large cohort of unsolved cases had not been explored prior to this study.

Overall, with decreasing cost and the dissemination of robust publicly available analysis pipelines, the increasing democratisation of high-throughput omics assays will be the driving force for their integration into routine clinical practice. In this regard, the approach and code utilised in this study (detailed in [118]) provide a blueprint for implementing multi-omics based Mendelian disease diagnostics.

5 | Conclusion and outlook

In this chapter, a conclusive summary is provided with respect to the four central objectives of this thesis, formulated in Chapter 1.4.1, and an outlook is provided for the next frontiers in mitochondrial disease, and more generally Mendelian disease, diagnostics.

5.1 Conclusive summary

In the course of this thesis, four studies were presented with the overarching objective to provide a genetic diagnosis to unsolved mitochondrial disease patients, and specifically:

1. To analyse the diagnostic power of WES in >2,000 paediatric suspected mitochondrial disease patients, genetically and clinically characterise the study population by integration of HPO encoded phenotype descriptors, and explore the value of functional data integration from patient-derived bio-material. The diagnostic rate of WES in paediatric suspected mitochondrial disease was found to be 54% (1,093/2,035), stratified into 73% (800/1,093) of solved patients with defects in mitochondrial disease genes and 27% (293/1,035) of solved patients with

defects in other Mendelian disease genes clinically manifesting as a phenocopy of mitochondrial disease. The phenotypes enriched in solved mitochondrial disease patients were identified as encephalopathy, MRI features of Leigh syndrome (basal ganglia and brain stem abnormalities), cardiomyopathy, elevated serum lactate, and mitochondrial RCC defects. Functional validation of VUS utilising patient-derived bio-material allowed their designation as P/LP in 14% (286/2,035) of solved cases, without which the diagnosis rate would have capped at approximately 35%.

2. To validate novel variants in a known disease gene, *FDXR*, by employing a yeast model as a valuable alternative in the absence of patient-derived bio-material.

   Seven novel missense and in-frame indel VUS in *FDXR* were functionally validated to be pathogenic in a yeast model, leading to an expansion of the associated phenotype severity described for the disease.

3. To validate a novel mitochondrial disease gene, *DNAJC30*, and provide insight in a novel biological pathway by characterisation of the disease pathomechanism in patient-derived cell lines and a cellular model.

   DNAJC30 was validated as a novel mitochondrial disease gene resolving the enigma of LHON in the absence of pathogenic mtDNA variants. Functional studies of patient-derived fibroblast cell lines and a *DNAJC30*-KO cellular model led to the discovery of a repair mechanism in mitochondrial CI. *DNAJC30* defects were found to reduce mitochondrial CI function due to impairing the turnover of specific subunits of the CI N-module exposed to higher risk of oxidative damage and requiring maintenance.

4. To integrate WES, HPO encoded phenotype descriptors, RNA-seq, and quantitative proteomic data to discover and simultaneously validate the genetic diagnosis in inconclusive WES/WGS cases.

   The integration of multiple levels of omics data validated prioritised VUS from WES/WGS in 14/121 unsolved cases and simultaneously discovered and validated the genetic diagnosis in a further 12 cases, resulting in an overall diagnostic success rate of 21% (26/121).

5.2 General outlook

This thesis focused on the application of WES, limiting diagnostic focus to the coding regions of the genome, and resulting in a diagnostic gap of approximately 50%. Looking to the future, further increasing the diagnostic success in suspected mitochondrial disease can be envisaged to include:

1. Movement toward WGS to determine the additional contribution of non-coding variation and structural variation in mitochondrial disease
2. Routine integration of complementary multi-omics methods, principally RNA-seq and quantitative proteomics, to WES/WGS data analysis in diagnostics to provide a high-throughput functional readout of variant consequence

WGS has the potential to overcome a number of shortfalls of WES. First, in the ability of long-read technology to sequence genomic regions posing a challenge to short read WES sequencing (∼150-300 bp). long-read WGS technology fragments genomic DNA to approximately 10 kb in length and uses PCR-free library preparation. This streamlines the workflow and eliminates amplification bias for coverage uniformity across the genome, thereby accurately genotyping regions typically subject to low depth, low alignment scores, or low base quality [217]. Long-read WGS is thereby clinically valuable in the detection of large deletions and insertions [218, 219], structural variants (e.g., chromosomal rearrangements including inversions, translocations, deletions, duplications in various neurodevelopmental and metabolic diseases [220, 221]), and in distinguishing stretches of highly repetitive genomic elements, such as repeat expansions (e.g., in fragile X syndrome [222], myotonic dystrophy [129], and in number of novel pathogenic repeat expansion discoveries [223, 224]). Moreover, it can resolve variant allele phasing [225] and distinguish potentially pathogenic variants in a disease gene from its pseudogene(s) [226]. Second, in the detection of non-coding variation (e.g., deep intronic and regulatory variants). In this regard, as WGS captures all regions of the 3.2 billion nucleotide genome, within which approximately four million variants are called, it comes with the caveat of more difficult variant interpretation, given the unknown functional consequence of non-coding variation. This will necessitate the integration complementary multi-omics methods.

Incomplete penetrance and variable phenotype expressivity suggest a contribution of additional genetic variants to the clinical manifestation of mitochondrial disease. Thereby, the next frontiers in mitochondrial disease genetics are forseen to include the identification of:

1. Genetic modifiers
2. Oligogenic inheritance of rare variants
3. Polygenic inheritance of common variants

A genetic modifier is defined as a variant allele modulating the effect of a second disease-causing variant allele, for example by modifying the penetrance, phenotype expression, and disease severity. So far, only a handful of genetic modifiers have been described in mitochondrial disease, primarily nuclear variants modifying the expression of pathogenic mtDNA variants. A prime example is in LHON, where modifier variants in *PRICKLE3* [227] and *YARS2* [228] have been reported to increase the penetrance of the m.11778G>A *MT-ND4* LHON-associated variant. Though examples in patients are limited, a study developing a compendium of almost 200 genetic modifiers of mitochondrial dysfunction, identified by genome-wide CRISPR-Cas screens, may in the future help to prioritise genetic variants in the explanation of incomplete penetrance, variable phenotype, and tissue-specific expression [229].

Thus far, diagnostic focus in human genetics has been on the detection of monogenic disease and Mendelian inheritance, predominantly due to fully penetrant rare variants segregating with the disease (see **Fig. 5.1**). Looking beyond Mendelian inheritance, there have been a small number of reports of digenic disease in mitochondrial disease, the simplest form of oligogenic inheritance. Digenic inheritance is defined as the inheritance of variants in two genes explaining the phenotype of the patient more clearly than the variant(s) at one locus alone. To date, DIDA, a curated digenic database (see Chapter 2.2.7) lists 258 digenic combinations across 169 disease genes, causative of 54 diseases [230]. Amongst these reports are just two patients with digenic inheritance of mitochondrial disease, with pathogenic mutations in two genes involved in the maintenance and stability of mtDNA, *POLG* and *TWINKLE* [231, 232] leading to CPEO with multiple deletions in the mtDNA. These discoveries were made by candidate gene approaches, and there remains a need to develop genome-wide statistical tools focusing on the digenic inheritance model for their detection going forward, such as a method for detecting gene to gene interactions from WES data recently described by [233].

Beyond searching in the rare variant space, the phenotypic variability of mitochondrial disease may be attributable to more complex genetic interplay, such as a collective interaction of "weaker" more common alleles between patients with the same primary rare genetic defect. This has recently been reported for a small number of Mendelian diseases, inclusive of familial hypercholesterolaemia, monogenic obesity, and sex-chromosome aneuploidies [234]. The study utilised data from large genome-wide association

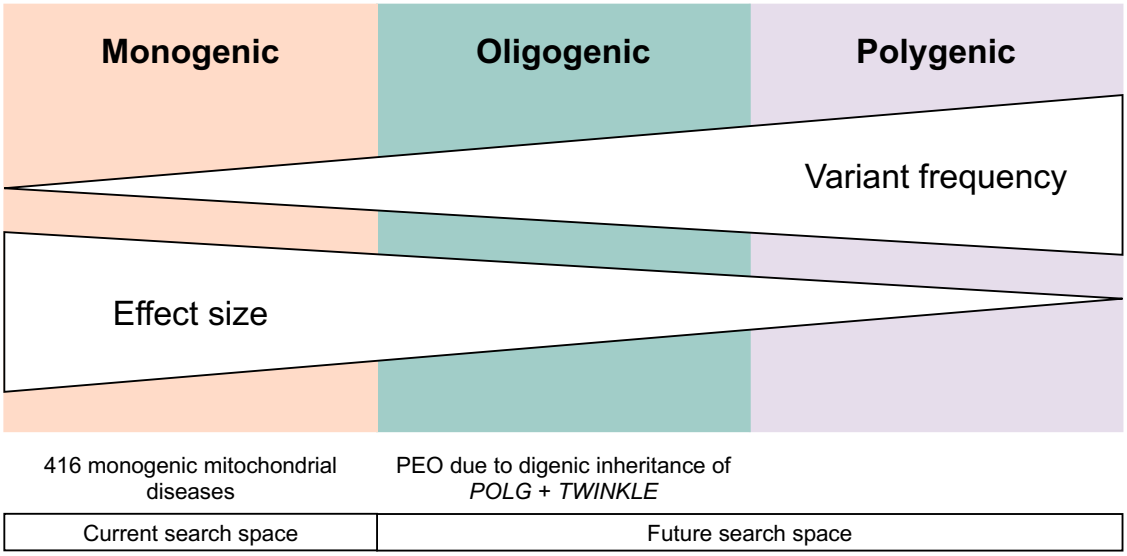


Figure 5.1: The next frontiers in mitochondrial disease genetics.
Inheritance depicted by variant allele frequency and variant effect size. Current examples in mitochondrial disease are listed below. The current monogenic search space has led to the discovery of defects in 416 mitochondrial disease genes. In the future the search space can be expanded to potentially identify oligogenic and polygenic inheritance in mitochondrial disease.

studies (GWAS) of relevant complex traits (cholesterol level, BMI, and height for the aforementioned diseases, respectively) to demonstrate the involvement of common polygenic variants in the variable expressivity of rare disease phenotypes. Such an approach could be considered in the future to determine the differential involvement of a phenotype or tissue between mitochondrial disease patients with primary pathogenic variants in the same disease gene, such as in *ACAD9* where patient may (∼85%) or may not manifest with cardiomyopathy [189]. To delineate these subtle interactions, a large volume of harmonised genotypic and phenotypic data is needed, underlining the importance of international collaboration and data sharing in the rare disease field, as exemplified by the WES study of >2,000 patients described in this thesis and by the creation of GENOMITexplorer.

Appendices

The appendices include a modified version of the Newcastle Paediatric Mitochondrial Disease Scale to compare patients across different ages, the Mitochondrial Disease Criteria developed by Morava et al., [50] modified for scoring by HPO terms, and the Mitochondrial Disease Criteria developed by Witters et al., [51] modified for scoring by HPO terms, a list of targeted gene panels for "Mitochondrial disease", and HPO criteria developed for Leigh syndrome.

Modified Newcastle Paediatric Mitochondrial Disease Scale

	Score	0-24 months	2-11 years	12-18 years
Vision	**0**	**Normal**. No parental concerns **/** Normal. Normal fixation and tracking	**Normal**. No parental / patient concerns **/ Normal.** Visual acuity better than or equal to 6/12 or normal fixation and tracking	**Normal.** No parental / patient concerns / Normal. Visual acuity better than or equal to 6/12 or normal fixation and tracking
	1	**Mild**. Limited eye or head movement to large objects or parental face in visual field **/ Mild.** Impaired fixation and / or tracking of small objects	**Mild**. Inattention to small objects in visual field **or** parent concerned about abnormality of visual behaviour **/ Mild.** Acuity worse than 6/12 but better than or equal to 6/18 or no fixation on small objects	**Mild.** Difficulty with small print or text on the television / **Mild.** Acuity worse than 6/12 but better than or equal to 6/18 or no fixation on small objects
	2	**Moderate**. No response to large objects or parental face in the visual field **/ Moderate.** Impaired fixation and / or tracking of familiar faces	**Moderate**. Visual impairment not fully corrected with glasses **or** inattention to large objects in visual field **/ Moderate.** Acuity worse than 6/18 but better than or equal to 6/60 or impaired fixation on large, brightly coloured objects	**Moderate.** Difficulty outside the home (e.g. bus numbers, road signs or shopping) / **Moderate.** Acuity worse than 6/18 but better than or equal to 6/60 or impaired fixation on large brightly coloured objects
	3	**Severe**. No response to light **/ Severe.** No response to light or registered blind	**Severe**. Not recognising faces **or** registered blind **or** using additional visual aids **/ Severe.** Acuity worse than 6/60 or no response to light or visual threat or unable to finger count	**Severe.** Difficulty recognising faces or unable to navigate without help or registered blind / **Severe.** Acuity worse than 6/60 or no response to light or visual threat or unable to finger count
Hearing	**0**	**Normal**	**Normal**	**Normal**
	1	**Mild**. Body, head or eye movement only to loud noise	**Mild**. Requires regular repetition / raised voice **or** not reacting to loud sounds	**Mild.** Mild deafness (e.g. missing words in presence of background noise)
	2	**Moderate**. No reaction to loud noise	**Moderate**. Hearing impaired but fully corrected with hearing aid	**Moderate.** Hearing impaired but fully corrected with hearing aid
	3	**Severe**. No hearing (even with aid)	**Severe**. Poor hearing even with aid	**Severe.** Poor hearing even with aid or reliant on non-verbal communication
Feeding	**0**	**Normal**	**Normal**	**Normal**
	1	**Mild**. Difficulties in sucking / coughing / anorexia / wheezy with feeds or occasional choking	**Mild**. Choking / vomiting / anorexia resulting in reduced intake **or** adaptation of age-appropriate diet	**Mild.** Choking / vomiting / anorexia resulting in reduced intake or swallowing difficulties requiring dietary adaptations
	2	**Moderate**. Supplementary enteral feeding or recurrent aspiration pneumonia	**Moderate**. Supplementary enteral feeding **or** recurrent aspiration pneumonia	**Moderate.** Supplementary enteral feeding or recurrent aspiration pneumonia
	3	**Severe**. Exclusive enteral feeding (gastrostomy / NG tube). Nil by mouth	**Severe**. Exclusive enteral feeding (gastrostomy / NG tube). Nil by mouth	**Severe.** Exclusive enteral feeding (gastrostomy / NG tube). Nil by mouth
Mobility	**0**	**Normal**. No concerns. Age appropriate mobility	**Normal**. No concerns. Age appropriate mobility	**Normal**
	1	**Mild**. Clumsy age appropriate mode of mobility	**Mild**. Difficulty walking upstairs or inclines	**Mild.** Difficulty walking upstairs or inclines
	2	**Moderate**. Mobile but through age inappropriate mode	**Moderate**. Requires support (stick / frame / callipers) to walk on the flat	**Moderate.** Requires support (stick / frame / callipers) to walk on the flat
	3	**Severe**. Immobile	**Severe**. Wheelchair / carrier dependent	**Severe.** Wheelchair dependent
Seizures	**0**	**None**	**None**	**None**
	1	**Mild**. Myoclonic or absence seizures only **or** <	**Mild**. Myoclonic or absence seizures only **or** <	**Mild.** Myoclonic or absence seizures only or <

		1 generalised tonic-clonic seizure/month	1 generalised tonic-clonic seizure/month	1 generalised tonic-clonic seizure each month
	2	**Moderate**. > 5 generalized tonic-clonic seizures/month **or** > 20 absence or myoclonic seizures/month	**Moderate**. > 5 generalized tonic-clonic seizures/month **or** > 20 absence or myoclonic seizures/month	**Moderate.** > 5 generalized tonic-clonic seizures/month or > 20 absence or myoclonic seizures/month
	3	**Severe**. Status epilepticus or intractable seizures	**Severe**. Status epilepticus	**Severe.** Status epilepticus
Encephalopathy	**0**	**None**	**None**	**None**
	1	**Mild**. Abnormal sleepiness / lethargy. Waking only for feeds	**Mild**. Single episode of personality change, excessive sleepiness, confusion or disorientation	**Mild.** Single episode of personality change, excessive sleepiness, confusion or disorientation
	2	**Moderate**. Recurrent episodes of mild encephalopathy (> 2/year)	**Moderate**. Obtunded **or** >2 encephalopathic episodes/year	**Moderate.** Obtunded or >2 encephalopathic episodes/year
	3	**Severe**. Life threatening encephalopathy - requires artificial ventilation	**Severe**. Life-threatening encephalopathy- requires artificial ventilation	**Severe.** Life-threatening encephalopathy- requires artificial ventilation
Gastrointestinal	**0**	**Normal**.	**Normal**	**Normal**
	1	**Mild**. Constipation or unexplained vomiting / diarrhoea > 3/week	**Mild**. Mild constipation **or** unexplaine d vomiting / diarrhoea < 1/week	**Mild.** Mild constipation or unexplained vomiting / diarrhoea < 1/week
	2	**Moderate**. Severe constipation (no relief with laxative treatment) **or** unexplained vomiting / diarrhoea every day **or** surgical intervention for dysmotility	**Moderate**. Moderate constipation (some relief with laxative treatment) **or** unexplained vomiting / diarrhoea > 3/week	**Moderate.** Moderate constipation (some relief with laxative treatment) or unexplained vomiting / diarrhoea > 3/week
	3	**Severe** Malabsorption / Failure to thrive	**Severe**. Severe constipation (no relief with laxative treatment) **or** unexplained vomiting / diarrhoea every day **or** surgical intervention for dysmotility	**Severe.** Severe constipation (no relief with laxative treatment) or unexplained vomiting / diarrhoea every day or surgical intervention for dysmotility
Endocrine	**0**	**Normal.**	**Normal**	**Normal**
	1	**Mild**. Biochemical evidence of impaired function	**Mild**. Biochemical evidence of impaired function	**Mild.** Biochemical evidence of impaired function
	2	**Moderate**. Endocrine failure requiring replacement therapy	**Moderate**. Endocrine failure requiring replacement therapy	**Moderate.** Endocrine failure requiring replacement therapy
	3	**Severe**. Endocrine decompensation (e.g. diabetic ketoacidosis, Addisonian crisis)	**Severe**. Decompensation (e.g. diabetic ketoacidosis, Addisonian crisis)	**Severe.** Endocrine decompensation (e.g. diabetic ketoacidosis, Addisonian crisis)
Respiratory	**0**	**Normal**	**Normal**	**Normal**
	1	**Mild**. Abnormal respiratory pattern not requiring therapy / hospitalization	**Mild**. Abnormal respiration not requiring hospitalization	**Mild.** Abnormal respiration not requiring hospitalization
	2	**Moderate**. Abnormal respiration requiring oxygen flow or hospitalisation but **not** ventilation	**Moderate**. Abnormal respiration requiring hospitalisation but **not** ventilation	**Moderate.** Abnormal respiration requiring hospitalisation but not ventilation
	3	**Severe**. Abnormal respiration requiring artificial ventilation	**Severe**. Abnormal respiration requiring artificial ventilation	**Severe.** Abnormal respiration requiring artificial ventilation
Cardiovascular	**0**	**Normal**	**Normal**	**Normal**
	1	**Mild**. Asymptomatic ECG change	**Mild**. Asymptomatic ECG change	**Mild.** Asymptomatic ECG change
	2	**Moderate**. Abnormal echocardiogram (e.g. cardiomegaly) **or** sustained / symptomatic arrhythmia on ECG	**Moderate**. Abnormal echocardiogram (e.g. cardiomegaly) **or** sustained / symptomatic arrhythmia on ECG	**Moderate.** Abnormal echocardiogram (e.g. cardiomegaly) or sustained / symptomatic arrhythmia on ECG
	3	**Severe**. Decompensated cardiomyopathy **or** requirin g pacing device / defibrillator / ablation	**Severe**. Decompensated cardiomyopathy **or** requirin g pacing device / defibrillator / ablation	**Severe.** Decompensated cardiomyopathy or requiring pacing device / defibrillator / ablation
Renal	**0**	**Normal**	**Normal**	**Normal**

	1	**Mild**. Impaired function but no change in diet or therapy required	**Mild**. Impaired function but no change in diet or therapy required	**Mild.** Impaired function but no change in diet or therapy required
	2	**Moderate**. Impaired function requiring restricted protein diet	**Moderate**. Impaired function requiring restricted protein diet	**Moderate.** Impaired function requiring restricted protein diet
	3	**Severe**. Failure requiring transplant / dialysis	**Severe**. Failure requiring transplant / dialysis	**Severe.** Failure requiring transplant / dialysis
Liver	0	**Normal**	**Normal**	**Normal**
	1	**Mild**. Mildly impaired Liver Function Tests (LFTs). Normal albumin and coagulation. No symptoms of hepatic failure	**Mild**. Mildly impaired Liver Function Tests (LFTs). No symptoms of hepatic Failure	**Mild.** Mildly impaired Liver Function Tests (LFTs). No symptoms of hepatic failure
	2	**Moderate**. Impaired LFTs with symptoms (e.g. jaundice, coagulation anomalies, oedema)	**Moderate**. Impaired LFTs with symptoms (e.g. jaundice, oedema)	**Moderate.** Impaired LTFs with symptoms (e.g. jaundice, oedema)
	3	**Severe**. Failure requiring hospitalisation **and / or** transplantation	**Severe**. Failure requiring hospitalisation **and / or** transplantation	**Severe.** Failure requiring hospitalisation and / or transplantation
Blood	0	**Normal**	**Normal**	**Normal**
	1	**Mild**. Anaemia only	**Mild**. Anaemia only	**Mild.** Anaemia only
	2	**Moderate**. Asymptomatic pancytopenia	**Moderate**. Asymptomatic pancytopenia	**Moderate.** Asymptomatic pancytopenia
	3	**Severe**. Pancytopenia requiring regular transfusion / transplantation	**Severe**. Pancytopenia requiring regular transfusion / transplantation	**Severe.** Pancytopenia requiring regular transfusion / transplantation
Growth	0	**Normal**. Following normal growth trajectory	**Normal**. Following normal growth trajectory	**Normal.** Following normal growth trajectory
	1	**Mild**. Weight less than second centile but growing parallel to it	**Mild**. Height or weight or both less than 2nd centile but growing parallel to it	**Mild.** Height or weight or both less than 2nd centile but growing parallel to it
	2	**Moderate**. Weight crossing one centile	**Moderate**. Height or weight or both crossing one centile	**Moderate.** Height or weight or both crossing one centile
	3	**Severe**. Weight crossing ≥ 2 centiles **or** less than 2nd centile with divergent trajectory	**Severe**. Height or weight or both crossing ≥ 2 centiles **or** less than 2nd centile with divergent trajectory	**Severe.** Height or weight or both crossing ≥ 2 centiles or less than 2nd centile with divergent trajectory
Development	0	**Normal**	**Normal**	**Normal**
	1	**Isolated** motor delay / delayed speech, language, or hearing with developmental progression	**Isolated** motor delay / delayed speech, language, or hearing with developmental progression	**Isolated** motor delay / delayed speech, language, or hearing with developmental progression
	2	**Isolated** motor delay / delayed speech, language, or hearing without developmental progression	**Isolated** motor delay / delayed speech, language, or hearing without developmental progression	**Isolated** motor delay / delayed speech, language, or hearing without developmental progression
	5	**Global** developmental delay with developmental progression	**Global** developmental delay with developmental progression	**Global** developmental delay with developmental progression
	6	**Global** developmental delay without developmental progression	**Global** developmental delay without developmental progression	**Global** developmental delay without developmental progression
	7	**Regression** in development	**Regression** in development	**Regression** in development
Ptosis and eye movement	0	**Normal**	**Normal**	**Normal**
	1	**Mild**. Gaze evoked nystagmus **or** unilateral ptosis **or** impaired eye movement at extremities	**Mild**. Gaze evoked nystagmus **or** unilateral ptosis **or** impaired eye movement at extremities	**Mild.** Gaze evoked nystagmus or unilateral ptosis or impaired eye movement at extremities
	2	**Moderate**. Intermittent nystagmus at rest **or** bilateral ptosis not obscuring pupils **or** restriction of >50% eye movement	**Moderate**. Intermittent nystagmus at rest **or** bilateral ptosis not obscuring pupils **or** restriction of >50% eye movement	**Moderate.** Intermittent nystagmus at rest or bilateral ptosis not obscuring pupils or restriction of >50% eye movement
	3	**Severe**. Continuous nystagmus at rest **or** bilateral ptosis obscuring pupils **or** only a flicker of eye movement	**Severe**. Continuous nystagmus at rest **or** bilateral ptosis obscuring pupils **or** only a flicker of eye movement	**Severe.** Continuous nystagmus at rest or bilateral ptosis obscuring pupils or only a flicker of eye movement

Myopathy	**0**	**Normal**	**Normal**	**Normal**
	1	**Mild**. Mild symmetrical weakness of hip and / or shoulder girdle only	**Mild**. Mild symmetrical weakness of hip and / or shoulder girdle only	**Mild.** Mild symmetrical weakness of hip and / or shoulder girdle only
	2	**Moderate**. Moderate symmetrical weakness (proximal > distal) limiting functional movement	**Moderate**. Moderate symmetrical weakness (proximal>distal) limiting mobility	**Moderate.** Moderate symmetrical weakness (proximal > distal) limiting mobility
	3	**Severe**. Wheelchair / carrier dependent **or** respiratory compromise due to myopathy.	**Severe**. Wheelchair / carrier dependent **or** respiratory compromise due to Myopathy	**Severe.** Wheelchair dependent or respiratory compromise due to myopathy
Ataxia	**0**	**Normal**	**Normal**	**Normal**
	1	**Mild**. Ataxic gait but walks unaided **or** mild upper limb dysmetria	**Mild**. Ataxic gait but walks unaided **or** mild upper limb dysmetria	**Mild.** Ataxic gait but walks unaided or mild upper limb dysmetria
	2	**Moderate**. Gait abnormality requiring assistance **or** severe upper limb Dysmetria	**Moderate**. Gait abnormality requiring assistance **or** severe upper limb Dysmetria	**Moderate.** Gait abnormality requiring assistance or severe upper limb dysmetria
	3	**Severe**. Wheelchair dependent **or** unable to feed due to ataxia	**Severe**. Wheelchair dependent **or** unable to feed due to ataxia	**Severe.** Wheelchair dependent or unable to feed due to ataxia
Pyramidal	**0**	**Normal**	**Normal**	**Normal**
	1	**Mild**. Unilateral pyramidal signs but retaining functional movement	**Mild**. Mild hemiplegia allowing unaided ambulation	**Mild.** Mild hemiplegia allowing unaided ambulation
	2	**Moderate**. Dense hemiplegia with little movement of affected side	**Moderate**. Moderate hemiplegia allowing ambulation with aids	**Moderate.** Moderate hemiplegia allowing ambulation with aids
	3	**Severe**. Bilateral pyramidal weakness with little or no movement	**Severe**. Wheelchair dependent due to hemi / tetraplegia	**Severe.** Wheelchair dependent due to hemi / tetraplegia
Extrapyramidal	**0**	**Normal.**	**Normal**	**Normal**
	1	**Mild**. Unilateral extrapyramidal posturing and increased tone	**Mild**. Focal dystonia **or** unilateral extrapyramidal tremor / bradykinesia	**Mild.** Focal dystonia or unilateral extrapyramidal tremor / bradykinesia
	2	**Moderate**. Bilateral extrapyramidal posturing and increased tone	**Moderate**. Generalised dystonia **or** bilateral extrapyramidal tremor / Bradykinesia	**Moderate.** Generalised dystonia or bilateral extrapyramidal tremor / bradykinesia
	3	**Severe**. Severe extrapyramidal posturing resulting in very little movement	**Severe**. Wheelchair dependent due to extrapyramidal disorder	**Severe.** Wheelchair dependent due to extrapyramidal disorder
Neuropathy	**0**	**Normal.**	**Normal.**	**Normal.**
	1	**Mild.** Areflexia only	**Mild.** Areflexia only	**Mild.** Areflexia only
	2	**Moderate.** Sensory ataxia or motor impairment (distal weakness) but mobile	**Moderate.** Sensory ataxia or motor impairment (distal weakness) but mobile	**Moderate.** Sensory ataxia or motor impairment (distal weakness) but mobile
	3	**Severe.** Reliant on mobility aids primarily due to neuropathy	**Severe.** Reliant on mobility aids primarily due to neuropathy	**Severe.** Reliant on mobility aids primarily due to neuropathy

Total NPMDS score is the sum of all system scores for the corresponding age range of the patient. Maximal score of 67.

Modified Mitochondrial Disease Criteria developed by Morava

Clincial signs and symptoms, 1 point/symptom (max. 4 points)			**Metabolic/ imaging studies (max. 4 points)**	**Tissue histopathology and biochemistry (max. 4 points)**
Muscular presentation (max. 2 points)	CNS presentation (max. 2 points)	Multisystem disease (max. 3 points)		
Ophthalmoplegia (HP:0000602)*	Neurodevelopmental abnormality (HP:0012759) **or** Intellectual disability (HP:0001249)	Hematological abnormality (HP:0001871) **or** Immune abnormality (HP:0002715)	Increased serum lactate (HP:0002151)*	Ragged-red muscle fibers (HP:0003200)**
Ptosis (HP:0000508) **or** Myopathic facies (HP:0002058)	Delayed speech and language development (HP:0000750)	(Digestive system abnormality (HP:0025031) **or** Decreased liver function (HP:0001410)* or Abnormality of the liver (HP:0001392)	Elevated lactate:pyruvate ratio (HP:0032653)	COX negative muscle fibers (HP:0003688)**
Exercise intolerance (HP:0003546) **or** Fatigue (HP:0012378)	Developmental regression (HP:0002376)	Endocrine abnormality (HP:0000818)	Increased serum alanine (HP:0003348)*	Abnormal mitochondrial morphology (HP:0008322)*
Muscle weakness (HP:0001324) **or** Myopathy (HP:0003198) **or** Muscular hypotonia (HP:0001252)	Stroke-like episode (HP:0002401)	Growth abnormality (HP:0001507)	Increased CSF lactate (HP:0002490)*	Decreased activity of mitochondrial respiratory chain (HP:0008972)* **or** Decreased activity of the pyruvate dehydrogenase complex (HP:0002928)*
Rhabdomyolsis (HP:0003201) **or** Increased serum creatine kinase (HP:0003236)	Migraine (HP:0002076)	Abnormality of the cardiovascular system (HP:0001626) **or** Cardiomyopathy (HP:0001638)*	Increased CSF alanine (HP:0500233)*	Mitochondrial depletion (HP:0030059)
Motor developmental delay (HP:0001270)	Seizure (HP:0001250) **or** Encepahlopathy (HP:0001298)	Abnormal renal physiology (HP:0012211)	Increased CSF protein (HP:0002922)	
EMG abnormality (HP:0003457)	Myoclonus (HP:0001336)	Abnormality of the eye (HP:0000478) **or** Visual impairment (HP:0000505) **or** Optic atrophy (HP:0000648)* **or** Leber optic atrophy (HP:0001112)*	Ethylmalonic aciduria (HP:0003219) **or** Methylmalonic aciduria (HP:0012120)	
	Cerebral visual impairment (HP:0100704)	Hearing impairment* (HP:0000365)	3-Methylglutaconic aciduria (HP:0003535)	
	Pyramidal signs (HP:0002493) **or** Spasticity (HP:0001257)	Peripheral neuropathy (HP:0009830)	Stroke-like picture (HP:0002401)	
	Extrapyramidal signs (HP:0002071) **or** Dystonia (HP:0001332)	Family history (HP:0032316)	Abnormality of the basal ganglia (HP:0002134)*	
	Ataxia (HP:0001251)		Abnormality of the brain stem (HP:0002363)*	
			Abnormality of the cerebral white matter (HP:0002500)	
			Elevated brain lactate level by MRS (HP:0012707)	

* these specific symptoms score 2 points.

** these specific symptoms score 4 points.

Score: 1 (unlikely), 2-4 (possible), 5-7 (probable), 8-12 (definite).

Modified Mitochondrial Disease Criteria developed by Witters

Clincial score, 1 point/symptom (max. 4 points)			**Metabolic and MRI score, 1 point/symptom (max. 4 points)**	
Muscular (max. 2 points)	Neurological (max. 2 points)	Multisystem (max. 3 points)	Metabolic (max. 4 points)	Imaging (max. 4 points)
Muscle weakness (HP:0001324) **or** Myopathy (HP:0003198) **or** Muscular hypotonia (HP:0001252)	Neurodevelopmental abnormality (HP:0012759) **or** Intellectual disability (HP:0001249)	Digestive system abnormality (HP:0025031)	Increased serum lactate (HP:0002151)*	Abnormality of the basal ganglia (HP:0002134)* **or** Abnormality of the brainstem (HP:0002363)*
EMG abnormality (HP:0003457)	Delayed speech and language development (HP:0000750)	Growth delay (HP:0001510) **or** Failure to thrive (HP:0001508)	Increased serum alanine (HP:0003348)	Stroke-like episode (HP:0002401)*
Motor delay (HP:0001270)	Dystonia (HP:0001332)	Endocrine abnormality (HP:0000818)	Ethylmalonic aciduria (HP:0003219) **or** Methylmalonic aciduria (HP:0012120)	Elevated brain lactate level by MRS (HP:0012707)
Exercise intolerance (HP:0003546)	Ataxia (HP:0001251)	Immune abnormality (HP:0002715)	3-Methylglutaconic aciduria (HP:0003535)	Abnormality of the cerebral white matter (HP:0002500)
Ophthalmoplegia (HP:0000602)	Spasticity (HP:0001257)	Visual impairment (HP:0000505) **or** Hearing impairment (HP:0000365)	Increased CSF lactate (HP:0002490) **or** Increased CSF alanine (HP:0500233)	Abnormality of the thalamus (HP:0010663)
	Peripheral neuropathy (HP:0009830)	Renal tubular dysfunction (HP:0000124)		Agenesis of the corpus callosum (HP:0001274)
	Seizure (HP:0001250) **or** Encephalopathy (HP:0001298)	Abnormality of the cardiovascular system (HP:0001626)		

* these specific symptoms score 2 points.
Score: 1 (unlikely), 2-4 (possible), 5-7 (probable), 8 (definite).

Targeted gene panels for "Mitochondrial Disease"

Panel identifier	Panel provider	Panel size
Mitochondrial Encephalopathy	MGZ Medical Genetics Center	131
Mitochondrial Diseases	MGZ Medical Genetics Center	168
Comprehensive mitochondrial disorders panel	Centogene AG - the Rare Disease Company	180
Comprehensive Mitochondrial Metabolic Panel	Knight Diagnostic Laboratories - Molecular Diagnostic Center	196
Mitochondrial Focused Nuclear Gene Panel	GeneDx	202
Mitochondrial Diseases	Asper Biogene	209
MitONE230	Upon request*	230
Combined Mito Genome Plus Mito Focused Nuclear Gene Panel	GeneDx	240
Mitochondrial diseases	CGC Genetics	268
MitoSure300	Upon request*	300
WES mitochondrial disorders	Translational Metabolic Laboratory	369
Metabolic Diseases incl. Mitochondriopathies	CeGaT	529
Nuclear-Mito NGS Panel	Fulgent Genetics	676

* custom panels utilised by the Unit of Medical Genetics and Neurogenetics, Fondazione IRCCS Istituto Neurologico Carlo Besta, contacts: Daniele.Ghezzi@istituto-besta.it, Costanza.Lamperti@istituto-besta.it.

HPO criteria for Leigh syndrome

Criteria 1	Criteria 2	Criteria 3
Neurodevelopmental delay (HP:0012758)	Abnormality of the basal ganglia (HP:0002134)	Increased serum lactate (HP:0002151)
or	or	or
Encephalopathy (HP:0001298)	Abnormality of the brain stem (HP:0002363)	Increased CSF lactate (HP:0002490)
or		
Developmental regression (HP:0002376)		

Patients must meet all three criteria to be considered indicative of Leigh syndrome. HPO terms were selected to reflect the clinical description of Leigh syndrome by [47].

Attachments

The attachments contain dedicated acknowledgments to everyone who supported me to complete the research included in this thesis, a list of published material, and an overview of all publications including first and co-authored articles both published and in preparation, submission, or revision.

Acknowledgments

Completing my Ph.D. research was an adventure, a step from medicine towards the uncharted waters of science, and an immensely rewarding experience that would not have been possible without the support and encouragement of many different people.

I would firstly like to express my gratitude to my Ph.D. advisor Prof. Dr. Thomas Meitinger, for the opportunity to join the captivating field of human genetics and for pushing the boundaries of my scientific thinking. It has been a privilege to be a member of his institute. My very deepest appreciation goes to Dr. Holger Prokisch, for accepting me into his research group, providing an immeasurable level of mentorship, and for his unfaltering encouragement to reach my highest potential. His creative scientific mind inspired the direction of my work and generated a continual drive to further evolve my scientific understanding and reasoning. I would also like to thank my Ph.D. mentors Prof. Dr. Thomas Klopstock and Prof. Dr. Peter Freisinger for their valued guidance through all stages of my Ph.D., for their contribution to my research development, and for their scientific and medical expertise, along with the fruitful collaborations we have shared.

Throughout my Ph.D., I was fortunate to meet with outstanding people, to share unforgettable moments, and most importantly true friendships. Anna Meier, Mirjana Gusic, and Enrica Zanuttigh were guiding lights both scientifically and personally. I am forever thankful for their time and understanding, along with introducing me to the pipette, for better or for worse. I would also like to thank all of my group members, former and present (Dimiitri Smirnov, Robert Kopajtich, Masaru Shimura, Vicente Yépez, Telka Wolstein, Riccardo Berutti, Caterina Terrile, Dewi Schlieben, Rui Ban, Manting Xu, Agnieszka Nadel, Laura Kremer, Marieta Borzes, and Michael Färberböck), along with other members of the Institute of Human Genetics (Arcangela Iuso, Thomas Schwarzmayr, Michael Zech), and Martina Kuhnert for her administrative support. They created a truly enjoyable environment to work in.

Furthermore, my thanks go to all of my collaboration partners and coauthors for their valuable contributions, rewarding scientific interactions, and for welcoming me into the mitochondrial disease community, especially to Dr. Ilka Wittig, Prof. Valerio Carelli, Dr. Kei Murayama, Prof. Robert McFarland, Prof. Robert Taylor, Dr. Johannes Mayr, Dr. Agnès Rötig, Dr. Costanza Lamperti, Dr. Daniele Ghezzi, Dr. Claudia Catarino, and Dr. Boriana Büchner. It has been a pleasure to work with them all. My special thanks go also to the patients from whom the samples and data underpinning my work was generously shared.

Finally, I am grateful to my friends, to my family, and to Markus Seeger. Their support was endless and unconditional, encouraging me always to follow my dreams, keeping me afloat and smiling through the challenges, and guiding me toward the ever more exciting times that are ahead.

Published material

The majority of the results shown in this thesis have been published in different research articles or are part of publications in preparation, in submission, or in revision at the time of writing this thesis.

- The results, figures, and tables presented in subsection 3.1 were in preparation for submission during the writing of this thesis.
- In subsection 3.2 the results, Fig. 3.15, 3.16, 3.18, 3.20, and 3.21, and Tab. 8, 9, and 10 were adapted from [37].
- In subsection 3.3 the results, Fig. 3.22 to 3.30, Fig. 3.31 to 3.34, and Tab. 11, 13, 14, and 15 were adapted from [123].
- In subsection 3.4 the results, Fig. 3.35 and 3.37 were adapted from the publication preprint [118] and were in review with *Nature Communications*.

List of publications

First author publications and preprints

1. R. Kopajtich*, D. Smirnov*, **S. L. Stenton***, (..), and H. Prokisch. "Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders". Preprint in: *MedRxiv*. (2021).

2. **S. L. Stenton**, J. Mayr, S. Wortmann, H. Prokisch. "Genomic Approaches for the Diagnosis of Inborn Errors of Metabolism". In: *Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases*. (2021).

3. **S. L. Stenton**, N. L. Sheremet, C. B. Catarino, (..), I. Wittig, and H. Prokisch. "Impaired complex I repair causes recessive Leber's hereditary optic neuropathy". In: *J. Clin. Invest*. (2021).

4. **S. L. Stenton**, Y. Zou, H. Cheng, H. Prokisch, and F. Fang. "Pediatric Leigh Syndrome: Neuroimaging Features and Genetic Correlations". In: *Ann. Neurol*. (2020).

5. **S. L. Stenton**, D. Piekutowska-Abramczuk, L. Kulterer, (..), and H. Prokisch. "Expanding the clinical and genetic spectrum of FDXR deficiency by functional validation of variants of uncertain significance". In: *Hum. Mutat*. (2020).

6. J. Zhou*, J. Li*, **S. L. Stenton***, X. Ren, S. Gong, F. Fang, and H. Prokisch. "NAD(P)HX dehydratase (NAXD) deficiency: A novel neurodegenerative disorder exacerbated by febrile illnesses". In: *Brain*. 143.2(2020) pp. E8.

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

8. **S. L. Stenton**, L. S. Kremer, R. Kopajtich, C. Ludwig, and H. Prokisch. "The diagnosis of inborn errors of metabolism by an integrative "multi-omics" approach: A perspective encompassing genomics, transcriptomics, and proteomics". In: *J. Inherit. Metab*. 43.1(2020), pp. 25-35.

9. **S. L. Stenton** and H. Prokisch. "The Clinical Application of RNA-sequencing in Genetic Diagnosis of Mendelian Disorders". In: *Clin. Lab. Med*. 40.2(2020), pp. 121-133.

10. **S. L. Stenton** and H. Prokisch. "The Clinical Application of RNA-sequencing in Genetic Diagnosis of Mendelian Disorders". In: *Adv. Mol. Pathol*. 1.1(2018), pp. 27-36.

11. **S. L. Stenton** and H. Prokisch. "Advancing genomic approaches to the molecular diagnosis of mitochondrial disease". In: *Essays Biochem*. 62.3(2018), pp. 399–408.

* equal contribution

In preparation, submission, or revision

12. **S. L. Stenton**, (..), and H. Prokisch. "Diagnosing paediatric mitochondrial disease: lessons from 2,000 exomes". *In submission.*

Co-author publications and preprints

13. M. Gusic, V. Yepez, (..), **S. L. Stenton**, (..), and H. Prokisch. "Clinical implementation of RNA-sequencing for Mendelian disease diagnostics". Preprint in: *MedRxiv*. (2021).

14. C. Alston, **S. L. Stenton**, G. Hudson, H. Prokisch, R. Taylor. "The genetics of mitochondrial disease: dissecting mitochondrial pathology using multi-omic pipelines". In: *J. Pathol.*. (2021).

15. J. Tan, M. Wagner, **S. L. Stenton**, T. M. Strom, S. B. Wortmann, H.Prokisch, T. Meitinger, K. Oexle, and T. Klopstock. "Lifetime risk of autosomal recessive mitochondrial disorders calculated from genetic databases". In: *EBioMedicine*. 54(2020).

16. L. Zhou, J. Deng, **S. L. Stenton**, J. Zhou, H. Li, C. Chen, H. Prokisch, and F. Fang. "Case Report: Rapid Treatment of Uridine-Responsive Epileptic Encephalopathy Caused by CAD Deficiency". In: *Front. Pharmacol.* 11(2020).

17. A. Esposito, A. Falace, M. Wagner, (..), **S. L. Stenton**, (..), and R. Guerrini. "Biallelic *DMXL2* mutations impair autophagy and cause Ohtahara syndrome with progressive course"". In: *Brain*. 142.12(2019), pp. 3876-3891.

References

[1] J. N. Timmis, M. A. Ayliff, C. Y. Huang, and W. Martin. "Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes". In: *Nat. Rev. Genet.* 5.2 (Feb. 2004), pp. 123–135.

[2] N. Lane and W. Martin. "The energetics of genome complexity". In: *Nature* 467.7318 (Oct. 2010), pp. 929–934.

[3] J. B. Spinelli and M. C. Haigis. "The multifaceted contributions of mitochondria to cellular metabolism". In: *Nat. Cell Biol.* 20.7 (July 2018), pp. 745–754.

[4] C. T. Walsh, B. P. Tu, and Y. Tang. "Eight Kinetically Stable but Thermodynamically Activated Molecules that Power Cell Metabolism". In: *Chem. Rev.* 118.4 (Feb. 2018), pp. 1460–1494.

[5] J. R. Friedman and J. Nunnari. "Mitochondrial form and function". In: *Nature* 505.7483 (2014), pp. 335–343.

[6] C. S. Ahn and C. M. Metallo. "Mitochondria as biosynthetic factories for cancer proliferation". In: *Cancer Metab.* 3.1 (Sept. 2015).

[7] L. A. Sena and N. S. Chandel. "Physiological roles of mitochondrial reactive oxygen species". In: *Mol. Cell* 48.2 (Oct. 2012), pp. 158–167.

[8] M. Schieber and N. S. Chandel. "ROS function in redox signaling and oxidative stress". In: *Curr. Biol.* 24.10 (May 2014).

[9] M. D. Brand. "The sites and topology of mitochondrial superoxide production". In: *Exp. Gerontol.* 45.7-8 (Aug. 2010), pp. 466–472.

[10] K. Szczepanowska et al. "A salvage pathway maintains highly functional respiratory complex I". In: *Nat. Commun.* 11.1 (Dec. 2020).

[11] L. D. Schlieben and H. Prokisch. "The Dimensions of Primary Mitochondrial Disorders". In: *Front. Cell Dev. Biol.* 8 (Nov. 2020).

[12] R. M. Andrews et al. "Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA [5]". In: *Nat. Genet.* 23.2 (1999), p. 147.

[13] R. E. Giles, H. Blanc, H. M. Cann, and C. Wallace aD. "Maternal inheritance of human mitochondrial DNA". In: *Proc. Natl. Acad. Sci. U. S. A.* 77.11 I (1980), pp. 6715–6719.

[14] M. Sato and K. Sato. "Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA". In: *Biochim. Biophys. Acta - Mol. Cell Res.* 1833.8 (Aug. 2013), pp. 1979–1984.

[15] P. Sutovsky. "Ubiquitin-dependent proteolysis in mammalian spermatogenesis, fertilization, and sperm quality control: Killing three birds with one stone". In: *Microsc. Res. Tech.* 61.1 (May 2003), pp. 88–102.

[16] S. Lutz-Bonengel and W. Parson. “No further evidence for paternal leakage of mitochondrial DNA in humans yet”. In: *Proc. Natl. Acad. Sci. U. S. A.* 116.6 (Feb. 2019), pp. 1821–1822.

[17] M. Schwartz and J. Vissing. “Paternal Inheritance of Mitochondrial DNA”. In: *N. Engl. J. Med.* 347.8 (Aug. 2002), pp. 576–580.

[18] S. Luo et al. “Biparental inheritance of mitochondrial DNA in humans”. In: *Proc. Natl. Acad. Sci. U. S. A.* 115.51 (Dec. 2018), pp. 13039–13044.

[19] W. Wei et al. “Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans”. In: *Nat. Commun.* 11.1 (Dec. 2020), pp. 1–11.

[20] C. W. Birky. “The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models”. In: *Annu. Rev. Genet.* 35.1 (Dec. 2001), pp. 125–148.

[21] S. K. Lehtinen et al. “Genotypic stability, segregation and selection in heteroplasmic human cell lines containing np 3243 mutant mtDNA”. In: *Genetics* 154.1 (2000), pp. 363–380.

[22] W. Wei et al. “Germline selection shapes human mitochondrial DNA diversity”. In: *Science (80-.).* 364.6442 (May 2019).

[23] R. LUFT, D. IKKOS, G. PALMIERI, L. ERNSTER, and B. AFZELIUS. “A case of severe hypermetabolism of nonthyroid origin with a defect in”. In: *J. Clin. Invest.* 41.9 (1962), pp. 1776–1804.

[24] M. Zeviani et al. “Deletions of mitochondrial DNA in Kearns-Sayre syndrome”. In: *Neurology* 38.9 (1988), pp. 1339–1346.

[25] I. J. Holt, A. E. Harding, and J. A. Morgan-Hughes. “Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies”. In: *Nature* 331.6158 (1988), pp. 717–719.

[26] D. C. Wallace et al. “Mitochondrial DNA mutation associated with Leber’s hereditary optic neuropathy”. In: *Science (80-.).* 242.4884 (1988), pp. 1427–1430.

[27] H. Endo et al. “Defective gene in lactic acidosis: Abnormal pyruvate dehydrogenase E1 α-subunit caused by a frame shift”. In: *Am. J. Hum. Genet.* 44.3 (1989), pp. 358–364.

[28] C. R. Ferreira, C. D. van Karnebeek, J. Vockley, and N. Blau. “A proposed nosology of inborn errors of metabolism”. In: *Genet. Med.* 21.1 (Jan. 2019), pp. 102–106.

[29] J. B. Stewart and P. F. Chinnery. “The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease”. In: *Nat. Rev. Genet.* 16.9 (Aug. 2015), pp. 530–542.

[30] P. Mishra and D. C. Chan. “Mitochondrial dynamics and inheritance during cell division, development and disease”. In: *Nat. Rev. Mol. Cell Biol.* 15.10 (Jan. 2014), pp. 634–646.

[31] M. Wagner et al. “Mitochondrial DNA mutation analysis from exome sequencing—A more holistic approach in diagnostics of suspected mitochondrial disease”. In: *J. Inherit. Metab. Dis.* 42.5 (Sept. 2019), pp. 909–917.

[32] V. Nesbitt et al. “The UK MRC Mitochondrial Disease Patient Cohort Study: Clinical phenotypes associated with the m.3243A>G mutation - Implications for diagnosis and management”. In: *J. Neurol. Neurosurg. Psychiatry* 84.8 (2013), pp. 936–938.

[33] J. M. Shoffner et al. “Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNALys mutation”. In: *Cell* 61.6 (June 1990), pp. 931–937.

[34] R. D. Ganetzky et al. “MT-ATP6 mitochondrial disease variants: Phenotypic and biochemical features analysis in 218 published cases and cohort of 14 new cases”. In: *Hum. Mutat.* 40.5 (May 2019), pp. 499–515.

[35] Y. S. Ng et al. “Pathogenic variants in MT-ATP6: A United Kingdom–based mitochondrial disease cohort study”. In: *Ann. Neurol.* 86.2 (Aug. 2019), pp. 310–315.

[36] C. Stendel et al. “Delineating MT-ATP6 -associated disease: From isolated neuropathy to early onset neurodegeneration”. In: *Neurol. Genet.* 6.1 (Feb. 2020), p. 393.

[37] S. L. Stenton and H. Prokisch. “Genetics of mitochondrial diseases: Identifying mutations to help diagnosis”. In: *EBioMedicine* 56 (June 2020), p. 102784.

[38] G. S. Gorman et al. “Mitochondrial diseases”. In: *Nat. Rev. Dis. Prim.* 2 (Oct. 2016).

[39] E. Barca et al. “Mitochondrial diseases in North America: An analysis of the NAMDC Registry”. In: *Neurol. Genet.* 6.2 (2020).

[40] G. S. Gorman et al. “Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease”. In: *Ann. Neurol.* 77.5 (May 2015), pp. 753–759.

[41] P. Y. Man et al. “The epidemiology of leber hereditary optic neuropathy in the North East of England”. In: *Am. J. Hum. Genet.* 72.2 (Feb. 2003), pp. 333–339.

[42] J. Tan et al. “Lifetime risk of autosomal recessive mitochondrial disorders calculated from genetic databases”. In: *EBioMedicine* 54 (Apr. 2020).

[43] P. Yu-Wai-Man et al. “A neurodegenerative perspective on mitochondrial optic neuropathies”. In: *Acta Neuropathol.* 132.6 (Dec. 2016), pp. 789–806.

[44] L. Jiang et al. “A Quantitative Proteome Map of the Human Body”. In: *Cell* 183.1 (Oct. 2020), 269–283.e19.

[45] N. Miyaue et al. “Repetitive brainstem lesions in mitochondrial DNA 11778G>A mutation of Leber hereditary optic neuropathy”. In: *eNeurologicalSci* 14 (Mar. 2019), pp. 74–76.

[46] G. Fruhman et al. “Atypical presentation of Leigh syndrome associated with a Leber hereditary optic neuropathy primary mitochondrial DNA mutation”. In: *Mol. Genet. Metab.* 103.2 (June 2011), pp. 153–160.

[47] S. Rahman et al. “Leigh syndrome: Clinical features and biochemical and DNA abnormalities”. In: *Ann. Neurol.* 39.3 (1996), pp. 343–351.

[48] J. Rahman, A. Noronha, I. Thiele, and S. Rahman. “Leigh map: A novel computational diagnostic resource for mitochondrial disease”. In: *Ann. Neurol.* 81.1 (Jan. 2017), pp. 9–16.

[49] A. Munnich et al. “Clinical presentations and laboratory investigations in respiratory chain deficiency”. In: *Eur. J. Pediatr.* 155.4 (1996), pp. 262–274.

[50] E. Morava et al. “Mitochondrial disease criteria: Diagnostic applications in children”. In: *Neurology* 67.10 (Nov. 2006), pp. 1823–1826.

[51] P. Witters et al. “Revisiting mitochondrial diagnostic criteria in the new era of genomics”. In: *Genet. Med.* 20.4 (Apr. 2018), pp. 444–451.

[52] C. Phoenix et al. “A scale to monitor progression and treatment of mitochondrial disease in children”. In: *Neuromuscul. Disord.* 16.12 (Dec. 2006), pp. 814–820.

[53] A. M. Schaefer et al. “Mitochondrial disease in adults: A scale to monitor progression and treatment”. In: *Neurology* 66.12 (June 2006), pp. 1932–1934.

[54] O. Shaham et al. “A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells”. In: *Proc. Natl. Acad. Sci. U. S. A.* 107.4 (Jan. 2010), pp. 1571–1575.

[55] A. Suomalainen et al. “FGF-21 as a biomarker for muscle-manifesting mitochondrial respiratory chain deficiencies: A diagnostic study”. In: *Lancet Neurol.* 10.9 (Sept. 2011), pp. 806–818.

[56] S. Yatsuga et al. “Growth differentiation factor 15 as a useful biomarker for mitochondrial disorders”. In: *Ann. Neurol.* 78.5 (Nov. 2015), pp. 814–823.

[57] S. B. Wortmann, J. A. Mayr, J. M. Nuoffer, H. Prokisch, and W. Sperl. “A Guideline for the Diagnosis of Pediatric Mitochondrial Disease: The Value of Muscle and Skin Biopsies in the Genetics Era”. In: *Neuropediatrics* 48.4 (Aug. 2017), pp. 309–314.

[58] R. J. Rodenburg. “Biochemical diagnosis of mitochondrial disorders”. In: *J. Inherit. Metab. Dis.* 34.2 (Apr. 2011), pp. 283–292.

[59] J. A. Mayr et al. “Spectrum of combined respiratory chain defects”. In: *J. Inherit. Metab. Dis.* 38.4 (July 2015), pp. 629–640.

[60] E. Ogawa et al. “Clinical validity of biochemical and molecular analysis in diagnosing Leigh syndrome: a study of 106 Japanese patients”. In: *J. Inherit. Metab. Dis.* 40.5 (Sept. 2017), pp. 685–693.

[61] R. J. Rodenburg et al. “A multi-center comparison of diagnostic methods for the biochemical evaluation of suspected mitochondrial disorders”. In: *Mitochondrion* 13.1 (Jan. 2013), pp. 36–43.

[62] G. S. Gorman, R. McFarland, J. Stewart, C. Feeney, and D. M. Turnbull. "Mitochondrial donation: from test tube to clinic". In: *Lancet* 392.10154 (Oct. 2018), pp. 1191–1192.

[63] A. Quadir, C. S. Pontifex, H. L. Robertson, C. Labos, and G. Pfeffer. "Systematic review and meta-analysis of cardiac involvement in mitochondrial myopathy". In: *Neurol. Genet.* 5.4 (Aug. 2019).

[64] D. Bick et al. "An online compendium of treatable genetic disorders". In: *Am. J. Med. Genet. Part C Semin. Med. Genet.* 187.1 (Mar. 2021), pp. 48–54.

[65] E. M. M. Hoytema van Konijnenburg et al. "Treatable inherited metabolic disorders causing intellectual disability: 2021 review and digital app". In: *Orphanet J. Rare Dis.* 16.1 (Dec. 2021), p. 170.

[66] T. Klopstock et al. "A randomized placebo-controlled trial of idebenone in Leber's hereditary optic neuropathy". In: *Brain* 134.9 (2011), pp. 2677–2686.

[67] F. Sanger, S. Nicklen, and A. R. Coulson. "DNA sequencing with chain-terminating inhibitors." In: *Proc. Natl. Acad. Sci. U. S. A.* 74.12 (1977), pp. 5463–5467.

[68] D. Altshuler, M. J. Daly, and E. S. Lander. "Genetic mapping in human disease". In: *Science (80-.).* 322.5903 (Nov. 2008), pp. 881–888.

[69] D. Botstein, R. L. White, M. Skolnick, and R. W. Davis. "Construction of a genetic linkage map in man using restriction fragment length polymorphisms". In: *Am. J. Hum. Genet.* 32.3 (1980), pp. 314–331.

[70] S. Anderson et al. "Sequence and organization of the human mitochondrial genome". In: *Nature* 290.5806 (1981), pp. 457–465.

[71] E. S. Lander et al. "Initial sequencing and analysis of the human genome". In: *Nature* 409.6822 (Feb. 2001), pp. 860–921.

[72] K. Neveling et al. "A Post-Hoc Comparison of the Utility of Sanger Sequencing and Exome Sequencing for the Diagnosis of Heterogeneous Diseases". In: *Hum. Mutat.* 34.12 (Dec. 2013), pp. 1721–1726.

[73] A. L. Frederiksen et al. "Tissue specific distribution of the 3243A→G mtDNA mutation". In: *J. Med. Genet.* 43.8 (Aug. 2006), pp. 671–677.

[74] E. W. Sommerville, P. F. Chinnery, G. S. Gorman, and R. W. Taylor. "Adult-onset Mendelian PEO Associated with Mitochondrial Disease." In: *J. Neuromuscul. Dis.* 1.2 (Jan. 2014), pp. 119–133.

[75] A. Broomfield et al. "Paediatric single mitochondrial DNA deletion disorders: an overlapping spectrum of disease". In: *J. Inherit. Metab. Dis.* 38.3 (May 2015), pp. 445–457.

[76] S. E. Calvo et al. "High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex i deficiency". In: *Nat. Genet.* 42.10 (Oct. 2010), pp. 851–858.

[77] S. E. Calvo et al. “Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing”. In: *Sci. Transl. Med.* 4.118 (Jan. 2012).

[78] V. Vasta, J. L. Merritt, R. P. Saneto, and S. H. Hahn. “Next-generation sequencing for mitochondrial diseases: A wide diagnostic spectrum”. In: *Pediatr. Int.* 54.5 (Oct. 2012), pp. 585–601.

[79] J. T. DaRe, V. Vasta, J. Penn, N. T. B. Tran, and S. H. Hahn. “Targeted exome sequencing for mitochondrial disorders reveals high genetic heterogeneity”. In: *BMC Med. Genet.* 14.1 (Nov. 2013), p. 118.

[80] D. S. Lieber et al. “Targeted exome sequencing of suspected mitochondrial disorders”. In: *Neurology* 80.19 (May 2013), pp. 1762–1770.

[81] A. Legati et al. “New genes and pathomechanisms in mitochondrial disorders unraveled by NGS technologies”. In: *Biochim. Biophys. Acta - Bioenerg.* 1857.8 (Aug. 2016), pp. 1326–1335.

[82] M. J. Bamshad et al. “Exome sequencing as a tool for Mendelian disease gene discovery”. In: *Nat. Rev. Genet.* 12.11 (Nov. 2011), pp. 745–755.

[83] A. M. Sulonen et al. “Comparison of solution-based exome capture methods for next generation sequencing”. In: *Genome Biol.* 12.10 (Sept. 2011).

[84] T. B. Haack et al. “Exome sequencing identifies ACAD9 mutations as a cause of complex i deficiency”. In: *Nat. Genet.* 42.12 (Dec. 2010), pp. 1131–1134.

[85] M. J. Bamshad, D. A. Nickerson, and J. X. Chong. “Mendelian Gene Discovery: Fast and Furious with No End in Sight”. In: *Am. J. Hum. Genet.* 105.3 (Sept. 2019), pp. 448–455.

[86] M. M. Clark et al. “Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases”. In: *npj Genomic Med.* 3.1 (Dec. 2018).

[87] T. B. Haack et al. “Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing”. In: *J. Med. Genet.* 49.4 (Apr. 2012), pp. 277–283.

[88] R. W. Taylor et al. “Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies”. In: *JAMA - J. Am. Med. Assoc.* 312.1 (July 2014), pp. 68–77.

[89] A. Ohtake et al. “Diagnosis and molecular basis of mitochondrial respiratory chain disorders: Exome sequencing for disease gene identification”. In: *Biochim. Biophys. Acta - Gen. Subj.* 1840.4 (2014), pp. 1355–1359.

[90] S. B. Wortmann, D. A. Koolen, J. A. Smeitink, L. van den Heuvel, and R. J. Rodenburg. “Whole exome sequencing of suspected mitochondrial patients in clinical practice”. In: *J. Inherit. Metab. Dis.* 38.3 (May 2015), pp. 437–443.

[91] M. Kohda et al. “A Comprehensive Genomic Analysis Reveals the Genetic Landscape of Mitochondrial Respiratory Chain Complex Deficiencies”. In: *PLOS Genet.* 12.1 (Jan. 2016). Ed. by G. S. Barsh, e1005679.

[92] E. Pronicka et al. “New perspective in diagnostics of mitochondrial disorders: Two years’ experience with whole-exome sequencing at a national paediatric centre”. In: *J. Transl. Med.* 14.1 (June 2016), p. 174.

[93] S. Puusepp et al. “Effectiveness of whole exome sequencing in unsolved patients with a clinical suspicion of a mitochondrial disorder in Estonia”. In: *Mol. Genet. Metab. Reports* 15 (June 2018), pp. 80–89.

[94] T. E. Theunissen et al. “Whole exome sequencing is the preferred strategy to identify the genetic defect in patients with a probable or possible mitochondrial cause”. In: *Front. Genet.* 9.OCT (Oct. 2018), p. 400.

[95] P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, and M. Kircher. “CADD: Predicting the deleteriousness of variants throughout the human genome”. In: *Nucleic Acids Res.* 47.D1 (Jan. 2019), pp. D886–D894.

[96] R. Vaser, S. Adusumalli, S. Ngak Leng, M. Sikic, and P. C. Ng. “SIFT missense predictions for genomes”. In: *Nat. Protoc.* 11 (2015).

[97] G. M. Cooper and J. Shendure. “Needles in stacks of needles: Finding disease-causal variants in a wealth of genomic data”. In: *Nat. Rev. Genet.* 12.9 (Sept. 2011), pp. 628–640.

[98] N. J. Lake, A. G. Compton, S. Rahman, and D. R. Thorburn. “Leigh syndrome: One disorder, more than 75 monogenic causes”. In: *Ann. Neurol.* 79.2 (Feb. 2016), pp. 190–203.

[99] T. B. Haack et al. “Phenotypic spectrum of eleven patients and five novel MTFMT mutations identified by exome sequencing and candidate gene screening”. In: *Mol. Genet. Metab.* 111.3 (Mar. 2014), pp. 342–352.

[100] K. J. Karczewski et al. “The mutational constraint spectrum quantified from variation in 141,456 humans”. In: *Nature* 581.7809 (May 2020), pp. 434–443.

[101] M. Willemsen et al. “Females with PDHA1 gene mutations: A diagnostic challenge”. In: *Mitochondrion* 6.3 (June 2006), pp. 155–159.

[102] K. Thompson et al. “Recurrent De Novo Dominant Mutations in SLC25A4 Cause Severe Early-Onset Mitochondrial Disease and Loss of Mitochondrial DNA Copy Number”. In: *Am. J. Hum. Genet.* 99.4 (Oct. 2016), pp. 860–876.

[103] V. Del Dotto et al. “SSBP1 mutations cause mtDNA depletion underlying a complex optic atrophy disorder”. In: *J. Clin. Invest.* 130.1 (Jan. 2020), pp. 108–125.

[104] J. F. McRae et al. “Prevalence and architecture of de novo mutations in developmental disorders”. In: *Nature* 542.7642 (Feb. 2017), pp. 433–438.

[105] S. C. Sallevelt et al. “De novo mtDNA point mutations are common and have a low recurrence risk”. In: *J. Med. Genet.* 54.2 (July 2016), pp. 114–124.

[106] S. Richards et al. “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”. In: *Genet. Med.* 17.5 (May 2015), pp. 405–424.

[107] Q. Li and K. Wang. “InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines”. In: *Am. J. Hum. Genet.* 100.2 (Feb. 2017), pp. 267–280.

[108] N. Sobreira, F. Schiettecatte, D. Valle, and A. Hamosh. “GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene”. In: *Hum. Mutat.* 36.10 (Oct. 2015), pp. 928–930.

[109] A. A. Philippakis et al. “The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery”. In: *Hum. Mutat.* 36.10 (Oct. 2015), pp. 915–921.

[110] A. M. Wenger, H. Guturu, J. A. Bernstein, and G. Bejerano. “Systematic reanalysis of clinical exome data yields additional diagnoses: Implications for providers”. In: *Genet. Med.* 19.2 (Feb. 2017), pp. 209–214.

[111] L. J. Ewans et al. “Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders”. In: *Genet. Med.* 20.12 (Dec. 2018), pp. 1564–1574.

[112] G. Costain et al. “Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing”. In: *Eur. J. Hum. Genet.* 26.5 (May 2018), pp. 740–744.

[113] V. Shashi et al. “A comprehensive iterative approach is highly effective in diagnosing individuals who are exome negative”. In: *Genet. Med.* 21.1 (Jan. 2019), pp. 161–172.

[114] J. Li et al. “Reanalysis of whole exome sequencing data in patients with epilepsy and intellectual disability/mental retardation”. In: *Gene* 700 (June 2019), pp. 168–175.

[115] P. Liu et al. “Reanalysis of Clinical Exome Sequencing Data”. In: *N. Engl. J. Med.* 380.25 (June 2019), pp. 2478–2480.

[116] S. Nambot et al. “Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: Substantial interest of prospective annual reanalysis”. In: *Genet. Med.* 20.6 (June 2018), pp. 645–654.

[117] L. S. Kremer and H. Prokisch. “Identification of disease-causing mutations by functional complementation of patient-derived fibroblast cell lines”. In: *Methods Mol. Biol.* Vol. 1567. Humana Press Inc., 2017, pp. 391–406.

[118] R. Kopajtich et al. “Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders”. In: *medRxiv* (2021).

[119] V. A. Yépez et al. “OCR-Stats: Robust estimation and statistical testing of mitochondrial respiration activities using seahorse xf analyzer”. In: *PLoS One* 13.7 (July 2018).

[120] R. J. Wanders et al. “Translational Metabolism: A multidisciplinary approach towards precision diagnosis of inborn errors of metabolism in the omics era”. In: *J. Inherit. Metab. Dis.* 42.2 (Mar. 2019), pp. 197–208.

[121] C. L. Alston et al. “Biallelic Mutations in TMEM126B Cause Severe Complex I Deficiency with a Variable Clinical Phenotype”. In: *Am. J. Hum. Genet.* 99.1 (July 2016), pp. 217–227.

[122] L. S. Kremer et al. “Genetic diagnosis of Mendelian disorders via RNA sequencing”. In: *Nat. Commun.* 8.1 (June 2017), pp. 1–11.

[123] S. L. Stenton et al. “Impaired complex I repair causes recessive Leber’s hereditary optic neuropathy”. In: *J. Clin. Invest.* (Jan. 2021).

[124] A. Paul et al. “FDXR Mutations Cause Sensorial Neuropathies and Expand the Spectrum of Mitochondrial Fe-S-Synthesis Diseases”. In: *Am. J. Hum. Genet.* 101.4 (Oct. 2017), pp. 630–637.

[125] S. L. Stenton et al. “Expanding the clinical and genetic spectrum of FDXR deficiency by functional validation of variants of uncertain significance”. In: *Hum. Mutat.* (2020).

[126] M. Ma et al. “Disease-associated variants in different categories of disease located in distinct regulatory elements”. In: *BMC Genomics* 16.8 (June 2015).

[127] L. G. Riley et al. “The diagnostic utility of genome sequencing in a pediatric cohort with suspected mitochondrial disease”. In: *Genet. Med.* 22.7 (July 2020), pp. 1254–1261.

[128] S. L. Stenton, L. S. Kremer, R. Kopajtich, C. Ludwig, and H. Prokisch. “The diagnosis of inborn errors of metabolism by an integrative “multi-omics” approach: A perspective encompassing genomics, transcriptomics, and proteomics”. In: *J. Inherit. Metab. Dis.* 43.1 (Jan. 2020), pp. 25–35.

[129] B. B. Cummings et al. “Improving genetic diagnosis in Mendelian disease with transcriptome sequencing”. In: *Sci. Transl. Med.* 9.386 (Apr. 2017).

[130] L. Frésard et al. “Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts”. In: *Nat. Med.* 25.6 (June 2019), pp. 911–919.

[131] H. D. Gonorazky et al. “Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease”. In: *Am. J. Hum. Genet.* 104.3 (Mar. 2019), pp. 466–483.

[132] V. A. Yépez et al. “Detection of aberrant gene expression events in RNA sequencing data”. In: *Nat. Protoc.* 16.2 (Feb. 2021), pp. 1276–1296.

[133] C. Mertes et al. “Detection of aberrant splicing events in RNA-seq data using FRASER”. In: *Nat. Commun.* 12.1 (Dec. 2021).

[134] F. Aguet et al. “The GTEx Consortium atlas of genetic regulatory effects across human tissues”. In: *Science (80-.).* 369.6509 (Sept. 2020), pp. 1318–1330.

[135] I. Papatheodorou et al. “Expression Atlas update: From tissues to single cells”. In: *Nucleic Acids Res.* 48.D1 (Jan. 2020), pp. D77–D83.

[136] J. K. Aicher, P. Jewell, J. Vaquero-Garcia, Y. Barash, and E. J. Bhoj. “Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq”. In: *Genet. Med.* 22.7 (July 2020), pp. 1181–1190.

[137] F. Brechtmann et al. “OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data”. In: *Am. J. Hum. Genet.* 103.6 (Dec. 2018), pp. 907–917.

[138] C. L. Alston, S. L. Stenton, G. Hudson, H. Prokisch, and R. W. Taylor. “The genetics of mitochondrial disease: dissecting mitochondrial pathology using multi-omic pipelines”. In: *J. Pathol.* (Feb. 2021), path.5641.

[139] N. Sahni et al. “Widespread macromolecular interaction perturbations in human genetic disorders”. In: *Cell* 161.3 (Apr. 2015), pp. 647–660.

[140] N. J. Lake et al. “Biallelic Mutations in MRPS34 Lead to Instability of the Small Mitoribosomal Subunit and Leigh Syndrome”. In: *Am. J. Hum. Genet.* 101.2 (Aug. 2017), pp. 239–254.

[141] N. N. Borna et al. “Mitochondrial ribosomal protein PTCD3 mutations cause oxidative phosphorylation defects with Leigh syndrome”. In: *Neurogenetics* 20.1 (Mar. 2019), pp. 9–25.

[142] T. D. Jackson et al. “The TIM22 complex regulates mitochondrial one-carbon metabolism by mediating the import of Sideroflexins”. In: *bioRxiv* (Feb. 2020), p. 2020.02.06.937920.

[143] V. Narasimhan et al. “BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data”. In: *Bioinformatics* 32.11 (June 2016), pp. 1749–1751.

[144] C. W. Rehder et al. “American College of medical genetics and genomics: Standards and guidelines for documenting suspected consanguinity as an incidental finding of genomic testing”. In: *Genet. Med.* 15.2 (Feb. 2013), pp. 150–152.

[145] P. Kumar, S. Henikoff, and P. C. Ng. “Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm”. In: *Nat. Protoc.* 4.7 (June 2009), pp. 1073–1082.

[146] I. A. Adzhubei et al. “A method and server for predicting damaging missense mutations”. In: *Nat. Methods* 7.4 (Apr. 2010), pp. 248–249.

[147] A. Dobin et al. “STAR: Ultrafast universal RNA-seq aligner”. In: *Bioinformatics* 29.1 (Jan. 2013), pp. 15–21.

[148] S. Anders, P. T. Pyl, and W. Huber. “HTSeq–a Python framework to work with high-throughput sequencing data”. In: *Bioinformatics* 31.2 (Jan. 2015), pp. 166–169.

[149] M. I. Love, W. Huber, and S. Anders. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biol.* 15.12 (Dec. 2014), p. 550.

[150] G. Yu, L. G. Wang, Y. Han, and Q. Y. He. “ClusterProfiler: An R package for comparing biological themes among gene clusters”. In: *Omi. A J. Integr. Biol.* 16.5 (May 2012), pp. 284–287.

[151] Y. Benjamini and Y. Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. In: *J. R. Stat. Soc. Ser. B* 57.1 (Jan. 1995), pp. 289–300.

[152] J. Zecha et al. “TMT labeling for the masses: A robust and cost-efficient, in-solution labeling approach”. In: *Mol. Cell. Proteomics* 18.7 (2019), pp. 1468–1478.

[153] P. Yu et al. “Trimodal Mixed Mode Chromatography That Enables Efficient Offline Two-Dimensional Peptide Fractionation for Proteome Analysis”. In: *Anal. Chem.* 89.17 (Sept. 2017), pp. 8884–8891.

[154] S. Tyanova, T. Temu, and J. Cox. “The MaxQuant computational platform for mass spectrometry-based shotgun proteomics”. In: *Nat. Protoc.* 11.12 (Dec. 2016), pp. 2301–2319.

[155] M. E. Ritchie et al. “Limma powers differential expression analyses for RNA-sequencing and microarray studies”. In: *Nucleic Acids Res.* 43.7 (Jan. 2015), e47.

[156] I. Wittig, H. P. Braun, and H. Schägger. “Blue native PAGE”. In: *Nat. Protoc.* 1.1 (June 2006), pp. 418–428.

[157] H. Heide and I. Wittig. “Methods to analyse composition and dynamics of macromolecular complexes”. In: *Biochem. Soc. Trans.* 41.5 (Oct. 2013), pp. 1235–1241.

[158] A. N. A. Agip et al. “Cryo-em structures of complex i from mouse heart mitochondria in two biochemically defined states”. In: *Nat. Struct. Mol. Biol.* 25.7 (June 2018).

[159] M. Giurgiu et al. “CORUM: The comprehensive resource of mammalian protein complexes - 2019”. In: *Nucleic Acids Res.* 47.D1 (Jan. 2019), pp. D559–D563.

[160] H. Liu and J. H. Naismith. “An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol.” In: *BMC Biotechnol.* 8.1 (Dec. 2008), p. 91.

[161] F. Mölder et al. “Sustainable data analysis with Snakemake”. In: *F1000Research* 10 (Jan. 2021), p. 33.

[162] L. Wachutka. *Leonhard Wachutka / wBuild · GitLab*. 2019.

[163] D. Greene, S. Richardson, and E. Turro. "OntologyX: A suite of R packages for working with ontological data". In: *Bioinformatics* 33.7 (Apr. 2017), pp. 1104–1106.

[164] P. Godard and M. Page. "PCAN: Phenotype consensus analysis to support disease-gene association". In: *BMC Bioinformatics* 17.1 (Dec. 2016), p. 518.

[165] J. L. Maag. "Gganatogram: An R package for modular visualisation of anatograms and tissues based on ggplot2 [version 2; peer review: 2 approved]". In: *F1000Research* 7 (2018).

[166] Y. Peng et al. "Biallelic mutations in the ferredoxin reductase gene cause novel mitochondriopathy with optic atrophy". In: *Hum. Mol. Genet.* 26.24 (Dec. 2017), pp. 4937–4950.

[167] J. Slone et al. "Biallelic mutations in FDXR cause neurodegeneration associated with inflammation". In: *J. Hum. Genet.* 63.12 (Dec. 2018), pp. 1211–1222.

[168] L. K. Beilschmidt and H. M. Puccio. "Mammalian Fe-S cluster biogenesis and its implication in disease". In: *Biochimie* 100.1 (2014), pp. 48–60.

[169] O. Stehling, C. Wilbrecht, and R. Lill. "Mitochondrial iron-sulfur protein biogenesis and human disease". In: *Biochimie* 100.1 (2014), pp. 61–77.

[170] M. Cnop, H. Mulder, and M. Igoillo-Esteve. "Diabetes in Friedreich ataxia". In: *J. Neurochem.* 126.SUPPL.1 (Aug. 2013), pp. 94–102.

[171] A. Filla, G. DeMichele, G. Caruso, R. Marconi, and G. Campanella. "Genetic data and natural history of Friedreich's disease: a study of 80 Italian patients". In: *J. Neurol.* 237.6 (Oct. 1990), pp. 345–351.

[172] A. T. Tebbenkamp et al. "The 7q11.23 Protein DNAJC30 Interacts with ATP Synthase and Links Mitochondria to Brain Development". In: *Cell* 175.4 (Nov. 2018), 1088–1104.e23.

[173] L. C. Gandolfo, M. Bahlo, and T. P. Speed. "Dating rare mutations from small samples with dense marker data". In: *Genetics* 197.4 (2014), pp. 1315–1327.

[174] G. H. Hardy. "Mendelian proportions in a mixed population". In: *Science (80-.).* 28.706 (1908), pp. 49–50.

[175] G. Hudson et al. "Clinical expression of leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background". In: *Am. J. Hum. Genet.* 81.2 (2007), pp. 228–233.

[176] M. Pala et al. "Mitochondrial DNA signals of late glacial recolonization of europe from near eastern refugia". In: *Am. J. Hum. Genet.* 90.5 (May 2012), pp. 915–924.

[177] V. Carelli, F. N. Ross-Cisneros, and A. A. Sadun. "Mitochondrial dysfunction as a cause of optic neuropathies". In: *Prog. Retin. Eye Res.* 23.1 (2004), pp. 53–89.

[178] T. Rosenberg et al. "Prevalence and genetics of leber hereditary optic neuropathy in the Danish population". In: *Investig. Ophthalmol. Vis. Sci.* 57.3 (Mar. 2016), pp. 1370–1375.

[179] H. H. Kampinga and E. A. Craig. “The HSP70 chaperone machinery: J proteins as drivers of functional specificity”. In: *Nat. Rev. Mol. Cell Biol.* 11.8 (Aug. 2010), pp. 579–592.

[180] E. L. Huttlin et al. “The BioPlex Network: A Systematic Exploration of the Human Interactome”. In: *Cell* 162.2 (2015), pp. 425–440.

[181] E. L. Huttlin et al. “Architecture of the human interactome defines protein communities and disease networks”. In: *Nature* 545.7655 (2017), pp. 505–509.

[182] S. Guerrero-Castillo et al. “The Assembly Pathway of Mitochondrial Respiratory Chain Complex I”. In: *Cell Metab.* 25.1 (Jan. 2017), pp. 128–139.

[183] Y. Cheng and F. Perocchi. “ProtPhylo: Identification of protein-phenotype and protein-protein functional associations via phylogenetic profiling”. In: *Nucleic Acids Res.* 43.W1 (2015), W160–W168.

[184] D. F. Bogenhagen and J. D. Haley. “Pulse-chase SILAC- based analyses reveal selective oversynthesis and rapid turnover of mitochondrial protein components of respiratory complexes”. In: *J. Biol. Chem.* 295.9 (Feb. 2020), pp. 2544–2554.

[185] D. G. MacArthur et al. “Guidelines for investigating causality of sequence variants in human disease”. In: *Nature* 508.7497 (Apr. 2014), pp. 469–476.

[186] M. J. Guillen Sacoto et al. “De Novo Variants in the ATPase Module of MORC2 Cause a Neurodevelopmental Disorder with Growth Retardation and Variable Craniofacial Dysmorphism”. In: *Am. J. Hum. Genet.* 107.2 (Aug. 2020), pp. 352–363.

[187] W. Lissens et al. “Mutations in the X-linked pyruvate dehydrogenase (E1) α subunit gene (PDHA1) in patients with a pyruvate dehydrogenase complex deficiency”. In: *Hum. Mutat.* 15.3 (2000), pp. 209–219.

[188] L. Marti-Sanchez et al. “Delineating the neurological phenotype in children with defects in the ECHS1 or HIBCH gene”. In: *J. Inherit. Metab. Dis.* 44.2 (2020).

[189] B. M. Repp et al. “Clinical, biochemical and genetic spectrum of 70 patients with ACAD9 deficiency: is riboflavin supplementation effective?” In: *Orphanet J. Rare Dis.* 13.1 (Dec. 2018), p. 120.

[190] T. B. Haack et al. “Deficiency of ECHS1 causes mitochondrial encephalopathy with cardiac involvement”. In: *Ann. Clin. Transl. Neurol.* 2.5 (May 2015), pp. 492–509.

[191] M. Tetreault et al. “Whole-exome sequencing identifies novel ECHS1 mutations in Leigh syndrome”. In: *Hum. Genet.* 134.9 (Sept. 2015), pp. 981–991.

[192] S. Rahman and W. C. Copeland. “POLG-related disorders and their neurological manifestations”. In: *Nat. Rev. Neurol.* 15.1 (Jan. 2019), pp. 40–52.

[193] B. P. Coe et al. “Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity”. In: *Nat. Genet.* 51.1 (Jan. 2019), pp. 106–116.

[194] J. E. Posey et al. “Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation”. In: *N. Engl. J. Med.* 376.1 (Jan. 2017), pp. 21–31.

[195] D. Trujillano et al. “Clinical exome sequencing: Results from 2819 samples reflecting 1000 families”. In: *Eur. J. Hum. Genet.* 25.2 (Feb. 2017), pp. 176–182.

[196] C. E. French et al. “Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children”. In: *Intensive Care Med.* 45.5 (May 2019), pp. 627–636.

[197] D. M. Niyazov, S. G. Kahler, and R. E. Frye. “Primary Mitochondrial Disease and Secondary Mitochondrial Dysfunction: Importance of Distinction for Diagnosis and Treatment”. In: *Mol. Syndromol.* 7.3 (July 2016), pp. 122–137.

[198] M. P. Coelho et al. “Iron-sulfur cluster ISD11 deficiency (LYRM4 gene) presenting as cardiorespiratory arrest and 3-methylglutaconic aciduria”. In: *JIMD Rep.* 49.1 (June 2019), pp. 11–16.

[199] A. Rotig et al. “Aconitase and mitochondrial iron-sulphur protein deficiency in friedreich ataxia”. In: *Nat. Genet.* 17.2 (1997), pp. 215–217.

[200] D. E. Bassett, M. S. Boguski, and P. Hieter. “Yeast genes and human disease”. In: *Nature* 379.6566 (1996), pp. 589–590.

[201] G. R. Stuart, J. H. Santos, M. K. Strand, B. Van Houten, and W. C. Copeland. “Mitochondrial and nuclear DNA defects in Saccharomyces cerevisiae with mutations in DNA polymerase γ associated with progressive external ophthalmoplegia”. In: *Hum. Mol. Genet.* 15.2 (Jan. 2006), pp. 363–374.

[202] D. Ghezzi et al. “SDHAF1, encoding a LYR complex-II specific assembly factor, is mutated in SDH-defective infantile leukoencephalopathy”. In: *Nat. Genet.* 41.6 (June 2009), pp. 654–656.

[203] E. Baruffini et al. “MTO1 mutations are associated with hypertrophic cardiomyopathy and lactic acidosis and cause respiratory chain deficiency in humans and yeast”. In: *Hum. Mutat.* 34.11 (Nov. 2013), pp. 1501–1509.

[204] T. Rinaldi, C. Dallabona, I. Ferrero, L. Frontali, and M. Bolotin-Fukuhara. “Mitochondrial diseases and the role of the yeast models”. In: *FEMS Yeast Res.* 10.8 (Dec. 2010), pp. 1006–1022.

[205] J. P. Lasserre et al. “Yeast as a system for modeling mitochondrial disease mechanisms and discovering therapies”. In: *DMM Dis. Model. Mech.* 8.6 (June 2015), pp. 509–522.

[206] K. Dimitriadis et al. “Leber’s hereditary optic neuropathy with late disease onset: clinical and molecular characteristics of 20 patients”. In: *Orphanet J. Rare Dis.* 9.1 (Dec. 2014), p. 158.

[207] P. Yu-Wai-Man, P. G. Griffiths, and P. F. Chinnery. “Mitochondrial optic neuropathies - Disease mechanisms and therapeutic strategies”. In: *Prog. Retin. Eye Res.* 30.2 (Mar. 2011), pp. 81–114.

[208] S. Gerber et al. “Compound heterozygosity for severe and hypomorphic NDUFS2 mutations cause non-syndromic LHON-like optic neuropathy”. In: *J. Med. Genet.* 54.5 (May 2017), pp. 346–356.

[209] S. Dröse and U. Brandt. “Molecular mechanisms of superoxide production by the mitochondrial respiratory chain”. In: *Adv. Exp. Med. Biol.* 748 (2012), pp. 145–169.

[210] J. Hirst and M. M. Roessler. “Energy conversion, redox catalysis and generation of reactive oxygen species by respiratory complex i”. In: *Biochim. Biophys. Acta - Bioenerg.* 1857.7 (July 2016), pp. 872–883.

[211] F. Distelmaier et al. “Mitochondrial complex i deficiency: From organelle dysfunction to clinical disease”. In: *Brain* 132.4 (2009), pp. 833–842.

[212] R. J. Janssen, L. G. Nijtmans, L. P. van den Heuvel, and J. A. Smeitink. “Mitochondrial complex I: Structure, function and pathology”. In: *J. Inherit. Metab. Dis.* 29.4 (Aug. 2006), pp. 499–515.

[213] K. Fiedorczuk and L. A. Sazanov. “Mammalian Mitochondrial Complex I Structure and Disease-Causing Mutations”. In: *Trends Cell Biol.* 28.10 (Oct. 2018), pp. 835–867.

[214] F. A. Urra, F. Muñoz, A. Lovy, and C. Cárdenas. “The mitochondrial Complex(I)ty of cancer”. In: *Front. Oncol.* 7.JUN (June 2017), p. 118.

[215] R. Stefanatos and A. Sanz. “Mitochondrial complex I: A central regulator of the aging process”. In: *Cell Cycle* 10.10 (May 2011), pp. 1528–1532.

[216] P. Grabowski et al. “Proteome analysis of human neutrophil granulocytes from patients with monogenic disease using data-independent acquisition”. In: *Mol. Cell. Proteomics* 18.4 (2019), pp. 760–772.

[217] E. L. van Dijk, Y. Jaszczyszyn, D. Naquin, and C. Thermes. “The Third Revolution in Sequencing Technology”. In: *Trends Genet.* 34.9 (Sept. 2018), pp. 666–681.

[218] J. D. Merker et al. “Long-read genome sequencing identifies causal structural variation in a Mendelian disease”. In: *Genet. Med.* 20.1 (Jan. 2018), pp. 159–163.

[219] H. Miao et al. “Long-read sequencing identified a causal structural variant in an exome-negative case and enabled preimplantation genetic diagnosis”. In: *Hereditas* 155 (2018), p. 32.

[220] M. Cretu Stancu et al. “Mapping and phasing of structural variation in patient genomes using nanopore sequencing”. In: *Nat. Commun.* 8.1 (Dec. 2017).

[221] U. R. Dutta et al. “Breakpoint mapping of a novel de novo translocation t(X;20)(q11.1;p13) by positional cloning and long read sequencing”. In: *Genomics* 111.5 (Sept. 2019), pp. 1108–1114.

[222] S. Ardui et al. “Detecting AGG Interruptions in Females With a FMR1 Premutation by Long-Read Single-Molecule Sequencing: A 1 Year Clinical Experience”. In: *Front. Genet.* 9.MAY (May 2018), p. 150.

[223] H. Ishiura et al. “Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy”. In: *Nat. Genet.* 50.4 (Apr. 2018), pp. 581–590.

[224] S. Zeng et al. “Long-read sequencing identified intronic repeat expansions in SAMD12 from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy”. In: *J. Med. Genet.* 56.4 (Apr. 2019), pp. 265–270.

[225] M. Wilbe et al. “A novel approach using long-read sequencing and ddPCR to investigate gonadal mosaicism and estimate recurrence risk in two families with developmental disorders”. In: *Prenat. Diagn.* 37.11 (Nov. 2017), pp. 1146–1154.

[226] G. Frans et al. “Conventional and Single-Molecule Targeted Sequencing Method for Specific Variant Detection in IKBKG while Bypassing the IKBKGP1 Pseudogene”. In: *J. Mol. Diagnostics* 20.2 (Mar. 2018), pp. 195–202.

[227] J. Yu et al. “PRICKLE3 linked to ATPase biogenesis manifested Leber’s hereditary optic neuropathy”. In: *J. Clin. Invest.* 130.9 (Sept. 2020), pp. 4935–4946.

[228] P. Jiang et al. “The exome sequencing identified the mutation in YARS2 encoding the mitochondrial tyrosyl-tRNA synthetase as a nuclear modifier for the phenotypic manifestation of Leber’s hereditary optic neuropathy-associated mitochondrial DNA mutation”. In: *Hum. Mol. Genet.* 25.3 (Feb. 2016), pp. 584–596.

[229] T. L. To et al. “A Compendium of Genetic Modifiers of Mitochondrial Dysfunction Reveals Intra-organelle Buffering”. In: *Cell* 179.5 (Nov. 2019), 1222–1238.e17.

[230] A. M. Gazzo et al. “DIDA: A curated and annotated digenic diseases database”. In: *Nucleic Acids Res.* 44.D1 (Jan. 2016), pp. D900–D907.

[231] P. Da Pozzo et al. “Sporadic PEO caused by a novel POLG variation and a Twinkle mutation: digenic inheritance?” In: *Neurol. Sci.* 36.9 (Sept. 2015), pp. 1713–1715.

[232] G. Van Goethem et al. “Digenic progressive external ophthalmoplegia in a sporadic patient: Recessive mutations in POLG and C10orf2/twinkle [2]”. In: *Hum. Mutat.* 22.2 (2003), pp. 175–176.

[233] G. Kerner et al. “A genome-wide case-only test for the detection of digenic inheritance in human exomes”. In: *Proc. Natl. Acad. Sci. U. S. A.* 117.32 (Aug. 2020), pp. 19367–19375.

[234] M. T. Oetjens, M. A. Kelly, A. C. Sturm, C. L. Martin, and D. H. Ledbetter. “Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders”. In: *Nat. Commun.* 10.1 (Dec. 2019), pp. 1–10.