Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Lehrstuhl für Entwicklungsgenetik

Discovery and validation of coding and non-coding pathogenic variants in mitochondrial disorders

Laura Sophie Kremer

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzende/-r:	Prof. Dr. Martin Hrabě de Angelis
Prüfende/-r der Dissertation:	1. Priv.-Doz. Dr. Thomas Floss
	2. Prof. Dr. Bernhard Küster
	3. Priv.-Doz. Dr. Johannes Mayr

Die Dissertation wurde am 28.06.2017 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 02.11.2017 angenommen.

For Tim.

Summary

Mitochondrial disorders are a heterogeneous group of genetic disorders caused by defects in mitochondrial oxidative phosphorylation. With an estimated prevalence of 1 in 5000 live births they depict one of the largest groups of inborn errors of metabolism. Their clinical presentation is characterized by a vast variation ranging from single organ involvement to multi-organ symptoms. This is accompanied by an extremely broad range of distinct molecular disorders with more than 250 reported disease-associated genes so far. A genotype-phenotype correlation can be observed only in a limited number of cases. Molecular diagnosis therefore requires a comprehensive diagnostic approach represented by whole exome sequencing (WES). WES is the massively parallel sequencing of all exonic, hence protein coding regions of the genome containing about 85% of the known Mendelian disease-causing variants.

During the first years of my PhD studies, I contributed to the identification and validation of causative variants in 10 novel and 5 known disease-associated genes in suspected mitochondrial disease patients using WES. In the first part of this thesis, I describe in detail 2 of these contributions.

In the first study, WES revealed homozygous predicted loss-of-function variants in two genes encoding mitochondrial proteins, *MTO1* and *LYRM7*. At the point of study, *MTO1* was an established and *LYRM7* a suggested mitochondrial disease-associated gene. To evaluate the pathogenic relevance of the variants and to examine whether the patient possibly suffers from two distinct or overlapping diseases, functional studies were employed. Western blot analysis and respiratory chain complex activity measurements upon overexpression of wild-type *MTO1* and *LYRM7* cDNA in patient derived fibroblast cell lines clearly demonstrated the pathogenicity of the *LYRM7* variant. While the experiments indicated that the variant in *MTO1* most likely is benign, a negative impact cannot be excluded.

In the second study, a joint analysis of several patients presenting with an overlapping clinical phenotype established disease-causing variants in *TANGO2*, a gene previously not associated with any disease. A postulated role of TANGO2 in the distribution of the Golgi membrane was not confirmed in my studies as immunostaining of the Golgi did not reveal a difference in organization of the Golgi between patient and control cells. However, by accessing the

cellular oxygen consumption I found normal activity of respiratory chain complexes but impaired ß-oxidation in the patient derived fibroblasts which supported the metabolic suspicion of a defect in ß-oxidation. The diagnosis of ß-oxidation defects, depicting non-classical, secondary mitochondrial disorders, has strong implications for possible therapeutic options.

In the second part of my PhD studies, I focused on cases where WES was inconclusive. Indeed, about half of the patients with suspected mitochondrial disorders remain undiagnosed after WES. We assumed that incomplete capture of variants, especially non-coding variants, and failure to prioritize variants contributes to such inconclusive WES cases. Whereas the former can be overcome by whole genome sequencing (WGS), the vast number of variants generated by WGS and the poor understanding of the non-coding genome further obscure the prioritization of the causative variant. RNA sequencing (RNA-seq), in turn, might ease the prioritization of variants by unravelling their effects on RNA abundance and sequence. I therefore selected 105 patient derived fibroblast cell lines from solved and unsolved cases and performed RNA-seq using a standardized protocol. By manual inspection of the RNA-seq data from a patient with respiratory chain complex I deficiency I detected aberrant expression of the respiratory chain complex I (RCCI) assembly factor *TIMMDC1*, a gene previously not annotated with disease risk. I subsequently identified a deep intronic variant likely involved in the activation of a cryptic intronic splice site resulting in aberrant splicing and finally nonsense-mediated decay (NMD). Using functional validation assays, I showed pathogenic relevance of TIMMDC1 deficiency and identified altogether 3 families carrying the same homozygous pathogenic variant. These findings guided a systematic analysis performed in close collaboration with computational biology department of the Technische Universität München. The systematic analysis detected a median of one aberrantly expressed gene, five aberrant splicing events, and six mono-allelically expressed rare variants per sample. This small number of events allowed manual inspection and validation providing a diagnosis for 10% (5 of 48) of previously unsolved cases. Importantly, our approach enabled the identification of causative non-coding variants in genes not previously associated with any disease as well as identification of causative variants in genes implicated in diseases going beyond mitochondrial disorders.

Zusammenfassung

Mitochondriale Erkrankungen sind eine heterogene Gruppe von Erbkrankheiten, die durch Defekte in der mitochondrialen oxidativen Phosphorylierung verursacht werden. Mit einer geschätzten Prävalenz von 1 in 5000 Lebendgeburten stellen sie eine der größten Gruppen von Stoffwechselerkrankungen dar. Ihre klinische Präsentation wird von einer großen Variation charakterisiert, welche von der Betroffenheit einzelner Organe bis hin zu der Beeinträchtigung multipler Organe reichen kann. Dies wird von einer extrem großen Bandbreite an genetischen Ursachen begleitet, bisher werden mehr als 250 publizierte Gene mit der Erkrankung in Verbindung gebracht. Eine klare Korrelation zwischen dem Genotyp und dem Phänotyp liegt jedoch nur in den wenigstens Fällen vor. Für die molekulargenetische Diagnose ist daher eine uneingeschränkte diagnostische Methode wie z.B. die Sequenzierung des gesamten Exoms („whole exome sequening“, WES) notwendig. WES ist die massive parallele Sequenzierung aller exonischen, also proteinkodierenden Bereiche des Genoms, die laut Vorhersagen 85% aller Varianten enthalten die für bekannte monogenetische Erkrankungen ursächlich sind.

Während der ersten Jahre meiner Promotion habe ich zur Identifizierung und Validierung von kausalen Varianten in 10 bis dahin nicht krankheitsassoziierten Genen, als auch in 5 krankheitsassoziiert Genen mittels WES beigetragen. Im ersten Teil dieser Doktorarbeit beschreibe ich 2 dieser Beiträge im Detail.

In der ersten Studie führte die Analyse mittels WES zu der Identifizierung von homozygoten Varianten in zwei Genen die mitochondriale Proteine kodieren, *MTO1* und *LYRM7*, die laut Vorhersagen zum Funktionsverlust des kodierten Proteins führen. Zum Zeitpunkt der Studie waren Mutationen in *MTO1* bereits mehrere Male als krankheitsverursachend beschrieben, während Mutationen in *LYRM7* in nur einem Fall mit einer mitochondrialen Erkrankung assoziiert wurden. Um die pathogene Relevanz der Varianten zu klären und zu untersuchen, ob der Patient möglicherweise von 2 separaten oder überlappenden Erkrankungen betroffen war, verwendeten wir funktionelle Tests. Überexpression von naiver cDNA von *MTO1* und *LYRM7* in Fibroblasten des Patienten und anschließende Western Blot-Analyse und Messungen der Aktivität der Atmungskettenkomplexe ergaben eine klare Pathogenität der *LYRM7* Variante. Die Experimente wiesen darauf hin, dass die *MTO1* Variante

wahrscheinlich gutartig ist. Ein negativer Effekt kann jedoch nicht komplett ausgeschlossen werden.

In der zweiten Studie wurden durch eine gemeinsame Analyse von mehreren Patienten mit ähnlichem Phänotyp krankheitsverursachende Varianten in *TANGO2* identifiziert, einem bis dahin nicht krankheitsassoziierten Gen. Eine postulierte Rolle von TANGO2 in der Verteilung der Golgi-Membran konnte durch meine Untersuchungen nicht bestätigt werden, da Immunfärbung des Golgis in Patientenzelllinien und Kontrollzelllinien keinen Unterschied in der Organisation des Golgis ergab. Mittels Analyse des zellulären Sauerstoffverbrauches konnte ich normale Aktivität der Atmungskettenkomplexe zeigen, jedoch war in den Patientenzelllinien eine verminderte ß-Oxidation nachweisbar was den metabolischen Verdacht auf einen Defekt der ß-Oxidation bestärkte. Die Diagnose eines ß-Oxidationsdefekts, welcher eine nicht-klassische, sekundäre Mitochondriopathie darstellt, hat starke Auswirkungen auf mögliche therapeutische Optionen.

Der zweite Teil dieser Arbeit zielte jedoch auf Fälle ab, bei denen die WES-Analyse unschlüssig war. Tatsächlich bleibt etwa die Hälfte aller Patienten mit Verdacht auf eine Mitochondriopathie nach WES ohne genetische Diagnose. Wir vermuteten, dass unschlüssige WES-Analyse auf die unvollständige Detektion von Varianten, speziell in nicht-kodierenden Bereichen, und unzureichende Priorisierung von Varianten zurückzuführen ist. Während Ersteres durch die Sequenzierung des gesamten Genoms („whole genome sequencing“, WGS) verbessert werden kann, wird die Priorisierung von Varianten dadurch allerdings auf Grunde der hohen Anzahl der identifizierten Varianten und dem unvollständigen Verständnisses über das nicht-kodierende Genom deutlich erschwert. RNA-Sequenzierung (RNA-seq) andererseits könnte die Priorisierung von Varianten erleichtern indem Einflüsse der Varianten auf die Menge und die Sequenz der RNA offenbart werden. Ich habe daher 105 Fibroblastenzelllinien von diagnostizierten und undiagnostizierten Patienten ausgewählt und eine standardisierte RNA-seq-Analyse vorgenommen. Mittels manueller Inspektion der Daten stellte ich abnormale Expression des Atmungskettenkomplex I (RCCI) Assemblierungsfaktors *TIMMDC1* fest, welcher zuvor noch nicht mit erhöhtem Krankheitsrisiko in Verbindung gebracht wurde. Weiterhin konnte ich eine intronische Variante identifizieren, die wahrscheinlich in der Aktivierung einer kryptischen intronischen Spleißstelle involviert ist was wiederum zu abnormalem Spleißen und letztendlich zu Nonsense-mediated mRNA

Decay (NMD) führt. Mit Hilfe von funktionellen Validierungsexperimenten konnte die pathogene Relevanz der TIMMDC1-Defiziens zeigen, sowie 3 weitere Patienten identifizieren. Die Ergebnisse stießen eine systematische Analyse an, die in enger Zusammenarbeit mit der Abteilung für Computational Biology der Technischen Universität München durchgeführt wurde. Die systematische Analyse erzielte einen Median von einem abnormal exprimierten Gen, fünf abnormal gespleißten Genen und sechs mono-allelisch exprimierten Genen pro Probe. Diese überschaubare Anzahl von Ereignissen ermöglichte eine manuelle Inspektion und Validierung die zur Diagnose von 10 % (5 von 48) der Patienten ohne vorherige Diagnose führte. Es ist hervorzuheben, dass unsere Methode sowohl die Identifikation von kausalen nicht-kodierenden Varianten in vormals nicht krankheitsassoziierten Genen erlaubt, als auch die Identifikation der pathogenen Variante in Genen deren Implikation über mitochondriale Erkrankung hinaus geht.

Table of contents

Abbreviations

aa	amino acid
Ab	Antibody
acetyl-CoA	acetyl coenzyme A
ad	autosomal dominant
ATP	adenosine triphosphate
bp	base pair
BN-PAGE	blue native polyacrylamide gel electrophoresis
BSA	bovine serum albumin
BWA	Burrows-Wheeler Aligner
BWT	Burrows-Wheeler Transform
c.	cDNA sequence position
cDNA	complementary DNA
chr	chromosome
CNS	central nervous system
CNV	copy number variant
CPEO	chronic progressive external ophthalmoplegia
CSF	cerebrospinal fluid
C-terminus	carboxy-terminus
Da	dalton
dbSNP	Single Nucleotide Polymorphism Database
dH2O	deionised water
DHPLC-H2O	denaturing high-performance liquid chromatography water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
ds	double-stranded
ECL	enhanced chemiluminescence
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
e.g.	exempli gratia

et al.	et alii
ExAC	Exome Aggregation Consortium
FAD	flavin adenine dinucleotide
FAF-BSA	fatty acid free bovine serum albumin
FASTQ	text format for representing sequencing reads
FBS	foetal bovine serum
FCCP	carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone
FGF21	fibroblast growth factor 21
FPKM	fragments per kilobase million
G418	geneticin disulphate
GAC	Genome Analysis Center, Helmholtz Zentrum München
Gb	gigabase
GDF15	growth differentiation factor 15
gDNA	genomic DNA
gnomAD	Genome Aggregation Database
GTEx	Genotype-Tissue Expression project
HEK293	human embryonic kidney 293 cells
hg19	human genome assembly GRCh37 (February 2009), UCSC
HGMD	Human Gene Mutation Database
HRP	horseradish peroxidase
IGV	Integrative Genomics Viewer
IHG	Institute of Human Genetics, Helmholtz Zentrum München
IMM	inner mitochondrial membrane
IMS	intermembrane space
indel	small insertion and deletion variation
iPSC	induced pluripotent stem cells
kb	kilobase
kDa	kilodalton
KSS	Kearns-Sayre syndrome
L	litre
LDH	lactate dehydrogenase
LHON	Leber hereditary optic neuropathy

m.	mitochondrial DNA sequence position
MAE	Mono-allelic expression
Mb	megabase
MELAS	mitochondrial myopathy, encephalopathy, lactic acidosis, stroke-like episodes
MERRF	myoclonic epilepsy with ragged-red fibers
MNGIE	mitochondrial neurogastrointestinal encephalopathy
MRI	Magnetic Resonance Imaging
mRNA	messenger ribonucleic acid
mtDNA	mitochondrial DNA
NA	not available
NAD	nicotinamide adenine dinucleotide
NARP	neuropathy, ataxia and retinitis pigmentosa
NCBI	National Center for Biotechnology Information
NGS	next-generation sequencing
NM	RefSeq mRNA sequence
NMD	nonsense-mediated decay
NP	RefSeq protein sequence
N-terminus	amino-terminus
OMIM	Online Mendelian Inheritance in Man
OMM	outer mitochondrial membrane
ON	overnight
OXPHOS	oxidative phosphorylation
p.	protein sequence position
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PE	paired-end
PEO	progressive external ophthalmoplegia
PMSF	phenylmethylsulfonylfluorid
PVDF	polyvinylidene fluoride
RefSeq	NCBI Reference Sequence Database

RC	respiratory chain
RCC	Respiratory chain complex
RCCI	Respiratory chain complex I
RCCII	Respiratory chain complex II
RCCIII	Respiratory chain complex III
RCCIV	Respiratory chain complex IV
RCCV	Respiratory chain complex V
RFLP	restriction fragment length polymorphisms
RIN	RNA integrity number
RNA	ribonucleic acid
RNA-seq	RNA sequencing
ROS	reactive oxygen species
RPKM	reads per kilobase million
rRNA	ribosomal RNA
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction
rxn	reaction
SBG	SERVA Blue G
SD	standard deviation
SDS	sodium dodecyl sulphate
SNV	single nucleotide variant
TBE	Tris-borat-EDTA
TBST	Tris-buffered saline and Tween 20
TCA	tricarboxylic acid
T_m	melting temperature
Tris	2-amino-2(hydroxymethyl)-1,3-propandiol
tRNA	transfer RNA
U	unit
UQ	ubiquinone
UQH_2	ubiquinol
UCSC	University of California, Santa Cruz
UTR	untranslated region

v/v	volume per volume
VUS	variant of unknown significance
WES	whole exome sequencing
WGA	wheat germ agglutinin
WGS	Whole genome sequencing
wt	wild-type
w/v	weight per volume

Publications

Peer-reviewed publications included in this thesis

Kremer LS, L'hermitte-Stead C, Lesimple P, Gilleron M, Filaut S, Jardel C, Haack TB, Strom TM, Meitinger T, Azzouz H, Tebib N, Ogier de Baulny H, Touati G, Prokisch H, Lombès A. Severe respiratory complex III defect prevents liver adaptation to prolonged fasting. J Hepatol. 2016 Aug;65(2):377-85. doi: 10.1016/j.jhep.2016.04.017. Epub 2016 May 2. PubMed PMID: 27151179.

Kremer LS, Distelmaier F, Alhaddad B, Hempel M, Iuso A, Küpper C, Mühlhausen C, Kovacs-Nagy R, Satanovskij R, Graf E, Berutti R, Eckstein G, Durbin R, Sauer S, Hoffmann GF, Strom TM, Santer R, Meitinger T, Klopstock T, Prokisch H, Haack TB. Biallelic Truncating Mutations in TANGO2 Cause Infancy-Onset Recurrent Metabolic Crises with Encephalocardiomyopathy. Am J Hum Genet. 2016 Jan 19. pii: S0002-9297(15)00504-2. doi: 10.1016/j.ajhg.2015.12.009. [Epub ahead of print] PubMed PMID: 26805782.

Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, Haack TB, Graf E, Schwarzmayr T, Terrile C, Konafikova E, Repp B, Kastenmüller G, Adamski J, Lichtner P, Leonhardt C, Funalot B, Donati A, Tiranti V, Lombes A, Jardel C, Gläser D, Taylor RW, Ghezzi D, Mayr JA, Rötig A, Freisinger P, Distelmaier F, Strom TM, Meitinger T, Gagneur J, Prokisch H (2017) Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Commun. 2017 Jun 12;8:15824. doi: 10.1038/ncomms15824. PubMed PMID: 28604674.

Peer-reviewed publications not included in this thesis

Gai X, Ghezzi D, Johnson MA, Biagosch CA, Shamseldin HE, Haack TB, Reyes A, Tsukikawa M, Sheldon CA, Srinivasan S, Gorza M, **Kremer LS**, Wieland T, Strom TM, Polyak E, Place E, Consugar M, Ostrovsky J, Vidoni S, Robinson AJ, Wong LJ, Sondheimer N, Salih MA, Al-Jishi E, Raab CP, Bean C, Furlan F, Parini R, Lamperti C, Mayr JA, Konstantopoulou V, Huemer M, Pierce EA, Meitinger T, Freisinger P, Sperl W, Prokisch H, Alkuraya FS, Falk MJ, Zeviani M. Mutations in FBXL4, encoding a mitochondrial protein, cause early-onset mitochondrial encephalomyopathy. Am J Hum Genet. 2013 Sep

5;93(3):482-95. doi: 10.1016/j.ajhg.2013.07.016. Epub 2013 Aug 29. PubMed PMID: 23993194; PubMed Central PMCID: PMC3769923.

Hildick-Smith GJ, Cooney JD, Garone C, **Kremer LS**, Haack TB, Thon JN, Miyata N, Lieber DS, Calvo SE, Akman HO, Yien YY, Huston NC, Branco DS, Shah DI, Freedman ML, Koehler CM, Italiano JE Jr, Merkenschlager A, Beblo S, Strom TM, Meitinger T, Freisinger P, Donati MA, Prokisch H, Mootha VK, DiMauro S, Paw BH. Macrocytic anemia and mitochondriopathy resulting from a defect in sideroflexin 4. Am J Hum Genet. 2013 Nov 7;93(5):906-14. doi: 10.1016/j.ajhg.2013.09.011. Epub 2013 Oct 10. PubMed PMID: 24119684; PubMed Central PMCID: PMC3824126.

Haack TB, Gorza M, Danhauser K, Mayr JA, Haberberger B, Wieland T, **Kremer L**, Strecker V, Graf E, Memari Y, Ahting U, Kopajtich R, Wortmann SB, Rodenburg RJ, Kotzaeridou U, Hoffmann GF, Sperl W, Wittig I, Wilichowski E, Schottmann G, Schuelke M, Plecko B, Stephani U, Strom TM, Meitinger T, Prokisch H, Freisinger P. Phenotypic spectrum of eleven patients and five novel MTFMT mutations identified by exome sequencing and candidate gene screening. Mol Genet Metab. 2014 Mar;111(3):342-52. doi: 10.1016/j.ymgme.2013.12.010. Epub 2013 Dec 25. PubMed PMID: 24461907.

Tischner C, Hofer A, Wulff V, Stepek J, Dumitru I, Becker L, Haack T, **Kremer L**, Datta AN, Sperl W, Floss T, Wurst W, Chrzanowska-Lightowlers Z, De Angelis MH, Klopstock T, Prokisch H, Wenz T. MTO1 mediates tissue specificity of OXPHOS defects via tRNA modification and translation optimization, which can be bypassed by dietary intervention. Hum Mol Genet. 2014 Dec 30. pii: ddu743. [Epub ahead of print] PubMed PMID: 25552653.

Brea-Calvo G, Haack TB, Karall D, Ohtake A, Invernizzi F, Carrozzo R, **Kremer L**, Dusi S, Fauth C, Scholl-Bürgi S, Graf E, Ahting U, Resta N, Laforgia N, Verrigni D, Okazaki Y, Kohda M, Martinelli D, Freisinger P, Strom TM, Meitinger T, Lamperti C, Lacson A, Navas P, Mayr JA, Bertini E, Murayama K, Zeviani M, Prokisch H, Ghezzi D. COQ4 Mutations Cause a Broad Spectrum of Mitochondrial Disorders Associated with CoQ10 Deficiency. Am J Hum Genet. 2015 Feb 5;96(2):309-17. doi: 10.1016/j.ajhg.2014.12.023. PubMed PMID: 25658047; PubMed Central PMCID: PMC4320255.

Haack TB, Jackson CB, Murayama K, **Kremer LS**, Schaller A, Kotzaeridou U, de Vries MC, Schottmann G, Santra S, Büchner B, Wieland T, Graf E, Freisinger P, Eggimann S, Ohtake A,

Okazaki Y, Kohda M, Kishita Y, Tokuzawa Y, Sauer S, Memari Y, Kolb-Kokocinski A, Durbin R, Hasselmann O, Cremer K, Albrecht B, Wieczorek D, Engels H, Hahn D, Zink AM, Alston CL, Taylor RW, Rodenburg RJ, Trollmann R, Sperl W, Strom TM, Hoffmann GF, Mayr JA, Meitinger T, Bolognini R, Schuelke M, Nuoffer JM, Kölker S, Prokisch H, Klopstock T. Deficiency of ECHS1 causes mitochondrial encephalopathy with cardiac involvement. Ann Clin Transl Neurol. 2015 May;2(5):492-509. doi: 10.1002/acn3.189. Epub 2015 Mar 13. PubMed PMID: 26000322; PubMed Central PMCID: PMC4435704.

Haack TB, Staufner C, Köpke MG, Straub BK, Kölker S, Thiel C, Freisinger P, Baric I, McKiernan PJ, Dikow N, Harting I, Beisse F, Burgard P, Kotzaeridou U, Kühr J, Himbert U, Taylor RW, Distelmaier F, Vockley J, Ghaloul-Gonzalez L, Zschocke J, **Kremer LS**, Graf E, Schwarzmayr T, Bader DM, Gagneur J, Wieland T, Terrile C, Strom TM, Meitinger T, Hoffmann GF, Prokisch H. Biallelic Mutations in NBAS Cause Recurrent Acute Liver Failure with Onset in Infancy. Am J Hum Genet. 2015 Jul 2;97(1):163-9. doi: 10.1016/j.ajhg.2015.05.009. Epub 2015 Jun 11. PubMed PMID: 26073778; PubMed Central PMCID: PMC4572578.

Powell CA, Kopajtich R, D'Souza AR, Rorbach J, **Kremer LS**, Husain RA, Dallabona C, Donnini C, Alston CL, Griffin H, Pyle A, Chinnery PF, Strom TM, Meitinger T, Rodenburg RJ, Schottmann G, Schuelke M, Romain N, Haller RG, Ferrero I, Haack TB, Taylor RW, Prokisch H, Minczuk M. TRMT5 Mutations Cause a Defect in Post-transcriptional Modification of Mitochondrial tRNA Associated with Multiple Respiratory-Chain Deficiencies. Am J Hum Genet. 2015 Aug 6;97(2):319-28. doi: 10.1016/j.ajhg.2015.06.011. Epub 2015 Jul 16. PubMed PMID: 26189817; PubMed Central PMCID: PMC4573257.

Holzerova E, Danhauser K, Haack TB, **Kremer LS**, Melcher M, Ingold I, Kobayashi S, Terrile C, Wolf P, Schaper J, Mayatepek E, Baertling F, Friedmann Angeli JP, Conrad M, Strom TM, Meitinger T, Prokisch H, Distelmaier F. Human thioredoxin 2 deficiency impairs mitochondrial redox homeostasis and causes early-onset neurodegeneration. Brain. 2015 Dec 1. pii: awv350. [Epub ahead of print] PubMed PMID: 26626369.

Van Haute L, Dietmann S, **Kremer L**, Hussain S, Pearce SF, Powell CA, Rorbach J, Lantaff R, Blanco S, Sauer S, Kotzaeridou U, Hoffmann GF, Memari Y, Kolb-Kokocinski A, Durbin R, Mayr JA, Frye M, Prokisch H, Minczuk M. Deficient methylation and formylation of mt-

tRNA(Met) wobble cytosine in a patient carrying mutations in NSUN3. Nat Commun. 2016 Jun 30;7:12039. doi: 10.1038/ncomms12039. PubMed PMID: 27356879; PubMed Central PMCID: PMC4931328.

Kremer LS, Danhauser K, Herebian D, Petkovic Ramadža D, Piekutowska-Abramczuk D, Seibt A, Müller-Felber W, Haack TB, Płoski R, Lohmeier K, Schneider D, Klee D, Rokicki D, Mayatepek E, Strom TM, Meitinger T, Klopstock T, Pronicka E, Mayr JA, Baric I, Distelmaier F, Prokisch H. NAXE Mutations Disrupt the Cellular NAD(P)HX Repair System and Cause a Lethal Neurometabolic Disorder of Early Childhood. Am J Hum Genet. 2016 Oct 6;99(4):894-902. doi: 10.1016/j.ajhg.2016.07.018. PubMed PMID: 27616477; PubMed Central PMCID: PMC5065653.

Non-peer-reviewed publications not included in this thesis

Kremer LS, Prokisch H. Identification of Disease-Causing Mutations by Functional Complementation of Patient-Derived Fibroblast Cell Lines. Methods Mol Biol. 2017;1567:391-406. doi: 10.1007/978-1-4939-6824-4_24. PubMed PMID: 28276032.

1. Introduction

1.1. Mitochondria

1.1.1. Mitochondrial functions

Mitochondria are double membrane-enclosed organelles present in almost all eukaryotic cells which likely arose through an endosymbiotic uptake of an ancient α-proteobacterium by an ancient Lokiarchaeum [1-4]. They are largely known as the powerhouse of the cell due to their crucial function in energy transduction. They exploit the energy stored in fats, carbohydrates, and proteins to produce ATP in a process called oxidative phosphorylation. In the cytosol, fats are decomposed to acyl-CoA via lipolysis whereas carbohydrates are broken down to pyruvate via glycolysis and proteins by different routes of amino acid degradation ending in pyruvate, Krebs cycle intermediates, or acetyl-CoA. Pyruvate and acyl-CoA are subsequently transported into the mitochondrial matrix, where they are further degraded into acetyl-CoA by the pyruvate dehydrogenase complex and the enzymes of the fatty acid ß-oxidation. Acetyl-CoA, in turn, is further oxidized to CO_2 by the tricarboxylic acid (TCA) cycle. The energy derived from this oxidative breakdown of pyruvate and acyl-CoA to CO_2 is transferred as electrons to NAD^+ and FAD^+ generating NADH and $FADH_2$. These electrons are subsequently harnessed by the electron transport chain composed of respiratory chain complex I-IV (RCCI-IV) embedded in the inner mitochondrial membrane (IMM). The 2 electrons derived from NADH are transferred by RCCI (NADH:ubiquinone oxidoreductase) to reduce ubiquinone (UQ) to ubiquinol (UQH_2), while the 2 electrons derived from $FADH_2$ are carried to UQ via RCCII (succinate:ubiquinone oxidoreductase) or by other FADH2 metabolizing enzymes (e.g. ETFDH, G3PDH, SQRDL). Subsequently, UQH_2 is oxidized by RCCIII (ubiquinol:cytochrome c oxidoreductase) and electrons are donated to cytochrome c. The reduction equivalents are finally used by RCCIV (cyctochrome c oxidase) to bivalently reduce O_2 to H_2O. The energy released by this energetically favored electron transport is deployed to translocate protons across the IMM by RCCI, RCCIII, and RCCIV generating a proton electrochemical gradient. This gradient is, in turn, used by RCCV (ATP synthase) to synthesize ATP from ADP and inorganic phosphate. The electron transfer by RCCI-IV and ATP synthesis by RCCV jointly constitute the process referred to as oxidative phosphorylation (OXPHOS). Besides this crucial function in energy transduction,

mitochondria furthermore play an important role in supplying important cofactors for numerous cellular processes as they are the site of e.g. Fe-S cluster, heme, and steroid biosynthesis, as well as amino acid biosynthesis and inter-conversion [5, 6]. Moreover, mitochondria house enzymes of the urea cycle and are involved in Ca^{2+} homeostasis and apoptosis [7, 8]. A schematic representation of the mitochondrial functions is illustrated in Figure 1.

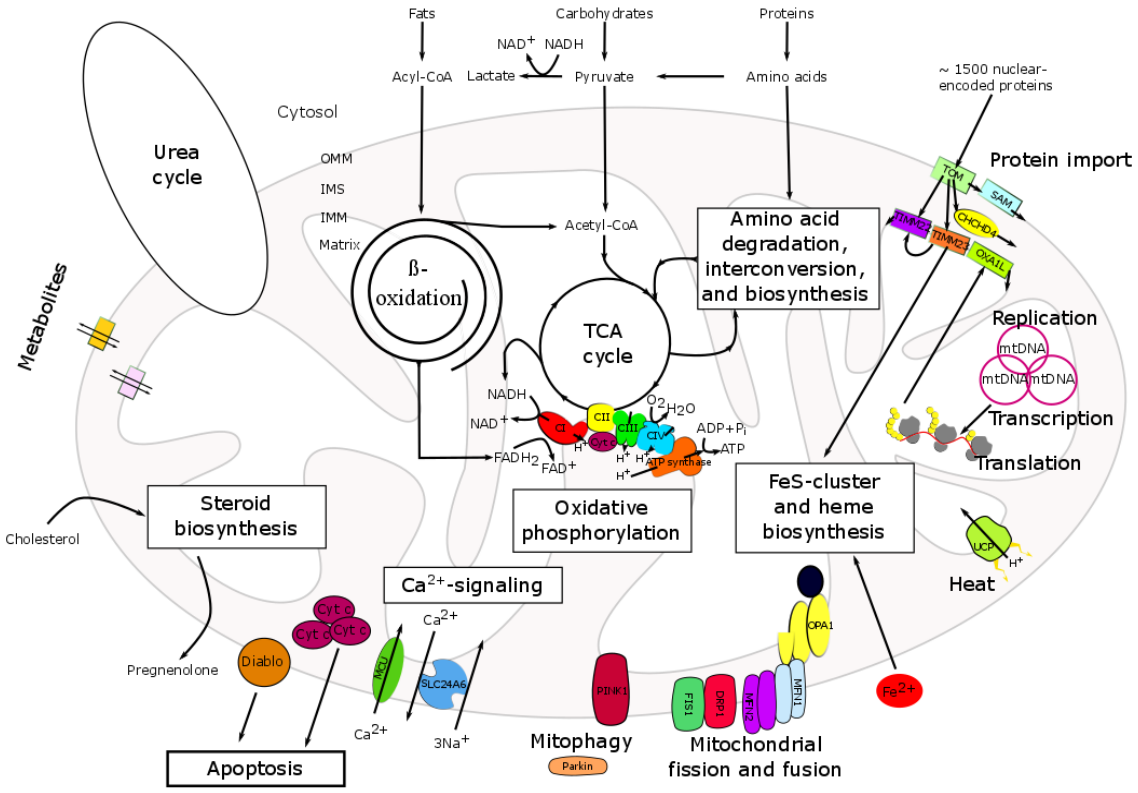

Figure 1 Schematic representation of a mitochondrion. Major metabolic pathways are depicted in boxes or encircled. Outer mitochondrial membrane (OMM) and inner mitochondrial membrane (IMM) are indicated with light grey lines, the intermembrane space (IMS) is enclosed by the OMM and IMM, the matrix is shaded in light grey.

1.1.2. Mitochondrial genomics

The workhorse of mitochondria, dedicated to fulfill the plethora of tasks stated above, is the mitochondrial proteome. It is comprised of approximately 1,500 proteins which are under the control of 2 genomes, the nuclear genome and the mitochondrial genome. While most of the

proteins are encoded by the nuclear genome and transported into mitochondria, the circular 16.6 kb mitochondrial genome (mtDNA) codes for 37 genes comprising 22 mitochondrial tRNAs, 2 mitochondrial rRNAs, and 13 subunits of OXPHOS complexes. The mtDNA is a remainder of the bacterial origin of mitochondria. The bulk of the genetic information originally encoded by the mtDNA was transferred to the nuclear genome in the course of evolution, some material remained within the mitochondria. The remaining genes encode highly hydrophobic proteins, making targeting and transport into the mitochondrion difficult. In addition, the retention of some genomic material within the mitochondria potentially enables an advantageous spatial control of the mitochondrial activity [9]. Each cell contains a few hundred to thousands of copies of the mtDNA molecule depending on the tissue or cell type. This multicopy nature of the mtDNA causes distinct features of mitochondrial genetics named homoplasmy and heteroplasmy. Homoplasmy is given when all mtDNA copies have the same genotype, heteroplasmy occurs if a mixture of different genotypes, e.g. wild-type and mutant mtDNA, coexist. While the diploid nuclear genes are inherited according to the Mendelian law, mtDNA is inherited solely by the oocytes via the mother [10].

1.2. Mitochondrial disorders

Mutations in genes encoding a mitochondrial protein can potentially cause disruption of the mitochondrial energy supply. Genetic diseases attributed to such an impairment of OXPHOS are generally referred to as mitochondrial. This primary defect in OXPHOS can not only be caused by mutations in genes encoding subunits or assembly factors of RCCs, but also by mutations in genes encoding proteins required for mtDNA replication, transcription, and translation, needed for the generation or transport of substrates in reactions upstream of the OXPHOS or cofactors of OXPHOS, and genes which encode proteins important for the homeostasis of mitochondria [11]. Among the physiological consequences of improper OXPHOS are decreased ATP production, imbalanced NAD^+/NADH pools, monovalent reduction of O_2 resulting in the generation of reactive oxygen species (ROS), Ca^{2+}, and halting of pathways feeding into the RC like TCA cycle and the fatty acid ß-oxidation [12, 13]. The later can yield in increased lactate and ketone body production, respectively. Abnormalities of mitochondrial function was first reported in 1959 in a patient with clinical symptoms comprising increased perspiration, weakness, and reduced body weight besides polyphagia [14, 15]. Interestingly, since the author reasoned that skeletal muscle depicts the

largest portion of the body's respiration and since a large amount of tissue was required to conduct biochemical measurements on mitochondria, mitochondria from skeletal muscle biopsies were chosen for the investigations. Analyzing the mitochondrial enzymatic activities in muscle biopsy is till now the state-of-the-art to biochemically diagnose mitochondriopathies. Ever since this first description, many more reports of mitochondrial disease patients followed and it has become evident that the phenotypic spectrum is extremely broad and clinical symptoms can involve any single tissue or organ as well as multiple organs at any age of disease onset. In 2001, Munnich and Rustin coined the term "any symptom, in any organ or tissue, at any age, with any mode of inheritance"[13]. Tissues highly depending on energy, like the central nervous system, cardiac and skeletal muscle, and liver, are more likely but not exclusively affected as depicted in Figure 2. The impairment of a specific tissue might originate from tissue-specific isoforms, tissue specific energy demands or regulations of the electron flux, as well as tissue-dependent threshold mutation level in case of heteroplasmy [16]. Childhood-onset disease has an estimated prevalence of 5 to 15 in 100,000 individuals and is often caused by recessive nuclear DNA mutations [17]. In contrast, adult-onset disorders mainly arise due to mtDNA mutations with a prevalence of 10 in 100,000 while the prevalence due to nuclear mutations is estimated to be only 3 in 100,000 individuals [17]. However, due to the diverse phenotypic spectrum, many patients might evade diagnosis resulting in an underestimation of disease prevalence. While the stated numbers are derived from cohort studies, population based studies revealed a carrier prevalence of up to 236 in 100,000 for a known pathogenic mutation (m.3243A>G) in the mtDNA [18]. Due to founder mutations and consanguinity in a specific population, population based studies can be biased and do not necessarily reflect the global distribution. Even though, with an average prevalence of 5-15 in 100,000, mitochondrial disorders constitute the largest and most heterogeneous groups of metabolic disorders [17, 19].

1.1. Diagnosis of mitochondrial disorders

1.1.1. Clinical diagnosis

Clinical diagnosis of mitochondrial disorders is strongly impaired by the broad variation of clinical symptoms and the symptomatic overlap with other disorders. Amongst the commonly recognized syndromes in childhood-onset disease are Leigh(-like) syndrome, the most

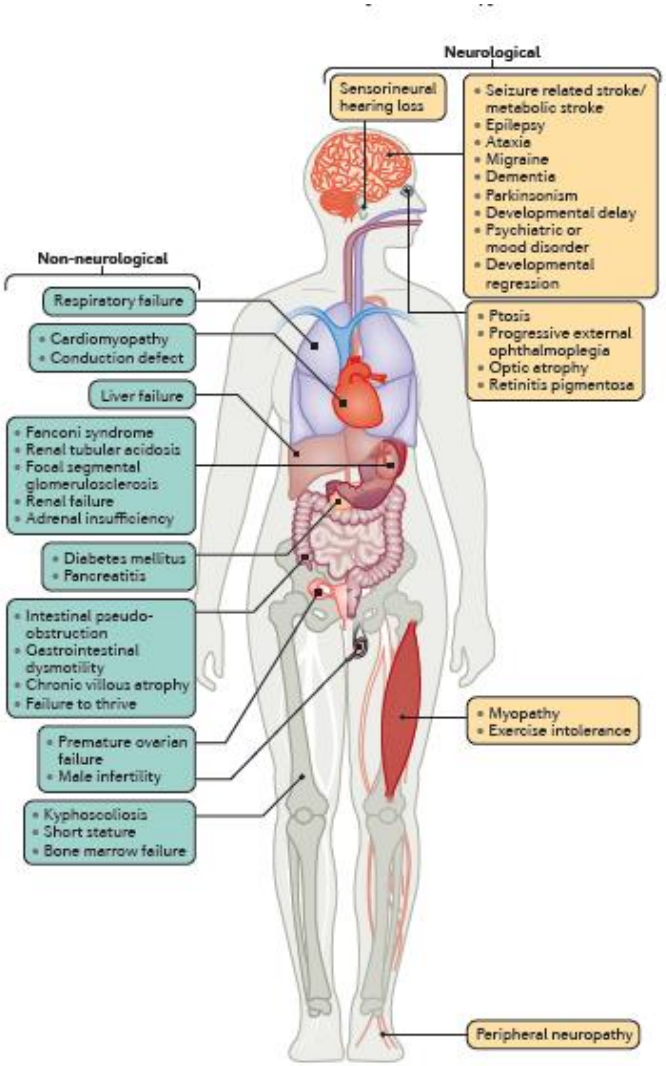

Figure 2 Phenotypic spectrum of mitochondrial disorders. The signs and symptoms are broadly divided into neurological symptoms on the right depicted in orange boxes and non-neurological symptoms on the left depicted in blue boxes. Figure taken from Gorman et al., 2016 [17].

frequent cause of childhood-onset disease, and Alpers syndrome. Leigh syndrome, also referred to as subacute necrotising encephalopathy, was first reported in 1951. It is a neurodegenerative disorder characterized by symmetrical lesions in the basal ganglia or the brain stem, which can be visualized by MRI [20]. Alpers syndrome presents with a triad of symptoms comprising seizures, liver dysfunction, and psychomotor retardation [21]. Syndromes associated with adult-onset disease include chronic progressive external ophthalmoplegia (CPEO), Kearns-Sayre syndrome (characterized by PEO, pigmentary retinopathy, and either or both, cardiomyopathy or cerebellar ataxia, KSS), LHON (Leber hereditary optic neuropathy), MERRF (myoclonic epilepsy with ragged-red fibers), MELAS

(mitochondrial myopathy, encephalopathy, lactic acidosis, stroke-like episodes), and NARP syndrome (neuropathy, ataxia and retinitis pigmentosa). The signs and symptoms of many patients, however, do not match any syndrome hampering a conclusive diagnosis as mitochondrial disorder. In general it has therefore been suggested to suspect a mitochondrial disorder upon the occurrence of the combined impairment of seemingly unrelated organs [22, 23]. For further reasoning metabolic or biochemical analyses are required.

1.1.2. Metabolic diagnosis

To compensate the ATP shortage caused by an impaired RC, the ATP production via glycolysis is promoted [24]. This yields in an excessive pyruvate production, which accumulates due to the halted TCA cycle and RC. At this stage, pyruvate can either be transaminated to alanine or reduced to lactate by the NADH-dependent lactate dehydrogenase (LDH), where in case of RC deficiency the increased $NADH/NAD^+$ ratio shifts the equilibrium from pyruvate to lactate [13]. Lactate is subsequently released into the blood stream. The increased lactate/pyruvate ratio in the blood can therefore serve as an indicator of mitochondrial disorder. Similarly, ketone bodies, comprising β-hydroxybutyrate, acetoacetate and acetone, can be employed as biomarkers for mitochondrial disorders [25]. Ketone bodies are produced in the liver from acetyl-CoA under conditions where acetyl-CoA cannot feed into the TCA cycle as the latter is halted, and secreted into the blood stream. Under physiological conditions this occurs upon starvation, when intracellular glucose levels are low resulting in a shortage of oxaloacetate, a crucial intermediate in the TCA cycle. Upon feeding, intracellular glucose levels increase, resulting in the formation of oxaloacetate from pyruvate and acetyl-CoA is subsequently predominantly feeds into the TCA cycle, the ketone body concentration in the blood decreases. Impaired RC, however, might halt the TCA cycle even though intracellular glucose levels would be sufficiently high after a meal. Therefore, in the case of mitochondrial disease, ketone body levels in the blood might rise despite feeding [26]. Further biomarkers of mitochondrial disorders frequently analyzed comprise creatine kinase, amino acids (e.g. alanine), acylcarnitines, and further organic acids and are often also accessed in cerebrospinal fluid (CSF) and urine [27]. The utility of these biomarkers is however limited by possible artefacts due to improper sample collection, tissue specific metabolic changes not detectable in blood, CSF, or urine, and poor sensitivity and specificity towards mitochondrial disorders [28]. Loeffen et al. indicated that even in the presence of a specific mitochondrial disorder, lactate is normal in 15% of the cases [29] while other studies

report a lactate sensitivity of 34-62% [28]. Respective metabolites might perform insufficiently when considering mitochondriopathies as a whole, but might allow the diagnosis of certain subgroups. Elevation of 3-methylglutaconic acid for example is indicative of mutations in *DNAJC19*, *OPA1*, *SERAC1*, *TAZ*, and *TMEM70* [30]. A recent diagnostic study revealed the usability of fibroblast growth factor 21 (FGF21) with a sensitivity and specificity of 92% for mitochondrial disorders presenting with muscle involvement [31]. A diagnostic study by Yatsuga et al. identified growth differentiation factor 15 (GDF15) as a promising new biomarker with a sensitivity of 98% and a specificity of 86% for mitochondrial disorders [32]. Future studies yet have to show the usability of these new biomarkers in routinely clinical practice.

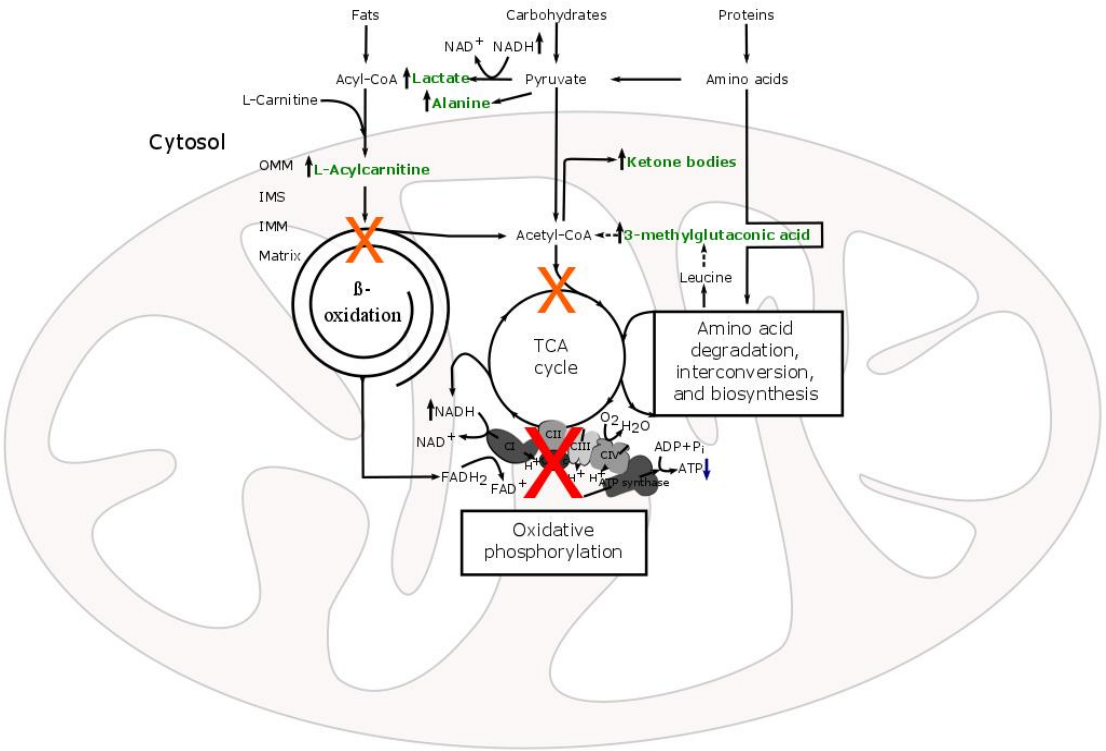

Figure 3 Metabolic biomarkers. Frequently employed metabolic biomarkers are depicted in green. Increase or decrease of the levels of the respective metabolite is indicated by a preceding or following upright arrow. Primary blockage of oxidative phosphorylation is indicated with a red cross, secondary blockage of the TCA cycle and ß-oxidation is indicated with an orange cross. Outer mitochondrial membrane (OMM) and inner mitochondrial membrane (IMM) are indicated with light grey lines, the intermembrane space (IMS) is enclosed by the OMM and IMM, the matrix is shaded in light grey.

1.1.3. Biochemical diagnosis

As mitochondrial disorders are characterized by faulty oxidative phosphorylation, biochemical measurements of the activities of enzymes directly or indirectly involved in oxidative phosphorylation by spectrophotometry and assessment of the overall mitochondrial respiration by polarography should provide a reliable read-out for the diagnosis of such a disorder [26]. The biochemical analysis might also help to classify mitochondriopathies according to the affected RCC or the nature of the defect, e.g. isolated RCC defect, combined RCC defect and such giving hints on the molecular cause. While biochemical investigations in mitochondria isolated from the skeletal muscle of the patient were already performed in the initial report by Luft in 1959 [14, 15], accurate assays for routinely diagnostics were only established in the 1970ies since the purification of sufficient amounts of intact mitochondria was tedious [33]. In 1985, DiMauro et al. outlined a biochemical strategy capable of distinguishing between impairment of the respiratory chain, defects of the TCA cycle, hampered substrate transport, disturbed substrate utilization, or defects in coupling [34]. Since then, biochemical assays on skeletal muscle have become the gold standard in the diagnosis of mitochondrial disorders while constantly being expanded and improved. Even nowadays though, the procedure has some critical drawbacks and results need to be evaluated carefully. While the analysis is mainly performed in skeletal muscle, skeletal muscle might not be the affected tissue and therefore not necessarily express any biochemical defect [19]. Biopsies on liver, heart, and CNS are possible, nevertheless by far more harmful and invasive to the patient. An additional obstacle is the inevitability of performing the biochemical analysis on frozen material in cases where the biopsy site and the site of the investigating laboratory are separated. The freezing can potentially damage the tissue or impact the metabolic state. This might lead to artificial results and it is therefore recommended to perform the biochemical measurements on fresh material. Even analysis on fresh tissue is however not invulnerable to artefacts which can be caused by improper sample handling or technical errors. In conclusion, biochemical investigations revolutionized the diagnostics of mitochondrial disorders but have their limitations. Negative findings do not exclude a mitochondriopathy, positive findings do not always prove a mitochondrial disorder [19]. This shortcoming might now be overcome by the upcoming of molecular genetic diagnostics.

1.1.4. Molecular genetic diagnosis

1.1.4.1. Single gene diagnosis

The molecular era began in 1977 with the introduction of Sanger sequencing [35]. Sanger sequencing allowed the analysis of the nucleotide sequence of a given piece of DNA of at that time up to 300 bp in size by using random incorporation of chain-terminating dideoxy nucleotides in a subset of reactions. While the first disease-associated gene in 1983 [36] and the first mtDNA abnormality in 1988 [37] were identified due to linkage analysis using restriction fragment length polymorphisms (RFLPs) which allowed pinpointing of the rough genomic location [38, 39], the combination with Sanger sequencing would soon revolutionize the field of molecular diagnostics and allow the localization of a pathogenic mutation with base resolution. As the human gDNA was not fully sequenced until 2003, this initial molecular era in respect to mitochondrial disorders was determined by the identification of mutations in mtDNA. With a rate of 10 discovery per year, more than 100 point mutations were listed by 2001 [40], while about 300 mutations in mtDNA had been described until 2012 [41]. Clinical syndromes were attributed to mtDNA point mutations or rearrangements of the mtDNA. In many cases however, the inheritance of the disease did not follow a solely maternal, hence mitochondrial, inheritance but the Mendelian law, therefore the pathogenic mutation had to reside in the nuclear DNA. In 1989 Zeviani et al. described a case of dominantly inherited disease, followed by a report of Moraes et al. in 1991 describing a case of likely recessive inheritance [42, 43]. However, the affected genes remained elusive. In 1995 Bourgeron finally reported the first mitochondrial disease-associated nuclear gene, SDHA [44], which was subsequently followed by many others. Assuming a genotype-phenotype correlation, the growing list of identified genes in combination with the respective clinical presentation prompted screening for candidate genes in a small scale. This small scale analysis was expanded to disease panels as sequencing costs were reduced. It became however evident, that the genotype-phenotype correlation for mitochondrial disorders is relatively weak. Mutations in the same gene or even the same mutation can give rise to many clinical phenotypes. The m.3243A>G mutation is associated with MELAS, but can also cause diabetes, hearing loss, CPEO, and Leigh syndrome [45]. Mutations in POLG, encoding the catalytic subunit of the mitochondrial DNA polymerase γ, can lead to Alpers syndrome, MNGIE (mitochondrial neurogastrointestinal encephalopathy), and CPEO [46]. Vice versa, a

given clinical phenotype can be caused by mutations in many genes. Leigh(-like) syndrome, for example, can originate from mutations in more than 75 genes [47].

1.1.4.2. Whole Exome Sequencing (WES)

The clinical and genetic heterogeneity of mitochondrial disorders, as well as their overlap with other disorders, asks for an untargeted diagnostic approach. This demand can be met by next generation sequencing techniques (NGS) employing massively parallel DNA sequencing in one reaction instead of setting-up single reactions. DNA is initially fragmented and ligated to adaptors of any kind. The DNA can subsequently be immobilized on a planar surface by bridge PCR or *in situ* polonies or on microbeads using emulsion PCR. Sequencing is performed in iterative cycles of an enzymatic synthesis reaction carried out either by a polymerase or a ligase and subsequent imaging [48]. While the introduction of NGS drastically reduced sequencing costs to 10^{-5} $ per nucleotide, routinely sequencing on genome-wide scale was still inadequate [49, 50]. As more than 85% of the identified disease-causing mutations are located in coding, hence exonic regions which constitute 2% of the genome, whole exome sequencing (WES) was therefore a cost effective alternative to whole genome sequencing (WGS) [51, 52]. Soon after the introduction of WES in 2009, the first molecular diagnosis gained by WES was reported for a patient suffering from congenital chloride diarrhea [53, 54]. The first mitochondrial disease gene identified by WES was published just one year later in 2010 by our group [55]. Since then, WES has developed to the gold standard of molecular diagnostics identifying around 300 new disease-associated genes every year [51]. Originally used mainly in a research setting, WES is now applied in routinely clinical diagnostics [56]. The break-through of WES cannot only be attributed to improvement in sequencing technology, but also to the development of bioinformatics pipelines which provide sequence alignment, annotation, variant calling, and further filtering [57-59]. The steps of a typical WES analysis are described in the following chapter.

1.2. WES analysis

1.2.1. Technical specifications

WES is the massively parallel sequencing of the exonic, hence protein coding regions of the genome. The exonic regions comprise only about 2% of the $3*10^9$ human genomic nucleotides but are predicted to harbor 85% of the Mendelian disease-causing variants [51,

52, 60]. To sequence the exonic regions of an individual's genomic DNA, the gDNA is first sheared into small fragments and subsequently ligated to adaptors. Exonic regions are then selectively captured by in-solution enrichment where the fragments are hybridized to biotinylated oligonucleotide baits [61, 62]. The hybridized fragments are pulled down by magnetic streptavidin beads and amplified by PCR. Massively parallel sequencing of the enriched and amplified sequences is performed. Commercial enrichment kits are offered from Agilent (SureSelect Human AllExon Kit), Illumina (TruSeq Exome Enrichment Kit), and Roche (Nimblegen SeqCap EZ Exome), differing in the bait type (DNA or RNA) and bait length, as well as the captured regions. Comparisons of the platforms are described elsewhere [62-64].

1.2.2. Variant detection

The raw WES data is stored as FASTQ file containing the short sequence reads with quality values for each base. The data is subsequently processed by aligning the sequences to a reference genome, calling single nucleotide variants (SNVs) and small insertions and deletions (indels) between the input sequence and the reference sequence, and subsequently filtering out low quality variants and annotation of the remainder high quality variants.

Alignment in this study was performed with the Burrows-Wheeler-Alignment (BWA) tool which uses the Burrows-Wheeler Transform (BWT) to map sequence reads to the reference genome by a backwards search. This search algorithm allows mismatches and gaps, therefore enabling the alignment of longer reads harboring indels [57].

Variant calling, in principle, is the piling-up of all sequenced bases aligning to a certain position and calculating the proportion of bases differing from the reference. Variants with a proportion higher than 30% are called as heterozygous, variants with a proportion higher than 80% are called as homozygous. This simplified depiction does however not account for properties like base or mapping quality, which especially distorts calling at low read depth. More sophisticated variant callers hence employ Bayesian models, such as SAMtools. In contrast, the recently developed GATK HaplotypeCaller first determines regions varying from the reference, so called active regions [65]. For each active region, it *de novo* assembles all possible haplotypes. Each individual read is then aligned to each possible haplotype to derive a per-read likelihood of a haplotype which is subsequently used to determine a per-read likelihood of an allele for each variant site. This is used to calculate the most likely genotype.

The GATK HaplotypeCaller was shown to reliably detect SNVs and outperform other callers for detecting SNVs and indels [66]. Despite this improvement, the detection of indels remains challenging.

A major obstacle of short read sequencing technologies as WES is the detection of structural variants and copy number variants (CNVs), which can range from 1 basepair to several megabases [67, 68]. Out of the 5 different strategies developed for CNV detection, which are read depth, split-read, paired-end, assembly, and a combination approach, read depth has proven especially useful for WES data as the size of the target region is between 100 and 300 bp [68, 69]. For read depth analysis, the normalized read depth of a chromosomal window, e.g. an exon, is compared to an expected read depth in this window based on a statistical model. The confidence of this statistical analysis increases with the window size, thus hampering evaluation of smaller CNVs. In this study, CNVs were called using ExomDepth which employs a Hidden Markov Model to compare the read depth of an exon in a given sample to the read depth of the respective exon in around 10 control samples [67]. Evaluation of ExomeDepth by Tan and colleagues showed that it had the highest sensitivity when compared to other tools, but also has a considerable high false positive rate [68]. Therefore, results need to be evaluated carefully.

1.2.3. Variant annotation

After variant calling, variants are filtered for minimal read depth and quality and annotated with additional information using annotating tools like ANNOVAR [70], SnpEff [71], or customized in-house tools as used in this study. Based on a gene definition file derived for example from the University of California Santa Cruz (UCSC) annotation database, the variant is annotated in regard to its genomic location as intergenic, 5'- UTR, exonic, splice site, intronic, or 3'-UTR variant. For exonic variants, the consequence of the variant on the protein sequence is predicted based on the mRNA sequence as synonymous, non-synonymous, frameshift, stop-gain, stop-loss, etc. [70, 71].

For the clinical interpretation of detected variants, frequency information and pathogenicity scores from public databases should be considered. The most comprehensive database for frequency information are ExAC and gnomAD [72]. Several tools for in silico prediction of conservation scores by aligning the human reference genome to multiple other mammalian or vertebrate genomes were developed [51, 73, 74]. Further lines of evidence indicating the

pathogenicity of a variant are prediction scores estimating the impact of the variant on protein function. These prediction tools, e.g. CADD [75], MutationTaster [76], PolyPhen-2 [77], and SIFT [78, 79] also include predictive models trained by using variants with known effects. Conserved regions are expected to be functionally important and mutations in such conserved regions are therefore assumed to be more likely deleterious than mutations in non-conserved regions. While this logic has been useful in some cases, one needs to be aware that there are exceptions. Harmful variants have also been identified in non-conserved regions. In contrast, the predicted deleterious effect of a variant in a conserved region might be compensated for by other variants [80]. Results from the prediction tools need to be judged with caution.

1.2.4. Prioritizing variants

A WES analysis yields on average between 20,000-23,000 SNVs which are mainly comprised of common polymorphisms and some sequencing errors [61, 81]. To identify the disease causative variant, sophisticated evaluation of the variants identified by WES is necessary. As outlined by MacArthur and coworkers, the causality of a variant for a given disease phenotype should be assessed in a two-step approach, where first the involvement of the candidate gene and subsequently the candidate variant(s) are evaluated using frequency information and statistical measures supplemented by experimental data [82]. Re-evaluation of published causal variants under the light of upcoming publicly available control cohorts revealed that a considerable amount of these causal variants was incorrectly assigned or lacked convincing evidence [83, 84]. MacArthur and colleagues therefore emphasize the need for such evaluations even if the causality of the candidate gene and candidate variants for the respective disorder have previously been reported.

In compliance with the above guidelines, candidate genes for a rare disorder like mitochondrial disease are first selected by filtering genes harboring non-synonymous variants for minor allele frequency. Whereas the scientific community defines variants with a minor allele frequency of less than 1% in a control population [82, 85] as rare, more stringent or more relaxed filtering can be employed in certain cases with the exact filter criteria largely depending on the disease under investigation [81]. For example, the Phe508del mutation in *CFTR* is responsible for two-thirds of the European cases of the recessive disorder cystic fibrosis, the most common lethal genetic disorder in Caucasians [81, 86]. With an allele frequency of 0.0106 in Europeans reported by ExAC, this mutation therefore exceeds the 1%

cut-off [72]. For our standard analysis, we apply a very stringent filtering using a MAF cut-off of 0.1% as there is only one mitochondrial disease-associated gene, *MTFMT*, harboring a more frequently observed pathogenic variant. Commonly employed control populations comprise private in-house database and publicly available databases as ExAC and gnomAD [72]. In-house cohorts are generally small, limiting robust statistical evaluation of variant frequencies. However, they enable the correction for systematic errors of the exome analysis pipeline and allow evaluation of high frequency variants present in isolated populations due to founder mutations. Publicly available databases usually comprise larger sample sizes, but need not always depict a global population and might not always only include healthy controls [87]. Moreover, they only provide single SNV frequencies and no individual level information like bi-allelic variants in a given gene. The filtering step for rare variants is highly efficient by yielding on average 100 private variants per sample but is hence seldom sufficient to reveal the causative variant on its own [81].

The list of candidate variants can be further enclosed by filtering for the assumed pattern of inheritance. Most mitochondrial disorders follow a recessive mode of inheritance and therefore only bi-allelic variants are considered causal. Especially in the case of consanguinity, the disease-causing mutation is probably homozygous. While autosomal dominant mutations are seen in adult onset PEO and KSS, most childhood onset mitochondrial diseases are inherited in a recessive fashion [88]. X-linked inheritance can be suspected if, in a family, males which are related to each other via a female are affected. Very few pathogenic *de novo* mutations in nuclear DNA have been described mitochondrial disease patients, whereas mtDNA deletions mostly occur sporadic [88-90]. Kong et al. reported a genome wide *de novo* mutation rate of 1.20×10^{-8} per nucleotide per generation, hence 1 SNV in the coding region [91]. However, identification of a *de novo* SNV out of the thousands of SNVs detect by WES is only possible by trio sequencing of the patient and the mother and the father.

Once variants are selected based on the expected type of inheritance, variants can further be prioritized based on reported disease associations listed in public databases as the Human Gene Mutation Database (HGMD) [92] or ClinVar [93]. As stated above, variants reported in public databases contain false positives as well as variants lacking conclusive evidence and need to be interpreted with caution. If no pathogenic variants are present, novel variants in

disease-associated genes listed in OMIM [94] can be prioritized. Additionally, experimental evidence can aid the selection of variants. E. g. for mitochondrial disorders, variants in genes encoding mitochondrial proteins or related to metabolic or biochemical findings are prioritized. The remainder variants can be prioritized using statistical significance calculations. These calculations depend on the mode of inheritance, the number of samples, gene size, mutation rate, and selective constraint [82]. MacArthur recommends a genome wide threshold of 1.7×10^{-6} (Bonferroni-corrected *P*-value) assuming a representative WES analysis on one individual, testing for 30,000 genes (21,000 protein-coding and 9,000 long noncoding RNA genes) [82]. Taking into account sample sizes, Wieland reported a genome wide threshold of 7.4×10^{-9} for bi-allelic variants in *ACAD9*, a RCCI assembly factor which is amongst the most frequently mutated genes responsible for isolated RCCI deficiencies [81, 95]. He detected bi-allelic variants in 15 of 623 mitochondrial disease patients and 7 of 3,969 samples suffering from other diseases. In addition to statistical means, conservation and prediction scores (see 1.2.3) can provide further indications for the causality of a variant as well as experimental data reported in the literature.

1.2.5. Validation of WES findings

For molecular diagnostics a validation of the likely causative variant identified by WES is mandatory [55]. At first, sequencing errors and sample mixing should be excluded by confirmation of the variant using Sanger sequencing. For variants of unknown significance (VUS), defined as variants which association with disease is unclear, further steps are necessary. Segregation of the variant in the family needs to be confirmed. Presence of segregation per se does not prove the causality of a variant as other benign variants might display the same haplotype. However, variants which do not segregate can immediately be dismissed. Additionally, conservation and prediction of functional impairment can hint at a causative role of the variant, but is not a mandatory prerequisite of pathogenic variants as described in 1.2.3. Final evidence of the effect of a variant on protein function can only arise from functional studies and functional studies are therefore extremely powerful.

Functional studies like RT-PCR, RNA-Seq, western blot analysis, or quantitative proteomics can be used to investigate the impact of a variant on transcript and protein stability. For mitochondrial disorders, biochemical assays as measurement of the activity of the RCC, measurement of the oxygen consumption rate, or assays tailored to the protein under

investigation can reveal a phenotype associated with the disease. Finally, rescuing the respective observed phenotype in the patient-derived cells by re-expression of a wild-type copy of the gene demonstrates the pathogenicity of the variant [55]. Furthermore, convincing evidence for a new genotype-phenotype correlation needs likely pathogenic variants in the same gene in unrelated patients sharing a common distinct phenotype. An overview of analysis workflow is illustrated in Figure 4.

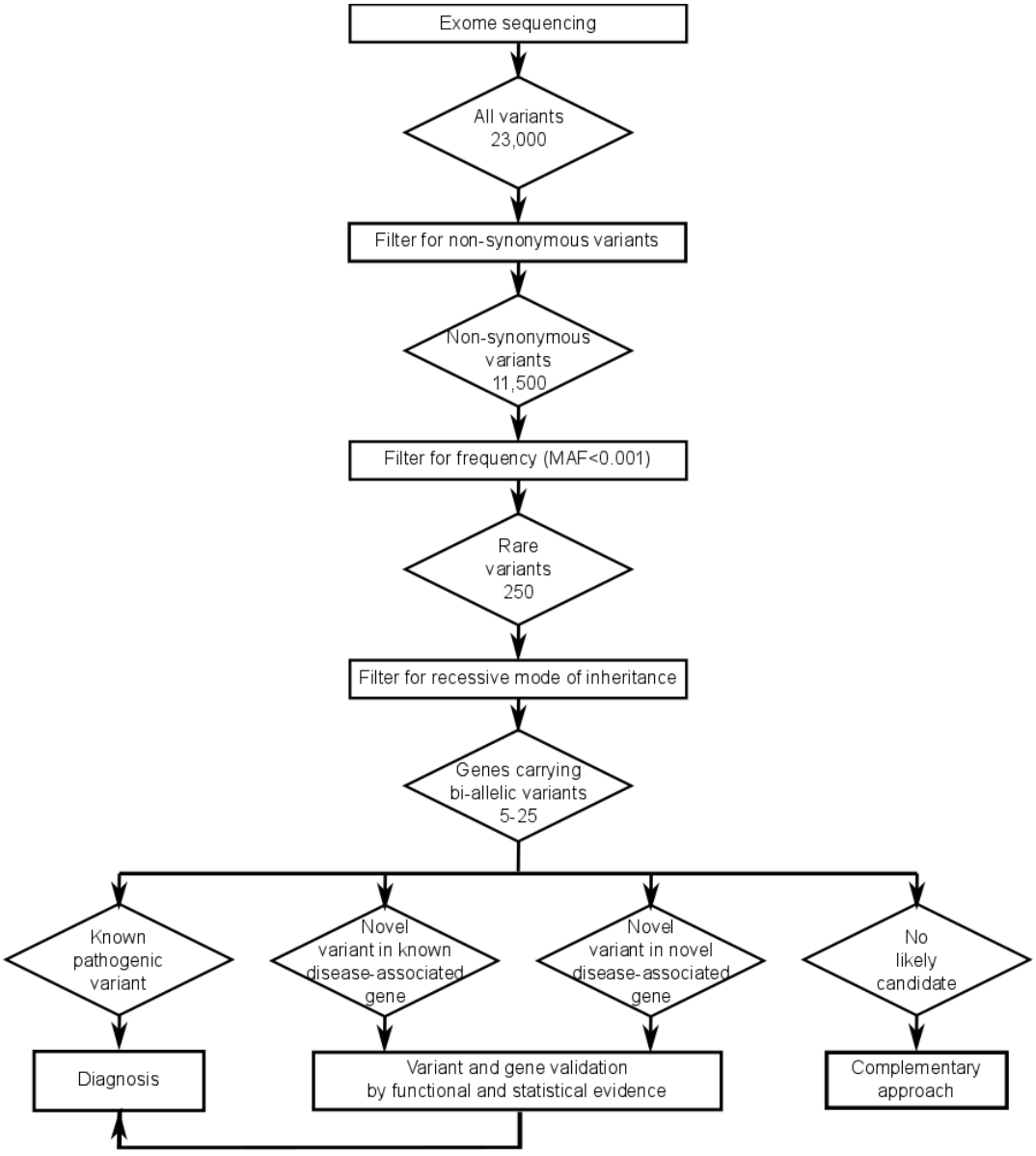

Figure 4 Flowchart for the identification and validation of pathogenic variants employing WES data. Numbers are representative for a standard WES analysis on patient blood or fibroblasts.

1.2.6. Diagnostic yield

WES has been successfully applied for a broad range of disorders. In heterogeneous cohorts a molecular diagnosis was achieved in about one third of the patients and stratified cohorts allowed detection in up to half of the patients depending on the investigated disease [96]. For mitochondrial disorders, in 2014, Taylor et al. reported a success rate of 53% for patients with combined OXPHOS defects hinting at impaired mitochondrial translation [97]. In another study in 2015, Wortmann et al. could diagnose 39% of a heterogeneous group of suspected mitochondrial disease patients whereas a success rate of 57% was achieved taking additional clinical, histochemical, biochemical, and neuroradiological information into account [56]. In a Japanese mitochondrial disease patient cohort, Ohtake et al. identified the molecular cause in 43% of the cases [98]. Taken together, the diagnostic outcome of WES leaves roughly half of the patients undiagnosed.

1.3. Complementary approaches for inconclusive WES

Inconclusive WES can originate from the incapability to capture the causative variant or the incapability to prioritize the causative variant. As WES is the sequencing of the coding genome solely, non-coding variants are not captured. But even coding variants might evade detection due to technical limitations mainly arising from the exon enrichment step. The insufficient capture of variants by WES can be overcome by WGS, which enables a complete coverage of the whole genome. As sequencing costs further decrease, WGS is becoming more widely used. However, WGS does not resolve the incapability to prioritize a variant. Quite the opposite, the incomplete understanding of the non-coding genome and vast number of detected variants by WGS render molecular diagnosis even more difficult. RNA sequencing (RNA-seq), the sequencing of the entire transcriptome, on the other hand, allows direct probing of the effect of variants, non-coding and coding, on RNA abundance and sequence. Variants in non-coding regulatory regions as promoters, enhancers, and suppressors but also variants in coding regions can result in aberrant expression of a gene which can be computed as expression outliers. Notably, expression outliers can also result from degradation of the RNA due to nonsense-mediated decay (NMD). Variants affecting splice sites or splice motifs can result in splicing defects detected as aberrant splice isoforms. Besides the direct detection of aberrant splicing, aberrant splicing can also result in expression outliers if the splice defect leads to the generation of a premature stop codon hence provoking NMD. If only one allele is

affected by aberrant expression or aberrant splicing, this can result in mono-allelic expression (MAE) of the other allele. So far, no systematic study was performed for the detection of aberrant expression, aberrant splicing, and MAE.

1.4. Treatment of mitochondrial disorders

Treatment of mitochondrial disorders has only been shown for a subset of mitochondrial disorders mostly comprising cofactor deficiencies and defects resulting in the accumulation of toxic metabolites [99]. For cases of cofactor deficiencies a beneficial effect of vitamin supplementation has been shown. For example, disease progression in Brown-Vialetto-Van Laere syndrome, a neurometabolic disorder caused by defects in biotin transporters, can be delayed or even stopped by riboflavin supplementation [100]. Accumulation of toxic metabolites on the other hand can be prevented by restricting substrate availability. Preliminary studies propose benefit of valine-restricted diet for defects in the valine catabolic pathway caused by mutations in *HIBCH* and *ECHS1* probably provoking secondary OXPHOS deficiency due to accumulation of toxic metabolites [101, 102]. For most patients though, therapeutic options are unfortunately limited to supportive care. To select cases amenable to treatment, molecular diagnosis is crucial. Nevertheless, also for the remainder, molecular diagnosis is necessary as living without diagnosis and recurrent cycles of genetic represent a major obstacle to patients. Furthermore, molecular diagnosis is fundamental for genetic counseling and prenatal testing. Finally, the development of new therapeutic strategies is mainly guided by treatment trials requiring patient stratification according to their molecular diagnosis [62].

1.5. Objectives

OMIM (Online Mendelian Inheritance in Man; http://www.ncbi.nlm.nih.gov/omim) catalogued 4951 disease-associated genetic loci by February 2017. With a rate of 300 new disease-associated genes per year and estimate of at least 7750 and probably 15300 potential disease-associated genes among the 25000-30000 genes in humans, there are still many disease-associations awaiting discovery [51]. For the genetic and phenotypic heterogeneous group of mitochondrial disorders, more than 250 disease-associated genes are reported but more than 1000 genes are predicted to encode mitochondrial proteins whose disruption could possibly cause disease [11]. Despite the tremendous success of WES as a diagnostic tool, around half of the suspected mitochondrial disease patients lack a molecular diagnosis. To

point out obstacles encountered when performing WES on suspected mitochondrial disease patients and provide solutions how to overcome these hurdles I pursued the following cases:

i. WES revealed homozygous variants in two genes encoding mitochondrial proteins, *MTO1* and *LYRM7*. It was unclear whether both variants were pathogenic and whether the patient possibly suffered from two distinct or overlapping diseases. Therefore, the causality of the variants had to be investigated.

ii. WES identified variants in a gene, *TANGO2*, previously not associated with any disease in several patients encoding a protein with unclear cellular localization and function. Functional assays needed to be developed to shed light onto the protein function and the respective pathomechansism.

iii. WES was not successful to provide a molecular diagnosis. As this could have been attributed to insufficient capture or prioritization of variants, the applicability of complementing approaches like WGS and RNA-seq was tested.

2. Material and Methods

2.1. Material

2.1.1. Nucleic acids

2.1.1.1. DNA

Genomic DNA of patients with suspected mitochondrial disorders belong to the DNA collection of IHG (Helmholtz Zentrum München, Munich, Germany). For all samples, informed consent was obtained.

2.1.1.2. cDNA

LYRM7 and *TIMMDC1* cDNA was purchased from DNASU Plasmid Repository (Tempe, AZ, USA) (Clone ID HsCD00514704 and Clone ID HsCD00442370, respectively). *MTO1* cDNA was purchased from GeneCopoeia (Rockville, MD, USA) (Cat# EX-H0858-Lv21).

2.1.2. Oligonucleotides

Oligonucleotides were synthesized by Metabion (Martinsried, Germany).

*TIMMDC1*_cDNA_forward	5'-CGCCATGGAGGTGCCG-3'
*TIMMDC1*_cDNA_reverse	5'-TCAGTCCTTGTCTTGTTTATCTATTA-3'
*LYRM7*_cDNA_forward	5'-CCACCATGGGACGGGCAGTCAAGGTTTTAC-3'
*LYRM7*_cDNA_reverse	5'-TCATTGCTTCTGAGTTGGTGCATCACA-3'
*TIMMDC1*_Intron5_forward	5'-GGGCATAATATATTCACAGTTGAGG-3'
*TIMMDC1*_Intron5_reverse	5'-ACAACAAAAGCAATGGCAGCA-3'
*ALDH18A1*_Exon14-15_forward	5'-TGTAAAACGACGGCCAGTTCTGTAAAAGGGAAGCTGCTG-3'
*ALDH18A1*_Exon14-15_reverse	5'-CAGGAAACAGCTATGACCGCTGTGCCTGGTCTAATTCC-3'
*ALDH18A1*_Exon16_forward	5'- TGTAAAACGACGGCCAGTTTGGGCTGTGGTTTTACAGG-3'

*ALDH18A1*_Exon16_reverse	5’- CAGGAAACAGCTATGACCGCAGGATCAGAA AGCAGC-3’

2.1.3. Cell lines

Primary patient fibroblast cell lines were derived from fresh skin biopsies. Normal human dermal fibroblasts (NHDF) from neonatal tissue (Lonza, Basel, Switzerland), and 293FT cells (Thermo Fisher Scientific, Waltham, MA, USA) were purchased.

2.1.4. Antibodies

Antibody	Manufacturer	Ordering number
CLPP	Abcam, Cambridge, UK	ab56455
MCOLN1	Abcam, Cambridge, UK	ab28508
NDUFA13	Abcam, Cambridge, UK	ab110240
NDUFB3	Abcam, Cambridge, UK	ab55526
NDUFB8	Abcam, Cambridge, UK	ab110242
MT-ND5	Abcam, Cambridge, UK	ab92624
TIMMDC1	Abcam, Cambridge, UK	ab171978
UQCRC2	Abcam, Cambridge, UK	ab14745
anti-rabbit HRP-conjugated Ab	Jackson Immuno Research Laboratories, West Grove, PA, USA	111-036-045
anti-mouse HRP-conjugated Ab	Jackson Immuno Research Laboratories, West Grove, PA, USA	115-036-062
anti-Giantin	Abcam, Cambridge, UK	ab80864
Alexa Fluor 568	Invitrogen, Eugene, OR, USA	A-11011
WGA-AF488 conjugate	Thermo Fisher Scientific, Waltham, MA, USA	W6748

2.1.5. Chemicals and solutions

Unless stated differently, chemicals and solutions were obtained either from Sigma-Aldrich (St. Louis, MO, USA) or Merck (Darmstadt, Germany).

2.2. Methods

2.2.1. Sequencing

2.2.1.1. Exome sequencing

Exonic regions from human DNA samples were enriched using the SureSelect Human All Exon kit from Agilent (Agilent Technologies, Santa Clara, CA, USA) and subsequently sequenced as 100 bp paired-end runs on a Illumina HiSeq2000 and Illumina HiSeq2500 (AG_50MB_v4 and AG_50MB_v5 exome kit samples) or as 76 bp paired-end runs on the Illumina GAIIx (AG_38MB_v1 and AG_50MB_v3 exome kit samples) (Illumina, San Diego, CA, USA). Reads were aligned to the human reference genome (UCSC Genome Browser build hg19) using Burrows-Wheeler Aligner (BWA, v.0.7.5a) and single nucleotide variants (SNVs) as well as small insertions and deletions (indels) were detected with SAMtools (version 0.1.19). On average, 22,500 SNVs were detected per individual. For further analysis, only non-synonymous SNVs with a minor allele frequency of less than 0.1% in our in-house database were considered. Assuming a recessive type of inheritance, on average 250 SNVs per individual were detected. Prioritization of variants in genes encoding mitochondrial proteins was eased by highlighting genes which are: 1) known mitochondrial disease-associated genes [11]; 2) encoding reported mitochondrial proteins [103]; 3) encoding predicted mitochondrial proteins [104].

2.2.1.2. RNA sequencing

RNA was isolated from whole-cell lysates using the AllPrep RNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Determination of the RNA integrity number (RIN) was performed using the Agilent 2100 BioAnalyzer (RNA 6000 Nano Kit, Agilent Technologies, Santa Clara, CA, USA). Starting with 1 µg of RNA, the library was prepared by poly(A) selection, followed by fragmentation and reverse transcription with the Elute, Prime, Fragment Mix (Illumina, San Diego, CA, USA). The resulting double-stranded cDNA was subjected to end repair, A-tailing, adaptor ligation, and subsequent library

enrichment according to the Low Throughput protocol of the TruSeq RNA Sample Prep Guide (Illumina, San Diego, CA, USA). Quality and quantity of the RNA libraries were assessed with the Agilent 2100 BioAnalyzer and the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA). Sequencing was performed as as 100 bp paired-end runs on an Illumina HiSeq2500 platform. RNA-seq reads were mapped to the hg19 genome assembly (UCSC Genome Browser build) using STAR (version 2.4.2a). FPKM were calculated as described [105]. For initial analysis, only transcripts detected in 95% of the samples were considered. Z-scores of log-transformed FPKM values were computed using Perseus (version Perseus_1.5.4.1.) [106] and ranked from lowest to highest for each patient. This set-up was used to generate and optimize algorithms allowing a systematic analysis performed by the group of Julien Gagneur as described elsewhere [107].

2.2.1.3. Sanger sequencing

DNA sequences were amplified by PCR using the Qiagen Taq DNA Polymerase Kit (Qiagen, Hilden, Germany) and a PeqStar thermal cycler (PeqLab Biotechnology, Erlangen, Germany). PCR reactions were performed in a 25 µL PCR reaction containing 25 ng gDNA, 0.5 U Taq DNA Polymerase, 1x PCR Buffer, 0.2 mM dNTPs, 1x Q-Solution, and 0.2 µM of each, forward and reverse primer using the following amplification conditions:

Step	Temperature [°C]	Time
Heat Lid	110	
Denature	95	5 min
Start Cycle (40x)		
Denature	95	30 s
Anneal	primer T_m - 5	30 s
Extend	72	1 min/kb
End Cycle		
Extend	72	10 min

1 µL of the PCR product was subjected to agarose gel electrophoresis to access quality and product size. The remainder was purified using the MultiScreen® PCRµ96 Filter Plate (Merck Millipore, USA) according to 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, CA, USA) as follows:

Sequencing reaction:

Component	5 µL reaction	Final concentration
BigDye Terminator v.3.1 Ready Reaction mix	1 µL	-
BigDye Terminator 5x Sequencing Buffer	1 µL	1x

10 µM forward or reverse primer	1 µL	2 µM
PCR product	1 µL	-
Ultrapure H_2O	1 µL	-

Sequencing program:

Step	Temperature [°C]	Time
Heat Lid	110	
Denature	96	1 min
Start Cycle (25x)		
Denature	96	10 s
Anneal	50	5 s
Extend	60	1 min 30 sec
End cycle		

For purification, the sequencing reaction was precipitated with 25 µL 100% ethanol for 15 min in the dark followed by centrifugation at 3000 g for 30 min at RT. The pellet was washed with 125 µL 70% ethanol, centrifuged at 2000 g for 10 min, and left to dry at RT in the dark. Subsequently, the pellet was resuspended in 25 µL ultrapure H_2O, transferred to a microtiter plate, and placed into the automated ABI 3730 sequencer. Resulting sequences were analysed using Staden Package (http://staden.sourceforge.net).

2.2.2. Cell culture

Primary patient fibroblast cell lines, NHDF, and 293FT cells were cultured in growth medium consisting of high glucose DMEM supplemented with 10% FBS, 1% penicillin/streptomycin, and 200 µM uridine (all Life Technologies, Carlsbad, CA, USA) at 37 °C and 5% CO2. All fibroblast cell lines have been tested negative for mycoplasma contamination using the MycoAlert Mycoplasma detection kit (Lonza, Basel, Switzerland).

2.2.3. Transduction and Transfection

Full-length *LYRM7* and *TIMMDC1* cDNA was overexpressed in fibroblast cell lines using the ViraPower HiPerform Lentiviral TOPO Expression Kit (Thermo Fisher Scientific, Waltham, MA, USA) as described previously [108]. In brief, both constructs were first cloned into the pLenti6.3 expression vector. The purchased cDNA was amplified by PCR using the Platinum® Taq DNA Polymerase High Fidelity Kit (Thermo Fisher Scientific, Waltham, MA, USA) in a 10 µL PCR reaction containing 5 ng cDNA, 0.2 U Platinum® Taq DNA Polymerase High Fidelity, 1x High Fidelity PCR Buffer, 0.2 mM dNTPs, 2 mM $MgSO_4$, and 0.2 µM of each, forward and reverse primer using the following amplification conditions:

Step	Temperature [°C]	Time
Heat Lid	110	
Denature	95	2 min
Start Cycle (30x)		
Denature	95	30 s
Anneal	60	30 s
Extend	68	1 min/kb
End Cycle		
Extend	68	10 min

The PCR product was analyzed using agarose gel electrophoresis and 1 µL was subsequently cloned into the pLenti6.3/V5-TOPO expression vector and transformed into One Shot® Stbl3™ Competent *E. coli* cells according to the manufacturer's instructions. 10 colonies were selected for Miniprep plasmid isolation using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and further Sanger sequencing. For subsequent transfection reactions, one sequence validated clone was subjected to Midiprep plasmid isolation using the QIAGEN Plasmid Midi Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

Full-length *MTO1* cDNA was directly obtained in the suitable expression vector and therefore immediately amplified using a Midiprep as described above. The *MTO1* cDNA was subsequently overexpressed in fibroblast cell lines using the Lenti-Pac™ FIV Expression Packaging Kit (GeneCopoeia, Rockville, MD, USA) as described below according to previous studies [55, 109].

For both systems, the expression vector was cotransfected with the respective packaging plasmid mix into 293FT cells using Lipofectamine 2000. The transfection mix was replaced with high glucose DMEM supplemented with 10% FBS 24 h after transfection. After an additional 72 h, the supernatant containing the viral particle was collected and used to subsequently transduce the fibroblast cell lines. Cells stably expressing the gene of interest were selected using 5 µg/mL Blasticidin (Thermo Fisher Scientific, Waltham, MA, USA), or 0.1 mg/mL G418 (Carl Roth, Karlsruhe, Germany), or a combination of both depending on the respective selection marker for 2 weeks.

2.2.4. Biochemical measurements

2.2.4.1. Seahorse

Cellular oxygen consumption can be used as a read-out of the mitochondrial fitness as the mitochondrial respiratory chain depicts the biggest oxygen consumer in the cell. Electron donors can derive from glucose, fatty acids and glutamine. In the following, two protocols are described, one to access the glucose-driven respiration, one to exploit respiration driven by the oxidation of fatty acids by ß-oxidation.

2.2.4.1.1. Mito Stress test – glucose-dependent respiration

The day before the assay, 20,000 fibroblasts/well were seeded in 80 µL of growth medium (see 2.2.2) in a XF 96-well cell culture microplate (Seahorse Bioscience, Agilent Technologies, Santa Clara, CA, USA) and incubated at 37 °C and 5% CO_2 ON. The sensor cartridge was rehydrate in the XF96 utility plate by adding 200 µl of XF calibrant solution per well. On the day of assay, cells were carefully washed once with and subsequently incubated in 180 µL bicarbonate-free DMEM (Life Technologies, Carlsbad, CA, USA) at 37 °C for 30 min prior measurement. The oxygen consumption rate (OCR) was measured using the XF96 Extracellular Flux Analyzer (Seahorse Biosciences). After an initial calibration step of 20 min, OCR was measured in repetitive cycles of 2 min mixing, 2 min waiting, 3 min measuring, 2 min mixing, 3 min measuring, 2 min mixing, and 3 min measuring. This measurement cycle was performed with no additions; after addition of oligomycin (1 µM) loaded in Port A; after addition of FCCP (0.4 µM) loaded in Port B; and after addition of rotenone (2 µM) loaded in Port C.

Following the measurement, cells were washed in PBS and subjected to cell number quantification using the CyQuant Cell Proliferation Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol.

2.2.4.1.2. ß-oxidation – palmitate-dependent respiration

Before the assay a solution of 1 mM pamitate/0.17 mM FAF-BSA was prepared as follows. A 44 mL solution of 2 mM palmitate in 150 mM sodium chloride was heated to 65 ºC to 70 ºC until the solution clarified. The solution was added to 50 mL of 0.34 mM FAF-BSA in 150 mM sodium chloride, pH 7.4, filtered. The combined solution was stirred for 1 h at 37 ºC, adjusted to pH 7.4 and aliquots were stored at -20 ºC. A 0.17 mM FAF-BSA solution was

prepared by mixing 50 mL of 0.34 mM FAF-BSA in 150 mM sodium chloride with 50 mL 150 mM NaCl. The solution was subsequently adjusted to pH 7.4, filtered, aliquoted, and stored at -20 ºC.

2 days before the assays, fibroblasts were seeded at a density of 20,000 cells/well in 80 µL growth medium (see 2.2.2.). The following day, cells were washed once and incubated in 160 µL MEM-Minimal Essential Medium supplemented with 1% pen/strep, 0.4 mM L-carnitine, and 0.4 % BSA at 37 °C and 5% CO_2. Carnitine and BSA were always added fresh the day of the medium change followed by filtration of the supplemented medium. The sensor cartridge was rehydrate in the XF96 utility plate by adding 200 µl of XF calibrant solution per well. On the day of assay, cells were carefully washed once and incubated in 180 µL KHB assay medium (111 mM NaCl, 4.7 mM KCl, 1.25 mM $CaCl_2$, 2 mM $MgSO_4$, 1.2 mM NaH_2PO_4 supplemented with 2.5 mM glucose, 0.4 mM carnitine, and 5 mM HEPES on the day of the assay, adjusted to pH 7.4 at 37 °C) at 37°C for 30 min prior measurement. After an initial calibration step of 20 min, OCR was measured with no additions in 4 cycle of 2 min mixing and 2 min waiting, followed by 3 cycles of 3 min measuring interspersed by 2 min mixing. The OCR was determined after addition of either Palmitate-BSA conjugate (181 µM pamitate/31 µM FAF-BSA) or BSA (31 µM FAF-BSA) as control loaded in Port A in 8 cycles of 2 min mixing, 2 min waiting, and 3 min measuring; after the addition of ETO (50 µM) loaded in Port B in one cycle of 2 min mixing, 2 min waiting, 3 min measuring, 2 min mixing, 3 min measuring, 2 min mixing, and 3 min measuring.

Following the measurement, cells were washed in PBS and subjected to cell number quantification using the CyQuant Cell Proliferation Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol.

2.2.5. Protein analytics

2.2.5.1. Immunofluorescence

Fibroblasts were plated at a density of 7500 cells/well on permanox slides (Lab-Tek Chamber ®Slides, Sigma-Aldrich Chemie GmbH Munich, Germany) and incubated overnight at 37 °C in 5% CO2. Subsequently the cells were washed with PBS twice and fixed in 4% PFA. For staining the Golgi with anti-Giantin (1:200), the fixed cells were permeabilized with 0.1% NP40 and stained with primary antibodies diluted in 2% BSA. As secondary antibody, anti-

rabbit Alexa Fluor 568 (1:500) was used. For trans-Golgi staining using WGA, the fixed cells were washed in HBSS, labelled with 5.0 µg/mL WGA-AF488 conjugate for 10 minutes, and washed twice in HBSS buffer. For both stainings, slides were mounted in ProLong Antifade Reagent containing DAPI (Invitrogen, Eugene, OR, USA). The Golgi area was analyzed using the image processing package Fiji (Schindelin J et al., 2012).

2.2.5.2. Immunoblotting

Cell lysates were prepared by resuspension of fresh or frozen fibroblast pellets in 100 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 on a rotating wheel for 1h at 4°C and subsequently disrupted by 10 strokes with a 0.30 x 8 mm syringe (Becton, Dickinson and Company, Franklin Lakes, NJ, USA). The mixture was centrifuged for 10 min at 15000 g at 4 °C and whole protein amount of the recovered supernatant was quantified using the Bradford method [110]. 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 proteins per sample were loaded on 4-12% precast gels (Lonza, Basel, Switzerland). Electrophoresis in 1x ProSieve EX Running buffer (Lonza, Basel, Switzerland) was started at 40 V for 30 min and continued at 120 V for 60 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 25V for 10 min. The membranes were blocked in 5% non-fat milk (Bio Rad) in TBS-T (150 mM NaCl, 30 mM Tris base, pH 7.4, 0.1% Tween 20) for 1 h and immunoblotted using primary antibodies (1:1000) for 1 h at RT or ON at 4°C. Signals were detected by incubation with HRP-conjugated goat anti-rabbit and goat anti-mouse secondary antibodies (1:5000, Jackson Immuno Research Laboratories) for 1 h and visualized using ECL (GE Healthcare Life Sciences, Chalfont St. Giles, UK).

2.2.5.3. Blue native PAGE (BN-PAGE)

For each fibroblast cell line, a fresh pellet derived from three 175 cm^2 tissue culture flasks was resuspended in PBS supplemented with 0.25 mM PMSF and 10 U/mL DNAse I followed by solubilisation using 2 mg digitonin/mg protein. The solution was incubated on ice for 15

min upon which 1 mL PBS added. The sample was centrifugation for 10 min at 10000 × g and 4 °C and the pellet was subsequently resuspended in 1x MB (750 mM ε-aminocaproic acid, 50 mM bis-Tris, 0.5 mM EDTA, pH 7.0) followed by whole protein quantification. Solubilisation of membrane proteins was performed using 0.5% (v/v) *n*-dodecyl-β-d-maltoside (DDM) for 1 h on ice at a protein concentration of 2 µg/µL. The mixture was centrifuged for 30 min at 10000 g at 4°C and whole protein amount of the recovered supernatant was quantified. SBG was added to a final concentration of 0.25% (v/v). 60 µg protein was loaded on NativePAGE 4-16% Bis-Tris gels (Thermo Fisher Scientific, Waltham, MA, USA). Anode buffer was comprised of 50 mM Bis-Tris, pH 7.0, blue cathode buffer was comprised of 15 mM Bis-Tris, 50 mM Tricine, pH 7.0, 0.02% SBG. Electrophoresis was started at 40 V for 30 min and followed by 130 V until the front line proceeded 2/3 of the gel. Thereafter, the blue cathode buffer was replaced by clear cathode buffer not containing SBG (15 mM Bis-Tris, 50 mM tricine, pH 7.0). Proteins were wet transferred to PVDF membranes and immunoblotted with primary antibodies (1:1000) against NDUFB8 (complex I) and UQCRC2 (complex III). The blots were incubated with HRP-conjugated goat anti-rabbit and goat anti-mouse secondary antibodies (1:5000, Jackson Immuno Research Laboratories) for 1 h and the signal was visualized using ECL.

2.2.5.4. Mass spectrometry

In parallel to RNA sequencing, a second aliquot of the fibroblast cell pellet was sent for quantitative mass spectrometry performed by Garwin Pichler and Matthias Mann at the Max-Planck Institute of Biochemistry, Martinsried. Sample preparation, data acquisition and processing have been described in detail previously [107]. Label free quantification (LFQ) values were visualized using Perseus (version Perseus_1.5.4.1.) [106]. Proteins which were not covered in at least 50% of the samples were discarded, hierarchical clustering was performed, and results were inspected manually. This set-up was used to generate and optimize algorithms allowing a systematic analysis performed by the group of Julien Gagneur as described in detail previously [107].

2.2.6. Metabolomics

Plasma samples of 143 patients with suspected mitochondrial disease as well as 97 age-matched controls were collected and prepared for a non-targeted metabolomics experiment conducted by the Genome Analysis Center of the Helmholtz Zentrum München (GAC). A

metabolomics platform based on mass spectrometry coupled to liquid chromatography established by Metabolon Inc. was used. Analytical protocols, identification of metabolites, and processing of the raw ion counts have been described in detail previously [107].

3. Results

3.1. Summary of publication 1

Severe respiratory complex III defect prevents liver adaptation to prolonged fasting

Kremer LS, L'hermitte-Stead C, Lesimple P, Gilleron M, Filaut S, Jardel C, Haack TB, Strom TM, Meitinger T, Azzouz H, Tebib N, Ogier de Baulny H, Touati G, Prokisch H, Lombès A. J Hepatol. 2016 Aug;65(2):377-85. doi: 10.1016/j.jhep.2016.04.017. Epub 2016 May 2. PubMed PMID: 27151179.

Through analysis of WES data of a patient suffering from recurrent metabolic crisis leading to death at 20 months of age homozygous variants in two genes encoding mitochondrial proteins, *MTO1* and *LYRM7*, were detected. MTO1 is crucial for proper mitochondrial translation and patients with *MTO1* mutations present with hypertrophic cardiomyopathy, lactic acidosis, and a combined RCCI and RCCIV defect. LYRM7 is a RCCIII assembly factor and at the time of the study only one patient harboring variants in LYRM7 was reported suggesting an association with severe encephalopathy. In contrast, our patient showed strong involvement of the liver. Since an evaluation of the pathological relevance of the variants based on the clinical symptoms was not elusive, we investigated the role of the variants using functional validation assays.

As the nonsense mutation in *MTO1* was predicted to result in removing 27 amino acids of the full length 692 amino acids and the frameshift variant in *LYRM7* was predicted to result in a severe truncation of the protein to only 28 amino acids, we first investigated the expression of the proteins by western blot analysis. A complete loss of LYRM7 in the patient derived fibroblasts was apparent, while MTO1 levels were normal. Whereas this clearly indicates a LYRM7 deficiency, normal levels of MTO1 do not exclude impairment of MTO1 function. To probe effects of the variants on OXPHOS, I generated two lentiviral expression vectors encoding wild-type *LYRM7* and wild-type *MTO1* respectively using two different selection markers. By lentiviral transduction, I subsequently generated patient and control cell lines expressing wild-type LYRM7, MTO1, and both in combination. By accessing the overall cellular oxygen consumption, I could not clarify the implications of the variants. Measurement of isolated respiratory chain complex III (RCCIII) activity performed by our collaborators showed rescue of the RCCIII defect present in the patient fibroblasts upon

LYRM7 transduction, no complementing effect was seen upon *MTO1* transduction. This clearly demonstrates the pathogenicity of the variants detected in LYRM7. Biochemical measurements revealed slightly decreased RCCIV activity in the patient fibroblasts which was not rescued upon *MTO1* transduction and could hence be an artefact. Together with the normal RCCI activity, this indicates that the variant in *MTO1* most likely does not have a negative impact on protein function and is benign.

In conclusion, we present a strategy to evaluate the causality of variants aiding to distinguish whether a patient possibly suffers from two distinct or overlapping diseases. For our patient, we could show a clear link between the *LYRM7* mutation and the RCCIII deficiency. In contrast, we could not detect any effect of the *MTO1* variant in patient derived fibroblasts. However, I cannot fully exclude a negative effect of the *MTO1* mutation as the effect might be tissue specific or develop at a later age.

Personal contribution:

• Experimental design of experiments (25%) together with co-authors.

• Generation of patient cell lines expressing wild-type cDNAs (100%).

• Data analysis and interpretation (25%) together with co-authors.

• Drafting and revising of the manuscript (25%) together with co-authors.

3.2. Summary of publication 2

Biallelic Truncating Mutations in *TANGO2* Cause Infancy-Onset Recurrent Metabolic Crises with Encephalocardiomyopathy

Kremer LS, Distelmaier F, Alhaddad B, Hempel M, Iuso A, Küpper C, Mühlhausen C, Kovacs-Nagy R, Satanovskij R, Graf E, Berutti R, Eckstein G, Durbin R, Sauer S, Hoffmann GF, Strom TM, Santer R, Meitinger T, Klopstock T, Prokisch H, Haack TB. Am J Hum Genet. 2016 Jan 19. pii: S0002-9297(15)00504-2. doi: 10.1016/j.ajhg.2015.12.009. [Epub ahead of print] PubMed PMID: 26805782.

As part of the Transnational Access project of the European Sequencing and Genotyping Infrastructure (ESGI), I selected and prepared DNA samples of 125 patients with suspected mitochondrial disorders for WES and subsequently analyzed the data. In one of these patients WES analysis yielded 14 genes carrying rare bi-allelic variants. None of the 14 genes was reported to be associated with any disease or predicted to encode a mitochondrial protein. By a joined analysis with other suspected mitochondrial disorder cases, we identified two additional patients with rare bi-allelic variants in one of these 14 genes, *TANGO2*. Strikingly, all 3 patients showed a consistent phenotype and metabolic findings of ketonuria and lactic acidosis hinting at impaired ß-oxidation.

Little was known about TANGO2. A role of TANGO2 in the distribution of the Golgi membrane was proposed based on genome-wide RNA-mediated interference screen in a *Drosophila* cell line. Immunostaining in several studies suggested localization in the cytosol, the Golgi, or the mitochondria. To ascertain the localization of the endogenous TANGO2, I performed cell fractionation followed by immunostaining of TANGO2 in fibroblast cell lines from patients and controls. Due to insufficient antibody functionality, these experiments were inconclusive. I subsequently investigated the postulated role of TANGO2 in the distribution of the Golgi membrane. Immunostaining the Golgi using Giantin and wheat germ agglutinin (WGA) did however not reveal a difference in organization of the Golgi between patient and control cells. As the patients presented with a metabolic profile indicating an impaired, I established and employed a ß-oxidation assay which supported the metabolic suspicion of a defect in ß-oxidation. However, more studies are needed to pinpoint the precise role of TANGO2 in ß-oxidation.

The diagnosis of ß-oxidation defects, depicting non-classical, secondary mitochondrial disorders, has strong implications for possible therapeutic options. For patients suffering from classical mitochondrial disorders a ketogenic diet, which is low on carbohydrates but rich on fats, can be beneficial as their pyruvate oxidation is disturbed. In contrast, patients with defect in ß-oxidation should receive nutrition low fats and rich on carbohydrates.

Personal contribution:

• Experimental design of experiments (65%) together with co-authors.

• Localization experiments of TANGO2 (100%).

• Golgi immunostaining (75%) together with co-authors.

• ß-oxidation measurements (100%).

• Data analysis and interpretation (50%) together with co-authors.

• Drafting and revising of the manuscript (25%) together with co-authors.

3.3. Summary of publication 3

Genetic diagnosis of Mendelian disorders via RNA sequencing

Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, Haack TB, Graf E, Schwarzmayr T, Terrile C, Konarikova E, Repp B, Kastenmüller G, Adamski J, Lichtner P, Leonhardt C, Funalot B, Donati A, Tiranti V, Lombes A, Jardel C, Gläser D, Taylor RW, Ghezzi D, Mayr JA, Rötig A, Freisinger P, Distelmaier F, Strom TM, Meitinger T, Gagneur J, Prokisch H (2017). Nat Commun. 2017 Jun 12;8:15824. doi: 10.1038/ncomms15824. PubMed PMID: 28604674.

In accordance with published data, our WES analysis successfully identifies the causative variant in about half of the suspected mitochondrial disease cases [17, 56, 97, 98, 111]. By implication, half of the patients do not receive a genetic diagnosis after WES. We reasoned that the inconclusive WES can be attributed to the incomplete capture of variants, especially non-coding variants, or the failure to prioritize the causative variants. Whereas the former can be overcome by whole genome sequencing (WGS), the vast number of variants generated by WGS and the poor understanding of the non-coding genome further obscure the prioritization of the causative variant. RNA sequencing (RNA-seq), in turn, might ease the prioritization of variants by unravelling their effects on RNA abundance and sequence. I therefore selected 105 patient derived fibroblast cell lines from solved and unsolved cases for RNA-seq. By manual inspection of the RNA-seq data I could detect aberrant expression of the RCCI assembly factor *TIMMDC1*, a gene previously not annotated with disease risk. I subsequently identified a deep intronic variant probably creating a novel splice site resulting in aberrant splicing and finally nonsense-mediated decay (NMD). Using BN-PAGE, I identified a defect in RCCI assembly in patient derived fibroblasts. Quantitative proteomics further confirmed a loss of TIMMDC1 and degradation of RCCI subunits due to the assembly defect. By western blot and BN-PAGE analysis, I demonstrated rescue of the assembly defect upon overexpression of wt-*TIMMDC1* cDNA. I subsequently identified 3 additional patients harboring the same variant presenting with an overlapping phenotype. I thus established *TIMMDC1* as a novel disease-associated gene. My findings guided a systematic analysis performed in close collaboration with computational biology department of the Technische Universität München. We detected a median of one aberrantly expressed gene, five aberrant splicing events, and six mono-allelically expressed rare variants. This small number of events

allowed manual inspection and validation providing a diagnosis for 10% (5 of 48) of previously unsolved cases. Importantly, our pipeline enabled the identification of causative non-coding variants in genes not previously associated with any disease as well as identification of causative variants in genes implicated in diseases other than mitochondrial disorders.

Personal contribution:

• Design of experiments (30%) together with co-authors.

• Growth and preparation of samples for RNA-seq, WGS, quantitative proteomics, metabolomics (100%).

• Prioritization of pathogenic variants using the RNA-seq and WES pipeline (90%) together with co-authors.

• Functional validation of disease-association of *TIMMDC1* (100%).

• Functional validation of variants in *ALDH18A1* and *CLPP* (70%) together with co-authors.

• Drafting and revising of the manuscript (30%) together with co-authors.

4. Discussion

Mitochondrial disorders are an extremely heterogeneous group of inborn errors of metabolism presenting with any combination of affected tissues and more than 250 identified disease-associated genes up to now comprising more than 100 discoveries in the NGS area, many more awaiting identification [11]. This broad clinical and genetic presentation of mitochondrial disorders makes their molecular diagnosis extremely difficult. Even though only in a few cases unraveling of the molecular cause allows therapeutic intervention, the identification of the underlying mutation is none the less crucial as it does not only provide the patient with a molecular diagnosis, but also is fundamental for genetic counseling and prenatal testing. Moreover, it renders additional, possibly more invasive testing redundant, might entitle for healthcare benefits, and might allow patient stratification which is the prerequisite for treatment trials and hence development of new therapeutic strategies [62].

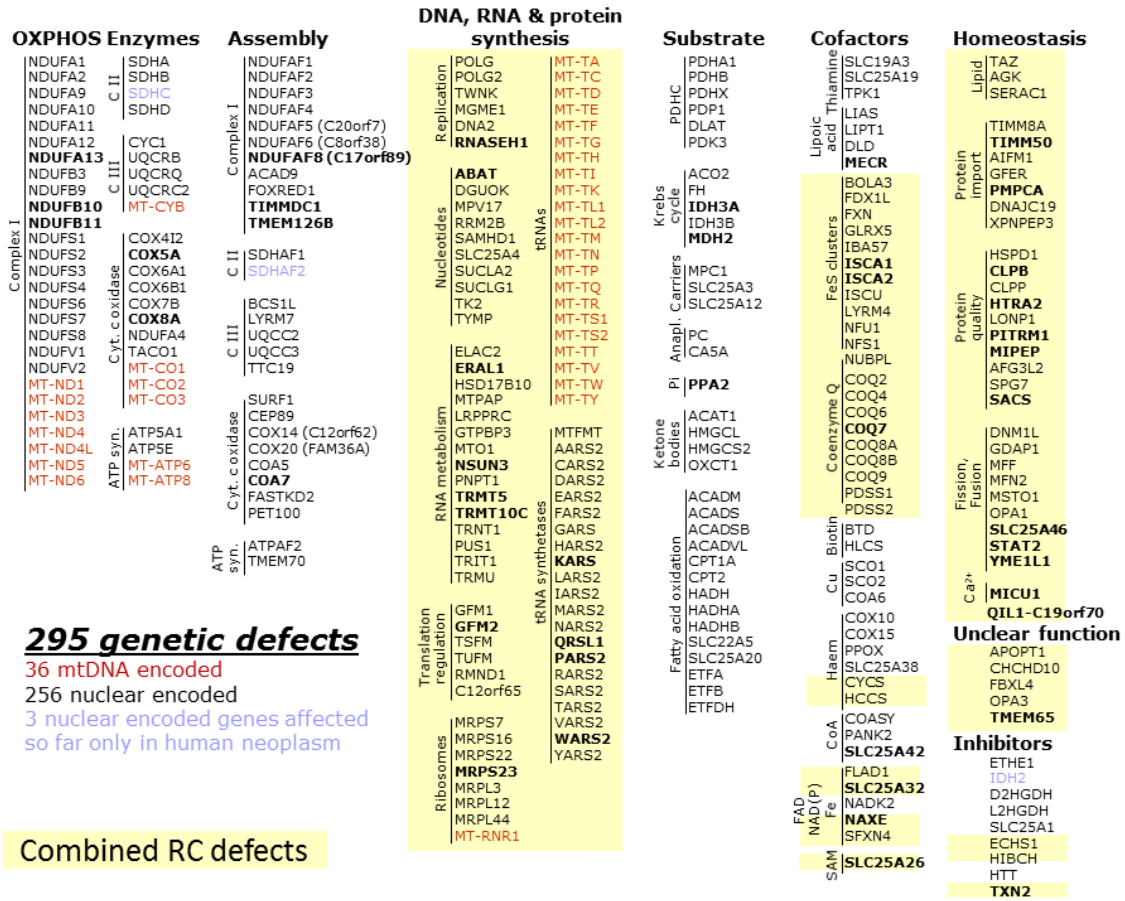

Figure 5 Mitochondrial disease-associated genes. Genes are grouped according to their implementation in mitochondria processes. Disease-associated genes identified since the report Mayr et al., [11] in 2015 are highlighted in bold. Figure is a courtesy of Dr. Johannes Mayr.

Diagnosis based on biochemical and metabolic investigations are often only specific for certain subgroups (see 1.1). WES, in contrast, has proven to be a comprehensive tool for the diagnosis of the overall group of mitochondrial disorders. WES is therefore not anymore used for research purposes only, but is on the rise in clinical diagnostics [56]. In the course of this PhD, WES was successfully applied to identify novel variants in genes previously associated with mitochondrial disease as well as genes not associated with disease. Two examples are discussed in 4.1 and 4.2.

Despite the tremendous impact of WES on molecular diagnosis, the diagnostic yield of WES analysis is far from complete. In the case of suspected mitochondrial disease patients, the causative variant is only revealed in only about 50% of the cases [56, 97]. This largely reflects the outcome at our institute. To scrutinize and elucidate this missing heritability and to develop strategies to improve the diagnostic outcome, shortages of WES and the usability of RNA-Seq to complement WES are discussed in 4.3.

4.1. Validating new variants in known disease-associated genes

We investigated the molecular basis of disease in the following individual: From the age of 6 months on, the patient suffered from recurrent metabolic crisis presenting with lactic acidosis and acute liver dysfunction [112]. At one year of age, she lapsed into coma during a severe episode lasting 3 days accompanied by irreversible neurological damage. She died at 20 months of age. Given the severe course and the presence of lactic acidosis, a mitochondrial disorder was suspected. A clear RCCIII defect was evident in biopsies from liver and fibroblasts of the patient. WES was performed and very stringent filtering for rare (0.1% of 5000 in-house controls) recessive variants revealed 22 candidate genes. Amongst these, 2 genes, *LYRM7* and *MTO1*, both harboring novel homozygous variants, were reported as mitochondrial disease-associated genes. The variant detected in *LYRM7* was not reported in ExAC, the variant in *MTO1* was detected in a heterozygous state in 11 out of 121348 individuals.

LYRM7 is a RCCIII assembly factor and 8 other patients harboring pathogenic variants in LYRM7 were reported by March 2017 [113, 114]. All patients for whom a muscle or skin biopsy was available presented with a reduced activity of RCCIII. Most of these patients suffered a neurological deterioration in the first years of life followed by recurrent episodes of encephalopathy and lactic acidosis. The episodes were partially accompanied by coma and led

to severe neurological damage or death. Importantly, MRI in all patients reported by Dallabona et al. showed a consistent and distinctive pattern of abnormalities [114].

In turn, pathogenic variants in the Mitochondrial Translation Optimization Factor 1 (*MTO1*) have been described in more than 10 patients [97, 115-118]. MTO1 is a mitochondrial, tRNA modifying enzyme that is crucial for proper mitochondrial translation [119]. Most patients with *MTO1* mutations present with hypertrophic cardiomyopathy, lactic acidosis, and a combined RCCI and RCCIV defect. Disease onset at birth was shown to result in death in the first year of life, while patients with disease onset in the first months or even years of live were reported to be alive, the oldest patient being 30 years [117, 118]. A mouse model of MTO1 deficiency resembles the cardiomyopathy and lactic acidosis seen in *MTO1* patients [120].

Considering the described clinical symptoms for *LYRM7* and *MTO1* patients, the clinical presentation of our patient closely resembled the clinical presentation of the *LYRM7* patients. As the *LYRM7* patients, our patient suffered from recurrent neurological crisis resulting in coma and neurological damage. In contrast to the *MTO1* patients, our patient did not show any heart involvement. Furthermore, biochemical analysis in liver and skin biopsies revealed an isolated RCCIII defect, which was also described for the published *LYRM7* patients. The activity of RCCI was normal and the activity of RCCIV only slightly decreased. While this clinical and biochemical data strongly hints at the variants in *LYRM7* to be causative, the pathogenicity of the VUS in *MTO1* cannot be excluded. Disease onset for *MTO1* patients can vary. It is therefore possible that due to the patient`s early death the MTO1-related symptoms were not yet evident. Consequently, we sought to functional investigate the pathogenicity of the novel variants in both, *LYRM7* and *MTO1*.

The homozygous frameshift variant in *LYRM7* results in a severe truncation of the protein to only 28 amino acids instead of the original 104 amino acids. Out of the 28 remaining amino acids only the first 17 are unchanged. The homozygous nonsense mutation in *MTO1* results in a stop codon at position 666 of the amino acid chain, removing 27 amino acids of the full length 692 amino acids. RNA-seq analysis revealed normal transcript levels of LYRM7 and MTO1. Western blot analysis in patient fibroblasts revealed complete loss of LYRM7, while MTO1 levels were normal. While this clearly indicated a LYRM7 deficiency, normal levels of MTO1 do not exclude impairment of MTO1 function. I therefore transduced the patient

fibroblast cell line with wild-type copies of *LYRM7* and *MTO1* and investigated the effect on the observed RCCIII defect. Using cellular oxygen consumption measurements, I could not distinguish impairment of RCCIII and RCCIV. Measurement of isolated RCC activity by our collaborators demonstrated a rescue of the RCCIII defect upon *LYRM7* transduction, no complementing effect was seen upon MTO1 transduction. This clearly demonstrated the pathogenicity of the variants detected in LYRM7. The slightly decreased RCCIV activity present in the patient fibroblasts, in contrast, was not rescued upon *MTO1* transduction and could hence be an artefact. Together with the normal RCCI activity, this indicated that the variant in *MTO1* most likely does not have a negative impact on protein function and is benign. However, this analysis was performed in fibroblast cell line, while mutations in *MTO1* seem to mainly affect the heart. To fully exclude any tissue dependent effect masking the impact of the *MTO1* variant, one would need to conduct these experiments in heart tissue of the patient. This is limited by availability of heart tissue and technical difficulties in performing the transfection procedure. Concordantly, impact of the *MTO1* variant on the patient's phenotype cannot fully be ruled out.

From the 8 published *LYRM7* patients, only for 2 patients liver abnormalities were reported, where 1 patient had an increased liver volume and 1 patient had elevated ammonia [113, 114]. In contrast, liver failure is frequently observed in childhood-onset mitochondrial diseases. Combined RCC deficiency is often seen in cases of globally changed liver function, whereas isolated RCC deficiencies, mainly RCCIII deficiency, is associated with distinct liver alterations [112]. In our patient, the presence of ketonuria despite insulin injection clearly hinted at impaired function of the RC in liver. Liver enzymes and coagulation factors, in contrast, were normal during the metabolic crisis, contradicting a liver failure. The metabolic crisis seemed to be induced by fasting. As described in 1.1.2, during fasting the TCA cycle is halted and electrons are derived from ß-oxidation and gluconeogenesis. This metabolic switch seems trigger the crisis. It is now tempting to speculate that nutritional care could prevent such crisis and hence the neuronal damage. This would be the case if the liver abnormalities are directly involved in the harming of the brain. However, normal brain function is reported for other CIII deficiency patients presenting with liver abnormalities arguing for a distinct role of *LYRM7* in the brain [112].

The case presented here exemplifies situations where WES identifies variants in more than one likely disease-associated gene. To evaluate the causality of the variants and to distinguish whether the patient possibly suffers from two distinct or overlapping diseases, careful validation of the pathogenicity of the variants is required. Posey et al. report 4.9% of their patients to show phenotypes attributed to multiple loci [121]. Their approach relies on the dissection of the clinical presentation based on clinical ontologies. This, however, requires a well-established genotype-phenotype correlation which is not readily available for rare disorders where only a few or even solely one patient is reported as was the case for *LYRM7*. In these cases, functional assays are absolutely essential.

This study describes how to validate novel variants in known and suspected disease-associated genes by performing re-expression of the wild-type copy of a gene in patient derived fibroblasts. Prerequisite of the complementation assay using the wild-type copy of the gene is the presence of a phenotype whose rescue can be used as a read-out. As a first tier, such phenotype can be depicted by a defect in mitochondrial respiration detected by decreased oxygen consumption. Rescue experiments assessing oxygen consumption are routinely conducted in our laboratory [108, 122]. However, a defect in mitochondrial respiration is only apparent in about two thirds of the more than 200 patient fibroblast cells I measured during the course of this PhD (data not shown) potentially due to tissue-specific manifestation of the defect. Furthermore, measurement of the mitochondrial respiration depicts a very general phenotype. For example, for the herein described patient, measurement of the overall oxygen consumption did not allow to distinguish between a defect in RCCIII or RCIV and was therefore not conclusive (data not shown). Careful evaluation of the specificity of an observed rescue is mandatory. Specific rescue of a decreased mitochondrial respiration can be ascertained by including control cell lines not harboring mutations in the gene under investigation. In the control cell lines, overexpression of the respective gene should not have a tremendous effect. Yet, rescue experiments provide stronger evidence if a unique phenotype is investigated as for example assembly of RCCI in the case of the *TIMMDC1* patients. This is also applicable for cases where overall decrease of mitochondrial respiration was not detectable whereas more subtle changes might still be evident. For such tailored rescue experiments the prior understanding of the protein function is crucial. Especially for new disease-associated gene discoveries however, the precise role of the encoded protein might not yet be understood. This lack of knowledge in combination with absence of a more general

phenotype call for an alternative strategy to validate new disease-associated genes as eluted to in the following chapter.

4.2. Validating genes not previously associated with disease

WES was performed in a patient presenting with episodes of impaired gait and muscle weakness accompanied by ketonuria from the age of 3 years on [123]. She developed cognitive and sensorineural hearing impairment. The episodes gradually worsened and with the age of 17 years muscle pain, rhabdomyolysis, and elevated plasma CK activities were present. Brain atrophy was evident in MRI taken at 15 years of age. Arrhythmia was detected from the age of 17 years on. Metabolic investigations were largely normal, besides repeatedly high ketone body levels. A slight elevation of the acylcarnitine C5OH was detected at the age of 15 years. RCC activities in muscle were normal. Due to the ketonuria, a defect in ß-oxidation was suspected. WES analysis yielded 14 genes carrying rare (0.1% of 5000 in-house controls) bi-allelic variants. None of the 14 genes was reported to be associated with any disease or predicted to encode a mitochondrial protein. For one of these genes, *TANGO2*, in another patient a homozygous stop variant had previously been identified but the causality remained unclear due to lack of further evidence. In the patient described here a homozygous CNV comprising a 34.6 kb deletion in *TANGO2* was detected. The frequency of this deletion in public databases could not be assessed due to limited annotation. In our in-house database, at that time containing 5300 individuals, the CNV was also seen in a heterozygous state 7 times and in a compound heterozygous state with a frameshift mutation in a third suspected mitochondrial disease patient. Strikingly, all 3 patients showed a consistent phenotype characterized by infancy-onset episodes of metabolic crisis, rhabdomyolysis, encephalopathy, brain atrophy, and arrhythmias. In all 3 patients, the metabolic finding of ketonuria and lactic acidosis hinted at impaired ß-oxidation.

Little is known about TANGO2. Depletion of *TANGO2* as part of a genome-wide RNA-mediated interference screen in a *Drosophila* cell line caused a fusion of the Golgi apparatus with the endoplasmatic reticulum (ER) [124]. Therefore, a role of TANGO2 in the distribution of the Golgi membrane was proposed. Immunostaining of V5-tagged TANGO2 suggested localization in the cytosol and the Golgi. In another study, bioinformatics prediction of localization and immunostaining using EGFP-tagged fusion proteins propose a mitochondrial localization of TANGO2 [125]. To study the localization of the endogenous

TANGO2, I therefore performed cell fractionation followed by immunostaining of TANGO2 in fibroblast cell lines from patients and controls. However, I was not able to detect TANGO2 in any fraction of any sample using 2 different antibodies, indicating insufficient antibody functionality. As the localization experiment failed, I investigated the postulated role of TANGO2 in the distribution of the Golgi membrane. Immunostaining the Golgi using Giantin and wheat germ agglutinin (WGA) did however not reveal a difference in organization of the Golgi between patient and control cells. As the patients presented with a metabolic profile indicating an impaired ß-oxidation, I measured the palmitate-dependent oxygen consumption rate as a read-out for the integrity of the mitochondrial respiration using the Seahorse XF96 Extracellular Flux Analyzer. Compared to control cell lines, patient cell lines showed a decreased oxygen consumption rate upon palmitate supply supporting the clinical suspicion of a defect in ß-oxidation. However, more studies are needed to ensure and pinpoint this presumed role of TANGO2.

Critical proof of the causality of the variants found in *TANGO2* is given by the repeated finding of the disease-associated gene in unrelated individuals sharing the same clinical presentation [82]. Simultaneous with our report, a second study provided further evidence as they identified 12 additional patients with bi-allelic mutations in *TANGO2* [126]. Importantly, the clinical presentation comprising episodic metabolic crisis with muscle weakness, rhabdomyolysis, brain abnormalities, and arrhythmias fully recapitulated the findings in our patients. Moreover, the CNV detected in our patients in a homozygous and a compound heterozygous state respectively, was also evident in in a homozygous state in 4 and in a compound heterozygous state in 1 patient from the second report. As the investigated subjects in both studies were of European origin, this CNV might depict a founder mutation. The CNV has been missed previously due to difficulties in the reliable detection of CNVs (see 1.2.2) while in this study the implementation of ExomDepth in the analysis pipeline allowed improved CNV calling.

The diagnosis of ß-oxidation defects, depicting non-classical mitochondrial disorders as OXPHOS is not directly impaired, has strong implications for possible therapeutic options. Patients suffering from classical mitochondrial disorders can potentially benefit from a ketogenic diet, which is low on carbohydrates but rich on fats, as their glucose metabolism is disturbed. This diet would however be harmful for patients with a defect in ß-oxidation as in

these patients the metabolism of fatty acids is hampered. The distinction of ß-oxidation defects is therefore crucial.

In conclusion, this study highlights how a gene previously not described as disease-associated can be validated by replication of the finding in additional independent cases presenting with identical symptoms. However, as presented above, genotype-phenotype correlation for mitochondrial disorders haven been shown to be poor with mutations in the same gene giving rise to different symptoms. Along with the rare occurrence with mutations in mitochondrial disease-associated genes, identification of additional independent patients might be hampered. If additional patients exist, national and international networks and registries of mitochondrial disease patients help their identification and joined investigation. Identification of the underlying molecular cause and the affected metabolic pathway can have tremendous effects on the benefit of possible treatment options.

4.3. Shortages of WES analysis

In accordance with published data, our WES analysis successfully identifies the causative variant in about 50% of the suspected mitochondrial disease cases [17, 56, 97, 98, 111]. Hence, 50% of the patients still lack a molecular diagnosis after WES analysis. Possible reasons for this missing heritability in cases of genetic disorders can be grouped into two categories: either the causative variant was not captured by WES due to technical limitations of the technique, or the variant was detected but not prioritized due to analytical failure.

4.3.1. Technical limitations

Even though WES is, in theory, the sequencing of the complete coding region of the genome, we do not fully understand the coding genome yet and the list of coding genes is not definite. The baits employed in the enrichment kits can therefore only target regions which have already been identified as exonic regions and regions not yet identified as exonic are hence missed.

Furthermore, baits capture target sequences with varying efficiency, with some sequences escaping capture entirely. Besides, efficiency of PCR amplification and enrichment of the fragments can vary highly, e.g. due to GC-rich regions, hence resulting in differences in the coverage of fragments. Even though the manufacturers of the exome enrichment kits steadily try to improve these shortcomings, the kits still fail to completely capture the targeted regions.

In this study, various versions of the Agilent Sure Select kit were used according to their release. The first kit targeted 38 Mb of sequence with 81.5% of the targeted bases covered at least 20x, the latest kit targets 50 Mb sequence with 95.6% of the targeted bases covered at least 20x [81]. This means that in the samples sequenced by us with the first kit almost 20% of the targeted coding sequence was not well covered. Amongst this 20% is, for example, the mitochondrial methionyl-tRNA formyl-transferase (*MTFMT*). *MTFMT* was not covered at all in the 38 Mb kit, in the 50 Mb kit all but exon 1 were covered sufficiently. In a patient sequenced using the 50 Mb kit, one heterozygous variant in *MTFMT* was found and re-sequencing of exon 1 with Sanger sequencing could identify the second heterozygous variant [127]. In the light of the reporting of incidental findings, the American College of Medical Genetics and Genomics (ACMG) investigated the performance of clinical exome sequencing [128]. They found that out of the 56 investigated actionable disease genes, in 6 different genes one exon was inadequately covered. Furthermore, they discovered insufficient coverage of more than half of the HGMD variant locations in 7 genes. In another study, Klein and colleagues showed that of the exons of genes known to be associated with neuropathy, 7% were covered insufficiently and 2% completely missed [129]. The lack of complete capture is therefore a drawback in WES analysis.

Finally, as WES is the sequencing of exonic regions, it is completely blind to variants in non-coding regions. Even though 85% of the Mendelian disease-causing variants are predicted to reside in the exome, this estimation might be biased by the poor understanding of the non-coding genome and the proceeding incapability of identifying causative variants therein.

Speculatively, even though recognized as a Mendelian disorder, it is impossible to rule out cases of bi- or polygenic inheritance as well as epigenetic mechanisms contributing to the occurrence of mitochondrial disorders. To my awareness, no such cases have been reported yet.

4.3.2. Analytical limitations

After the capture of a variant, sophisticated bioinformatic steps are required for its reliable alignment, calling, annotation, and prioritization. These steps are dependent on statistical and predictive tools and can hence be incorrect and sensitive to errors, especially in regions of repeats and homopolymers.

One error-prone process is the calling of indels. While the calling of one base pair substitutions is more or less trivial as only the base at the very position must be considered, the calling of indels requires joint analysis of the adjacent base pairs. This requirement posed a problem to previous calling tools which was resolved by the recently developed GATK HaplotypeCaller [65, 66]. The superiority of the GATK HaplotypeCaller is based on the employed *de novo* assembly of regions varying from a given reference. Re-analysis using GATK HaplotypeCaller of data previously analyzed using SAMtools could therefore reveal pathogenic indels in unsolved cases.

A further hurdle in WES analysis is the detection of CNVs, which are defined as deletions or duplications with a size of more than 1 kb. CNVs can result in altered gene dosage due to gene disruption or interference with regulatory sequences [130]. In the recent years, the association of CNVs with disease has become incrementally evident, e.g. for intellectual disability [131], cancer [132], or schizophrenia [133]. According to estimates, CNVs contribute to approximately 1% of sequence variation of a personal genome and make up for more base pair variations than SNVs [130, 134, 135]. CNVs are detected by computing the read depth at a given position. However, for WES data, differences in read depth can also be attributed to uneven coverage due to technical limitations rendering reliable detection of CNVs difficult [130]. Identification of CNVs in repetitive genomic regions is further complicated [136]. ExomDepth was reported to be the most sensitive tool for CNV detection and its usefulness in our hands was demonstrated by the successful identification of CNVs in *TANGO2* [67, 68, 123]. For data previously not analyzed using ExomDepth, CNVs might be missed and the analysis should be repeated accordingly.

The annotation of variants in regard to their genomic location (e.g. UTR, exonic, splice site, etc.) and their functional effect (e.g. synonymous, frameshift etc.) is based on predictions build upon given gene and transcript definition files. While the information in these files reflects the current understanding of genes and transcripts, the knowledge might not be final. In our RNA-seq study, we could demonstrate predicted missense mutations to be splice site mutations for *CLPP* and *MCOLN1* and therefore highlight how VUS can contribute to inconclusive WES analysis.

Misleading prioritization of variants can be due to a manifold of reasons. In this study, we did not perform trio sequencing which is however the prerequisite for the prioritization of

previously unknown *de novo* mutations. *De novo* mutations are the most deleterious form of variants as they have not been subjected to evolutionary selection [137]. Based on estimations from WGS studies on average 44 to 82 SNVs occur *de novo* in an individual´s genome out of which 1 to 2 affect the coding region [138]. Causative *de novo* variants identified by WES were first reported in 2010 and 2011 for malformation syndromes [139-141] but have since then been implicated in many other genetic diseases, most strikingly neurodevelopmental disorders [137, 142]. In 2014, Lee et al. performed clinical exome sequencing on trios of 410 patients with rare Mendelian disorders providing a diagnosis to 127 of these patients [143]. In 50% of the diagnosed cases the causative variant was *de novo*, while in 20% the variants were compound heterozygous and in 16% homozygous. In the same journal issue, Yang reported on diagnosing 504 patients with suspected genetic disorders by clinical exome sequencing out of which 208 cases were diagnosed with *de novo* mutations [144]. Concordantly, d*e novo* mutations were also found as the most frequent cause in a clinical exome study conducted in 2016 by Posey and coworkers [145]. D*e novo* mutations might therefore account for a considerable amount of our inconclusive WES cases. As *de novo* mutations, dominant mutations are difficult to prioritize. The recognition of dominant mutations is tremendously facilitated by sequencing of affected distant family members which was not performed in our study. As dominant mutations are frequently identified as the molecular cause in adult onset mitochondrial disorders, we must assume that dominant mutations might be the molecular grounds for certain adult onset unsolved cases. The contribution of dominant mutations on infantile-onset patients should be neglectable for cases with healthy parents. Prioritization is further hampered with increasing numbers of candidates and the presence of VUS as outlined above. Additionally, prioritizing genes based on the assumption of certain disease can prevent diagnosis as clinical spectra can overlap or the prevalent symptoms can be secondary. Especially in the case of mitochondrial disorders, non-mitochondrial protein encoding genes can cause primary mitochondrial defects as well as giving rising to unexpected secondary respiratory chain deficiencies which might impede variant prioritization [87].

4.4. Developing complementing strategies for inconclusive WES analysis

4.4.1. WGS

Molecular diagnosis using WES analysis fails in cases where the causative variant was not captured or not prioritized as outlined above. Insufficient capture can be resolved by

performing WGS. WGS, in contrast to WES, does not depend on an enrichment step and PCR-free WGS is much less biased due to differences in the GC content. WGS therefore provides nearby complete coverage of the entire genome. This not only allows detection of variants in non-coding regions but eases detection of structural variants as CNVs or inversions [52]. A recent comparison between WES and WGS by Meienberg et al. revealed that WGS completely covered the ACMG genes where WES only completely covered 75.56% of the genes and that WGS provided 100% coverage of the disease-causing mutations listed in HGMD whereas WES only covered 98.22% [146]. As sequencing costs further decrease, WGS will replace WES in the near future. Despite WGS allegedly providing the whole picture, the absence of the great success of WGS in identifying the molecular cause for the missing heritability cases with clear genetic disorders can most likely be attributed to the shear plethora of variants encountered during WGS combined with lack of knowledge on how to interpret the non-coding genome. WGS on average reveals between 3 and 5 million variants per sample mostly comprised on SNVs and small indels but also comprising around 2,500 structural variants [147, 148]. Analysis of WGS data from 32 of our unsolved WES cases did not result in a diagnosis so far. However, the analysis was focused on the coding region and CNVs only.

4.4.2. RNA-seq

To ease prioritization of candidates obtained by either WES or WGS, we pursued a complementing approach, RNA-seq [107]. While WES and WGS provide purely genetic information, subsequent RNA-seq allows direct probing of detected variants for their effect on RNA abundance and RNA sequence as depicted in Figure 6.

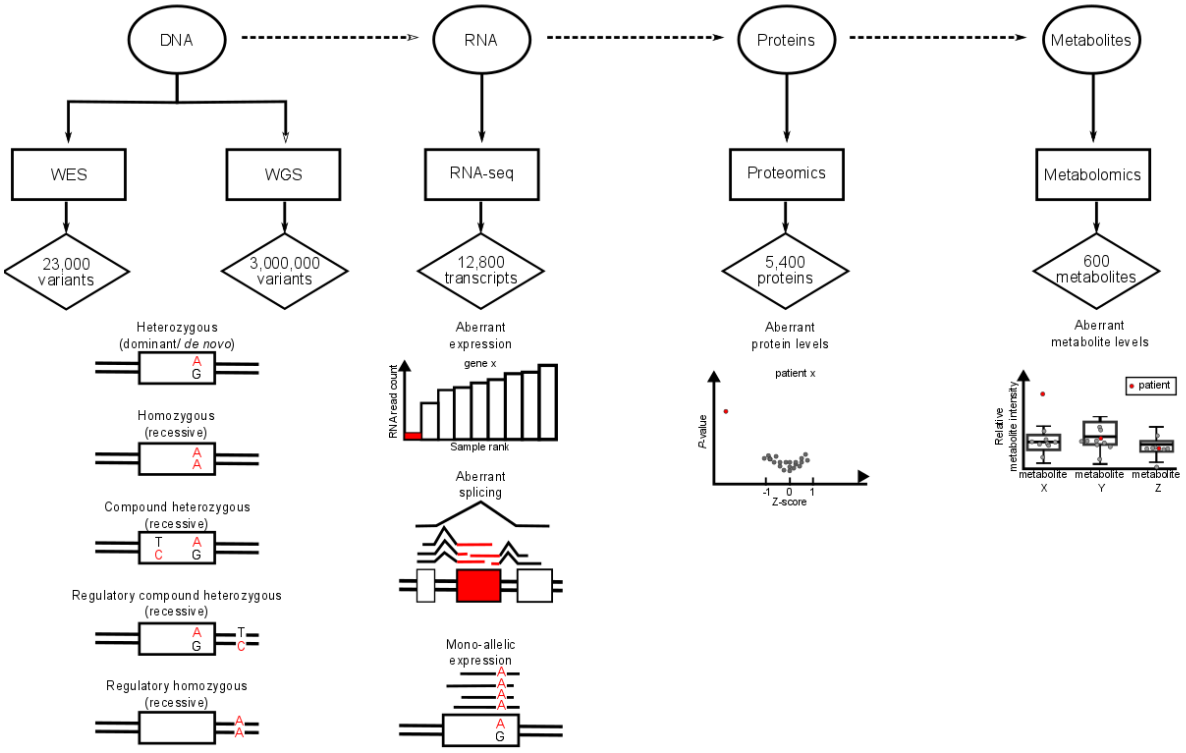

Figure 6 Rational of our approach based on the central dogma of molecular biology. Genetic information is passed onto distinct biopolymers and biochemical moieties depicted in circles as indicated by the dashed arrows. Techniques to investigate the given moiety are depicted in boxes. The diamonds illustrate typical results, where the number of variants are derived from WES and WGS analysis on blood or fibroblasts, the number of transcripts and proteins is derived from investigations on fibroblasts and the number of metabolites is representative for analysis on plasma. The lower part illustrates possible observations derived from the respective species and technique.

First, the impact of a variant on RNA abundance can be assessed by computing expression outliers using for example the Z-score or statistical analysis on whole gene sets [149-151]. The aberrant expression can be caused by variants in non-coding regulatory regions as promoters, enhancers, and suppressors but also variants in coding regions [149, 152]. Importantly, expression outliers are not only a consequence of altered transcription itself, but can also arise due to changes in RNA degradation, for example by nonsense-mediated decay (NMD).

Second, RNA-seq can disclose differences in RNA sequence in cases of mono-allelic expression (MAE). MAE is found upon transcriptional silencing or post-transcriptional degradation of one allele [153]. Depending on the tissue, 1.7% to 3.7% of SNVs have been found to be allele-specifically expressed [154]. MAE has been shown on genetic as well as

epigenetic grounds. In cases of TAR syndrome, allele-specific expression of alleles harboring non-coding variants was shown as a consequence of the deletion of the second allele [155]. Epigenetically determined MAE can be the result of three distinct processes: parent-of-origin imprinting during gametogenesis yields the exclusive expression from either the paternal or maternal allele throughout the whole tissue or organism [156]; random MAE is seen for X-chromosome inactivation, where random selection and silencing of one of the two X-chromosomes balances gene-dosage [157]; finally, random MAE is also seen for autosomal genes [158]. Disease attributed to each of these instances are reviewed by Fahrner and Bjornsson [159]. With only few exceptions of X-linked or dominant examples, mitochondrial disorders follow a recessive mode of inheritance and hence we did not consider mono-allelic variants. MAE of heterozygous variants however mimics the effects of homozygous variants and therefore fits the most likely inheritance pattern. Thus, MAE might indicate the prioritization of heterozygous variants which would be missed otherwise.

Third, altered RNA sequence can be attributed to aberrant splicing. Aberrant splicing is a well-known cause of Mendelian disorders [160-162]. Nonetheless, prediction of splicing defects based on genetic information is not only hampered by the poor understanding of the cis-regulatory elements involved in splicing, but also insufficient capture of such elements in non-coding regions by WES for example [163]. Direct detection of splice isoforms by RNA-seq tremendously facilitates the discovery of splicing defects and has helped to identify various instances aberrant splicing in Mendelian diseases. Exon skipping and creation of a novel exon due to an intronic variant has been reported in *DMD* [164, 165]. Intron retention due to a splice site variant has been shown in *LMNA* [166]. A variant in a splicing factor binding site yielding in exon skipping was seen in *SMN1* [167].

As no systematic study of aberrant expression, MAE, and aberrant splicing had been conducted up to then, we assessed its power by performing RNA-seq on 105 fibroblast cell lines from suspected mitochondrial disease patients. For benchmarking and increment of statistical power, this 105 cell lines included 57 genetically diagnosed patients and 48 patients for which WES analysis was inconclusive. To adopt the analysis to a rare and recessive disease setting, we applied strict filtering for rare events with strong effect sizes. For aberrant expression, we only considered events with a Hochberg adjusted *P*-value smaller than 0.05 and a |Z-score| larger than 3. For aberrant splicing, we filtered for events with a Hochberg

adjusted *P*-value smaller than 0.05. Eventually, for MAE, we filtered for heterozygous SNVs with a RNA-seq coverage of more than 10 reads and considered a rare (MAF < 0.001) variant mono-allelically expressed when more than 80% of the reads harbored that variant and the Hochberg adjusted *P*-value was smaller than 0.05. Using these criteria the median detection per sample for aberrant expression was 1 event, for aberrant splicing 5 events, and for MAE 6 events. This small number of events makes manual inspection and validation feasible.

For 40 out of the 57 genetically diagnosed patients WES was available and we investigated whether the variants revealed by WES had an effect on the RNA. Stop variants were diagnosed in 11 patients, 5 thereof in a homozygous state. Out of these 5 homozygous stop variants, our pipeline reported aberrant expression for 3 and aberrant splicing for 1. Upon manual inspection, the aberrant splicing was classified as a false positive. Furthermore, a clear effect was seen for the remainder variant not detected by our computational pipeline prompting us to further optimize the pipeline. For the 6 patients harboring stop variants in a compound heterozygous state, 3 showed MAE while 2 reside in the very last exon therefore probably not evoking NMD. Unexpectedly, a detectable RNA defect was only apparent for one of the 9 frameshift variants. In all other cases the expression level did not change significantly. In 7 out of 8 patients with a pathogenic splicing variant our pipeline reported aberrant splicing. In the remainder, aberrant splicing was seen by manual inspection and revealed intron retention. Intron retention depicts a special instance of aberrant splicing where no split reads are generated at the splice junction. As the tool we employed to quantify RNA splicing predicts aberrant splicing based on the generation of new split reads, it is blind to intron retention. Until this drawback of the tool has not been amended, intron retention presents a possible blind spot in our study. As anticipated, no RNA defect was seen for any of the disease-associated genes solely carrying bi-allelic missense mutations as missense variants more likely impact protein function than RNA expression.

For 36 out of the 48 undiagnosed patients, the developed pipeline provided a strong candidate gene, defined as known disease-associated or mitochondrial protein-encoding gene. Based on manual inspection I selected 5 candidates. Validation provided a diagnosis for all 5 patients (out of 48, 10%). Amongst these, in two patients a new disease-associated gene (*TIMMDC1*) was identified. In one patient, the suspicion of a mitochondrial disorder was not confirmed but mutations in a gene (*MCOLN1*) involved in mucolipidosis were shown to be causative. This

illustrates the global character of our approach as we are not restricted neither by previous knowledge on disease-association of a gene nor by the assumption of a certain disease class.

Interestingly, defective splicing resulting in a loss of function was the predominantly detected cause in our newly diagnosed patients. This is in line with the ever increasing recognition of splicing defects in both, Mendelian and common disorders [168-170]. The aberrant splicing detected in *TIMMDC1* was most likely caused by a deep intronic variant (c.596+2146A>G) that was not covered by WES. While the variant was called from WGS data, prioritization of this variant was only achieved by using the information obtained by RNA-seq analysis. RNA-seq can therefore provide crucial additional information on intronic positions possibly affecting splicing upon mutation. Importantly, the defective splicing seen in the case of *TIMMDC1* arose from a loci demonstrating weak splicing also in control samples. We could show that private exons resulting from such aberrant splicing were often generated at these loci with weak splicing of about 1%. Suggestively, variants close to these cryptic splice sites therefore more likely affect splicing. Prioritization of such variants could be eased by the generation of tissue-specific maps of cryptic splice sites by RNA-seq analysis of multiple tissues from healthy donors like GTEx [154].

4.4.3. Limitations of RNA-seq

The 10% success rate of our approach is preliminary and only includes validated variants, whereas the pipeline provided good candidates in 36 unsolved cases. During the preparation of this thesis, one additional case was solved based on our findings. Validation of further candidate variants is ongoing. However, some patients will likely remain without diagnosis after our RNA-seq analysis. This can be due to several reasons which I will elute to in the following.

First, we used a standard non-strand specific RNA-seq protocol. During non-strand specific RNA-seq library preparation, the information about the strand specific origin of each transcript is lost. This information is however crucial especially for overlapping genes as without strand information our pipeline cannot distinguish from which of the overlapping genes the transcript originated. For our analysis, we could therefore only consider transcripts originating from non-overlapping genes leading to the exclusion of transcripts because there genetic origin was ambiguous.

Second, we used very strict filtering in regard to coverage, frequency, and effect size to adopt to a rare and recessive disease setting. Therefore, our approach is only successful in identifying the disease cause if the rare variants provoke a strong RNA defect. For only 1 out of 9 frameshift mutations a significant effect on RNA was detected. Possibly, some of these variants exert a moderate change which could be detected by applying more relaxed filtering in combination with larger sample size.

Third, aberrant splicing was missed by our pipeline in one case due to intron retention, a blind spot of the employed prediction tool as LeafCutter computes aberrant splicing based on split reads [171]. In the case of intron retention no split reads are generated though and such events can therefore not be detected. Speculatively, the pipeline might also have missed intron retention events in other undiagnosed patients.

Forth, the primary affected tissues like heart, liver, brain, or skeletal muscle are rarely available. We therefore conducted our experiments in patient derived fibroblasts which are a side product of the muscles biopsies performed for biochemical diagnosis of mitochondrial disease patients (see 1.1.3). Even though the fibroblast cell lines are usually not a primarily affected tissue, the fitness of the fibroblasts' mitochondria is determined by the same genetic background as that of mitochondria from other tissues. While some genes with mitochondrial functions might be expressed in a tissue specific manor, many have been reported to be expressed in most tissues [172]. Functional consequences of pathogenic variants in these genes should hence also been detectable in fibroblasts. Nevertheless, our RNA-seq analysis in fibroblast cell lines detected only 12,680 reliably transcribed genes (≥ 10 reads in 5% of the samples) which depicts less than two thirds of all genes. We must therefore assume that performing RNA-seq in the affected tissue would provide a more complete picture and increase our diagnostic output to a certain extend. This is supported by a corroborating study for neuromuscular disorders, where the RNA-seq analysis was performed directly in muscle tissue yielding a success rate of more than 30% [169].

Fifth, as for WES findings, findings from RNA-seq data need to be validated carefully. The perpetrator of the observed RNA defect, hence the causal variant, should be identified from WES or WGS data to rule out any artefacts. If no variant can be observed in close proximity to the aberrant event and a sample swap was excluded, additional evidence from clinical and experimental data can provide critical evidence as shown for the *MGST1* case. MGST1 has a

postulated role in the reactive oxygen species (ROS) defense system. The clinical findings of a childhood-onset neurodegenerative closely resembled the symptoms of another patient suffering from a perturbation of the ROS defense system and increased ROS in patient fibroblasts supported the pathogenicity of the aberrant expression [173]. If no further insight can be gained on the functional level, repeated detection of a RNA defect in additional tissue as for example blood can ascertain the specificity of the observed event. Conceivable grounds beyond variants in close proximity to the gene can include aberrant levels of antisense transcripts, miRNAs, altered epigenetics, and many more.

5. Outlook

WES has proven as a successful tool allowing diagnosis of about half of the patients with suspected mitochondrial disorders. Due to this success rate, Wortmann et al. postulated a shift in the diagnostic workup of mitochondrial disorders away from biochemical investigations on muscle biopsies, the gold standard up to now, towards a first tier genetic testing in blood [174]. This paradigm shift not only spares the patients the burden of invasive testing unless absolutely necessary but directly leads to profound diagnosis enabling genetic counseling, therapeutic means, and prenatal testing amongst others.

Albeit the tremendous impact of WES, for half of the suspected mitochondrial disease patients no diagnosis is achieved. Inconclusive WES can be due to technical or analytical limitations. Some of these limitations, as for example the detection of indels and CNVs as well as the detection of non-coding variants, can be overcome by WGS. As sequencing costs further decrease, WGS will therefore completely replace WES in the near future. However, prioritization of variants, especially in the non-coding region, will still remain difficult. By directly probing variant effects on RNA, RNA-seq can complement the sole genetic information gained from WES or WGS. As part of this PhD study, I laid the foundation of RNA-seq analysis as a second tier approach, which provided so far a molecular diagnosis for 10% of the patients with inconclusive WES or even WGS. While further investigations on the herein generated data using improved prioritization pipelines and extended validation experiments most likely will yield further diagnoses, RNA-seq on fibroblast cell lines will still leave a substantial amount of patients without diagnosis.

Remaining diagnostic deficits can partially be attributed to limitations inherent to our approach. Due to our strict filtering, we possibly missed to prioritize moderate but nevertheless causative events. As we and likely others will continue performing RNA-seq, increasing sample size will allow a more sophisticated statistical analysis and therefore ease the detection of such moderate events. Furthermore, detection of intron retention was hampered by limitations of the analysis tool employed to detect aberrant splicing. Currently available tools were designed for the comparison between groups. Our collaborators at the computational biology department of the Technische Universität München therefore currently develop a prediction tool for aberrant splicing that is first of all adopted to a rare disease setting by using different statistical models. Additionally, instead of considering split reads at a gene level, the new tool probes each junction at an exon level to allow the detection of intron retention. The tool furthermore aims to improve the prioritization of events by integrating genetic information from WES or WGS allowing the direct query for causative variants. Finally, in this study a standard non-strand specific RNA-seq protocol was used. To obtain information on overlapping genes, antisense transcripts, and long non-coding RNAs, we will now conduct RNA-seq experiments using a strand-specific RNA-seq protocol.

RNA-seq is furthermore fruitless if the causative variant does not have an effect on RNA abundance and sequence as expected for missense mutations, many *de novo* mutations, and dominant inheritance. In turn, such mutations likely affect protein levels and activity, both of which are however challenging to measure globally in large scale. While the yield and reproducibility of quantitative protein level measurements by quantitative proteomic approaches is constantly improving, the coverage is still not complete. Therefore, quantitative proteomics alone does not yet allow systematic detection of causal protein level aberration. Nevertheless, by now thousands of proteins can be detected allowing the validation of variant effects in the respective genes on protein synthesis, stability, and degradation. In this study, quantitative proteomics on 31 fibroblast samples detected 5400 proteins out of which 4000 were reliably detected in at least 50% of the samples [107]. In the *TIMMDC1* patients we detected systematic decrease and absence of several RCCI subunits underpinning the role of TIMMDC1 as RCCI assembly factor. The instability of macromolecular complexes like RCCI and subsequent degradation of subunits caused by mutations in structural proteins might be a common mechanism. Such events might be picked up more easily due to the larger amount of observable events and more robust statistical models. Notably, quantitative

proteomics can also reveal effects of VUS as exemplified by the *ALDH18A1* patient carrying a VUS compound heterozygous with a nonsense variant in *ALDH18A1*. As only 2% remaining ALDH18A1 can be measured, this argues for an effect not only of the nonsense variant but also of the VUS on the protein level. Furthermore, as mitochondrial disorders present metabolic disorders, the routine diagnostic process for mitochondrial disorders often involves the measurement of a very few metabolites serving as biomarkers (see 1.1.2). These biomarkers, however, generally have a poor sensitivity and specificity towards mitochondrial disorders as a whole, probably attributed to their diversity in origin and hence manifestation. Even though, they can allow the metabolic diagnosis of certain subgroups. Whereas comprehensive quantitative metabolomics is not yet used in a routine diagnostic set-up, it has been applied successfully in the past to unravel novel biomarkers [175].[175]. While this was outside the scope of the herein presented investigations, we could demonstrate the usability of quantitative metabolomics for the prioritization of variants in single cases whereby the causality of the variants in *ALDH18A1*, initially rejected due to an at that point incoherent phenotype, could be re-evaluated.

Finally, our yield might be limited by peculiarities of the investigated tissue. While limitations attributed to tissue-specific expression might further be circumvented by performing RNA-seq on the affected tissue, in the majority of the cases the affected tissue probably will not be available. To elude this shortage, one could generate induced pluripotent stem cells (iPSC) from fibroblasts and differentiate the iPSC into the affected tissue, which is however laborious and not yet routinely performed. Further insights might be derived from performing RNA-seq on patient blood samples, which are readily available. Performing RNA-seq on blood could i) replace RNA-seq on fibroblasts as a less invasive procedure; ii) validate moderate events detected in fibroblasts; iii) detect additional tissue-specific events. I therefore suggest to investigate the applicability of RNA-seq on blood samples by parallel RNA-seq on multiple tissues.

Conclusively, the era of next generation sequencing continuously provides molecular diagnosis for patients. In combination with the clinical presentation of the patients, this allows for a better understanding and estimation of genotype-phenotype correlations. While for some genes, as for example *TANGO2*, the overlapping clinical presentation of the patients strongly facilitated the discovery of the causal variants, for deficiencies in many other genes, e.g.

ALDH18A1, a broadening of the clinical spectrum is evident. Ambiguous genotype-phenotype correlations together with the increasing amount of VUS generated by WGS demands for comprehensive evaluation and validation of variants in large-scale and high-throughput. RNA-seq, quantitative proteomics and quantitative metabolomics will therefore become vital to the diagnostic process, initiating the multi-omics era.

6. References

[1] Margulis, L., *Origins of species: acquired genomes and individuality.* Biosystems, 1993. **31**(2-3): p. 121-5.

[2] Andersson, S.G. and C.G. Kurland, *Origins of mitochondria and hydrogenosomes.* Curr Opin Microbiol, 1999. **2**(5): p. 535-41.

[3] Gray, M.W., G. Burger, and B.F. Lang, *The origin and early evolution of mitochondria.* Genome Biol, 2001. **2**(6): p. REVIEWS1018.

[4] Spang, A., J.H. Saw, S.L. Jorgensen, K. Zaremba-Niedzwiedzka, J. Martijn, A.E. Lind, R. van Eijk, C. Schleper, L. Guy, and T.J. Ettema, *Complex archaea that bridge the gap between prokaryotes and eukaryotes.* Nature, 2015. **521**(7551): p. 173-9.

[5] Lill, R. and U. Muhlenhoff, *Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases.* Annu Rev Biochem, 2008. **77**: p. 669-700.

[6] Lambeth, J.D., D.W. Seybert, J.R. Lancaster, Jr., J.C. Salerno, and H. Kamin, *Steroidogenic electron transport in adrenal cortex mitochondria.* Mol Cell Biochem, 1982. **45**(1): p. 13-31.

[7] Graier, W.F., M. Frieden, and R. Malli, *Mitochondria and Ca(2+) signaling: old guests, new functions.* Pflugers Arch, 2007. **455**(3): p. 375-96.

[8] Heath-Engel, H.M. and G.C. Shore, *Mitochondrial membrane dynamics, cristae remodelling and apoptosis.* Biochim Biophys Acta, 2006. **1763**(5-6): p. 549-60.

[9] Johnston, I.G. and B.P. Williams, *Evolutionary Inference across Eukaryotes Identifies Specific Pressures Favoring Mitochondrial Gene Retention.* Cell Syst, 2016. **2**(2): p. 101-11.

[10] Giles, R.E., H. Blanc, H.M. Cann, and D.C. Wallace, *Maternal inheritance of human mitochondrial DNA.* Proc Natl Acad Sci U S A, 1980. **77**(11): p. 6715-9.

[11] Mayr, J.A., T.B. Haack, P. Freisinger, D. Karall, C. Makowski, J. Koch, R.G. Feichtinger, F.A. Zimmermann, B. Rolinski, U. Ahting, T. Meitinger, H. Prokisch, and W. Sperl, *Spectrum of combined respiratory chain defects.* J Inherit Metab Dis, 2015. **38**(4): p. 629-40.

[12] Reinecke, F., J.A. Smeitink, and F.H. van der Westhuizen, *OXPHOS gene expression and control in mitochondrial disorders.* Biochim Biophys Acta, 2009. **1792**(12): p. 1113-21.

[13] Munnich, A. and P. Rustin, *Clinical spectrum and diagnosis of mitochondrial disorders.* Am J Med Genet, 2001. **106**(1): p. 4-17.

[14] Ernster, L., D. Ikkos, and R. Luft, *Enzymic activities of human skeletal muscle mitochondria: a tool in clinical metabolic research.* Nature, 1959. **184**: p. 1851-4.

[15] Luft, R., D. Ikkos, G. Palmieri, L. Ernster, and B. Afzelius, *A case of severe hypermetabolism of nonthyroid origin with a defect in the maintenance of mitochondrial respiratory control: a correlated clinical, biochemical, and morphological study.* J Clin Invest, 1962. **41**: p. 1776-804.

[16] Rustin, P., D. Chretien, T. Bourgeron, B. Gerard, A. Rotig, J.M. Saudubray, and A. Munnich, *Biochemical and molecular investigations in respiratory chain deficiencies.* Clin Chim Acta, 1994. **228**(1): p. 35-51.

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

[18] Manwaring, N., M.M. Jones, J.J. Wang, E. Rochtchina, C. Howard, P. Mitchell, and C.M. Sue, *Population prevalence of the MELAS A3243G mutation.* Mitochondrion, 2007. **7**(3): p. 230-3.

[19] Smeitink, J.A., *Mitochondrial disorders: clinical presentation and diagnostic dilemmas.* J Inherit Metab Dis, 2003. **26**(2-3): p. 199-207.

[20] Baertling, F., R.J. Rodenburg, J. Schaper, J.A. Smeitink, W.J. Koopman, E. Mayatepek, E. Morava, and F. Distelmaier, *A guide to diagnosis and treatment of Leigh syndrome.* J Neurol Neurosurg Psychiatry, 2014. **85**(3): p. 257-65.

[21] Saneto, R.P., B.H. Cohen, W.C. Copeland, and R.K. Naviaux, *Alpers-Huttenlocher syndrome.* Pediatr Neurol, 2013. **48**(3): p. 167-78.

[22] Munnich, A., A. Rotig, D. Chretien, J.M. Saudubray, V. Cormier, and P. Rustin, *Clinical presentations and laboratory investigations in respiratory chain deficiency.* Eur J Pediatr, 1996. **155**(4): p. 262-74.

[23] Wallace, D.C., *Diseases of the mitochondrial DNA.* Annu Rev Biochem, 1992. **61**: p. 1175-212.

[24] Koenig, M.K., *Presentation and diagnosis of mitochondrial disorders in children.* Pediatr Neurol, 2008. **38**(5): p. 305-13.

[25] Smeitink, J.A., M. Zeviani, D.M. Turnbull, and H.T. Jacobs, *Mitochondrial medicine: a metabolic perspective on the pathology of oxidative phosphorylation disorders.* Cell Metab, 2006. **3**(1): p. 9-13.

[26] Munnich, A., P. Rustin, A. Rotig, D. Chretien, J.P. Bonnefont, C. Nuttin, V. Cormier, A. Vassault, P. Parvy, J. Bardet, and et al., *Clinical aspects of mitochondrial disorders.* J Inherit Metab Dis, 1992. **15**(4): p. 448-55.

[27] Mitochondrial Medicine Society's Committee on, D., R.H. Haas, S. Parikh, M.J. Falk, R.P. Saneto, N.I. Wolf, N. Darin, L.J. Wong, B.H. Cohen, and R.K. Naviaux, *The in-depth evaluation of suspected mitochondrial disease.* Mol Genet Metab, 2008. **94**(1): p. 16-37.

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

[29] Loeffen, J.L., J.A. Smeitink, J.M. Trijbels, A.J. Janssen, R.H. Triepels, R.C. Sengers, and L.P. van den Heuvel, *Isolated complex I deficiency in children: clinical, biochemical and genetic aspects.* Hum Mutat, 2000. **15**(2): p. 123-34.

[30] Wortmann, S.B., L.A. Kluijtmans, R.J. Rodenburg, J.O. Sass, J. Nouws, E.P. van Kaauwen, T. Kleefstra, L. Tranebjaerg, M.C. de Vries, P. Isohanni, K. Walter, F.S. Alkuraya, I. Smuts, C.J. Reinecke, F.H. van der Westhuizen, D. Thorburn, J.A. Smeitink, E. Morava, and R.A. Wevers, *3-Methylglutaconic aciduria--lessons from 50 genes and 977 patients.* J Inherit Metab Dis, 2013. **36**(6): p. 913-21.

[31] Suomalainen, A., J.M. Elo, K.H. Pietilainen, A.H. Hakonen, K. Sevastianova, M. Korpela, P. Isohanni, S.K. Marjavaara, T. Tyni, S. Kiuru-Enari, H. Pihko, N. Darin, K. Ounap, L.A. Kluijtmans, A. Paetau, J. Buzkova, L.A. Bindoff, J. Annunen-Rasila, J. Uusimaa, A. Rissanen, H. Yki-Jarvinen, M. Hirano, M. Tulinius, J. Smeitink, and H. Tyynismaa, *FGF-21 as a biomarker for muscle-manifesting mitochondrial respiratory chain deficiencies: a diagnostic study.* Lancet Neurol, 2011. **10**(9): p. 806-18.

[32] Yatsuga, S., Y. Fujita, A. Ishii, Y. Fukumoto, H. Arahata, T. Kakuma, T. Kojima, M. Ito, M. Tanaka, R. Saiki, and Y. Koga, *Growth differentiation factor 15 as a useful biomarker for mitochondrial disorders.* Ann Neurol, 2015. **78**(5): p. 814-23.

[33] Dimauro, S., *A history of mitochondrial diseases.* J Inherit Metab Dis, 2011. **34**(2): p. 261-76.

[34] DiMauro, S., E. Bonilla, M. Zeviani, M. Nakagawa, and D.C. DeVivo, *Mitochondrial myopathies.* Ann Neurol, 1985. **17**(6): p. 521-38.

[35] Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.

[36] Gusella, J.F., N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Tanzi, P.C. Watkins, K. Ottina, M.R. Wallace, A.Y. Sakaguchi, and et al., *A polymorphic DNA marker genetically linked to Huntington's disease.* Nature, 1983. **306**(5940): p. 234-8.

[37] Holt, I.J., A.E. Harding, and J.A. Morgan-Hughes, *Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies.* Nature, 1988. **331**(6158): p. 717-9.

[38] Altshuler, D., M.J. Daly, and E.S. Lander, *Genetic mapping in human disease.* Science, 2008. **322**(5903): p. 881-8.

[39] Botstein, D., R.L. White, M. Skolnick, and R.W. Davis, *Construction of a genetic linkage map in man using restriction fragment length polymorphisms.* Am J Hum Genet, 1980. **32**(3): p. 314-31.

[40] Servidei, S., *Mitochondrial encephalomyopathies: gene mutation.* Neuromuscul Disord, 2001. **11**(8): p. 774-9.

[41] Schon, E.A., S. DiMauro, and M. Hirano, *Human mitochondrial DNA: roles of inherited and somatic mutations.* Nat Rev Genet, 2012. **13**(12): p. 878-90.

[42] Zeviani, M., S. Servidei, C. Gellera, E. Bertini, S. DiMauro, and S. DiDonato, *An autosomal dominant disorder with multiple deletions of mitochondrial DNA starting at the D-loop region.* Nature, 1989. **339**(6222): p. 309-11.

[43] Moraes, C.T., S. Shanske, H.J. Tritschler, J.R. Aprille, F. Andreetta, E. Bonilla, E.A. Schon, and S. DiMauro, *mtDNA depletion with variable tissue expression: a novel genetic abnormality in mitochondrial diseases.* Am J Hum Genet, 1991. **48**(3): p. 492-501.

[44] Bourgeron, T., P. Rustin, D. Chretien, M. Birch-Machin, M. Bourgeois, E. Viegas-Pequignot, A. Munnich, and A. Rotig, *Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency.* Nat Genet, 1995. **11**(2): p. 144-9.

[45] Nesbitt, V., R.D. Pitceathly, D.M. Turnbull, R.W. Taylor, M.G. Sweeney, E.E. Mudanohwo, S. Rahman, M.G. Hanna, and R. McFarland, *The UK MRC Mitochondrial Disease Patient Cohort Study: clinical phenotypes associated with the m.3243A>G mutation--implications for diagnosis and management.* J Neurol Neurosurg Psychiatry, 2013. **84**(8): p. 936-8.

[46] Tchikviladze, M., M. Gilleron, T. Maisonobe, D. Galanaud, P. Laforet, A. Durr, B. Eymard, F. Mochel, H. Ogier, A. Behin, T. Stojkovic, B. Degos, I. Gourfinkel-An, F. Sedel, M. Anheim, A. Elbaz, K. Viala, M. Vidailhet, A. Brice, C. Jardel, and A. Lombes, *A diagnostic flow chart for POLG-related diseases based on signs sensitivity and specificity.* J Neurol Neurosurg Psychiatry, 2015. **86**(6): p. 646-54.

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

[48] Shendure, J. and H. Ji, *Next-generation DNA sequencing.* Nat Biotechnol, 2008. **26**(10): p. 1135-45.
[49] Ng, S.B., E.H. Turner, P.D. Robertson, S.D. Flygare, A.W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E.E. Eichler, M. Bamshad, D.A. Nickerson, and J. Shendure, *Targeted capture and massively parallel sequencing of 12 human exomes.* Nature, 2009. **461**(7261): p. 272-6.
[50] Olson, M., *Enrichment of super-sized resequencing targets from the human genome.* Nat Methods, 2007. **4**(11): p. 891-2.
[51] Cooper, D.N., J.M. Chen, E.V. Ball, K. Howells, M. Mort, A.D. Phillips, N. Chuzhanova, M. Krawczak, H. Kehrer-Sawatzki, and P.D. Stenson, *Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics.* Hum Mutat, 2010. **31**(6): p. 631-55.
[52] Majewski, J., J. Schwartzentruber, E. Lalonde, A. Montpetit, and N. Jabado, *What can exome sequencing do for you?* J Med Genet, 2011. **48**(9): p. 580-9.
[53] Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E.M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D.B. Jaffe, E.S. Lander, and C. Nusbaum, *Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing.* Nat Biotechnol, 2009. **27**(2): p. 182-9.
[54] Choi, M., U.I. Scholl, W. Ji, T. Liu, I.R. Tikhonova, P. Zumbo, A. Nayir, A. Bakkaloglu, S. Ozen, S. Sanjad, C. Nelson-Williams, A. Farhi, S. Mane, and R.P. Lifton, *Genetic diagnosis by whole exome capture and massively parallel DNA sequencing.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19096-101.
[55] Haack, T.B., K. Danhauser, B. Haberberger, J. Hoser, V. Strecker, D. Boehm, G. Uziel, E. Lamantea, F. Invernizzi, J. Poulton, B. Rolinski, A. Iuso, S. Biskup, T. Schmidt, H.W. Mewes, I. Wittig, T. Meitinger, M. Zeviani, and H. Prokisch, *Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency.* Nat Genet, 2010. **42**(12): p. 1131-4.
[56] Wortmann, S.B., D.A. Koolen, J.A. Smeitink, L. van den Heuvel, and R.J. Rodenburg, *Whole exome sequencing of suspected mitochondrial patients in clinical practice.* J Inherit Metab Dis, 2015. **38**(3): p. 437-43.
[57] Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
[58] Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.* Bioinformatics, 2011. **27**(21): p. 2987-93.
[59] Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing, *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
[60] Botstein, D. and N. Risch, *Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease.* Nat Genet, 2003. **33 Suppl**: p. 228-37.
[61] Bamshad, M.J., S.B. Ng, A.W. Bigham, H.K. Tabor, M.J. Emond, D.A. Nickerson, and J. Shendure, *Exome sequencing as a tool for Mendelian disease gene discovery.* Nat Rev Genet, 2011. **12**(11): p. 745-55.
[62] Warr, A., C. Robert, D. Hume, A. Archibald, N. Deeb, and M. Watson, *Exome Sequencing: Current and Future Perspectives.* G3 (Bethesda), 2015. **5**(8): p. 1543-50.
[63] Sulonen, A.M., P. Ellonen, H. Almusa, M. Lepisto, S. Eldfors, S. Hannula, T. Miettinen, H. Tyynismaa, P. Salo, C. Heckman, H. Joensuu, T. Raivio, A.

Suomalainen, and J. Saarela, *Comparison of solution-based exome capture methods for next generation sequencing.* Genome Biol, 2011. **12**(9): p. R94.

[64] Mamanova, L., A.J. Coffey, C.E. Scott, I. Kozarewa, E.H. Turner, A. Kumar, E. Howard, J. Shendure, and D.J. Turner, *Target-enrichment strategies for next-generation sequencing.* Nat Methods, 2010. **7**(2): p. 111-8.

[65] McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo, *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.

[66] Hwang, S., E. Kim, I. Lee, and E.M. Marcotte, *Systematic comparison of variant calling pipelines using gold standard personal exome variants.* Sci Rep, 2015. **5**: p. 17875.

[67] Plagnol, V., J. Curtis, M. Epstein, K.Y. Mok, E. Stebbings, S. Grigoriadou, N.W. Wood, S. Hambleton, S.O. Burns, A.J. Thrasher, D. Kumararatne, R. Doffinger, and S. Nejentsev, *A robust model for read count data in exome sequencing experiments and implications for copy number variant calling.* Bioinformatics, 2012. **28**(21): p. 2747-54.

[68] Tan, R., Y. Wang, S.E. Kleinstein, Y. Liu, X. Zhu, H. Guo, Q. Jiang, A.S. Allen, and M. Zhu, *An evaluation of copy number variation detection tools from whole-exome sequencing data.* Hum Mutat, 2014. **35**(7): p. 899-907.

[69] Medvedev, P., M. Stanciu, and M. Brudno, *Computational methods for discovering structural variation with next-generation sequencing.* Nat Methods, 2009. **6**(11 Suppl): p. S13-20.

[70] Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.

[71] Cingolani, P., A. Platts, L. Wang le, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, and D.M. Ruden, *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.* Fly (Austin), 2012. **6**(2): p. 80-92.

[72] Lek, M., K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A.H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, T. Tukiainen, D.P. Birnbaum, J.A. Kosmicki, L.E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D.N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M.I. Kurki, A.L. Moonshine, P. Natarajan, L. Orozco, G.M. Peloso, R. Poplin, M.A. Rivas, V. Ruano-Rubio, S.A. Rose, D.M. Ruderfer, K. Shakir, P.D. Stenson, C. Stevens, B.P. Thomas, G. Tiao, M.T. Tusie-Luna, B. Weisburd, H.H. Won, D. Yu, D.M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J.C. Florez, S.B. Gabriel, G. Getz, S.J. Glatt, C.M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M.I. McCarthy, D. McGovern, R. McPherson, B.M. Neale, A. Palotie, S.M. Purcell, D. Saleheen, J.M. Scharf, P. Sklar, P.F. Sullivan, J. Tuomilehto, M.T. Tsuang, H.C. Watkins, J.G. Wilson, M.J. Daly, D.G. MacArthur, and C. Exome Aggregation, *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.

[73] Siepel, A., K.S. Pollard, and D. Haussler, *New Methods for Detecting Lineage-Specific Selection,* in *Research in Computational Molecular Biology: 10th Annual International Conference, RECOMB 2006, Venice, Italy, April 2-5, 2006.*

Proceedings, A. Apostolico, et al., Editors. 2006, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 190-205.

[74] Garber, M., M. Guttman, M. Clamp, M.C. Zody, N. Friedman, and X. Xie, *Identifying novel constrained elements by exploiting biased substitution patterns.* Bioinformatics, 2009. **25**(12): p. i54-62.

[75] Kircher, M., D.M. Witten, P. Jain, B.J. O'Roak, G.M. Cooper, and J. Shendure, *A general framework for estimating the relative pathogenicity of human genetic variants.* Nat Genet, 2014. **46**(3): p. 310-5.

[76] Schwarz, J.M., C. Rodelsperger, M. Schuelke, and D. Seelow, *MutationTaster evaluates disease-causing potential of sequence alterations.* Nat Methods, 2010. **7**(8): p. 575-6.

[77] Adzhubei, I.A., S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, and S.R. Sunyaev, *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

[78] Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic Acids Res, 2003. **31**(13): p. 3812-4.

[79] Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc, 2009. **4**(7): p. 1073-81.

[80] Kondrashov, A.S., S. Sunyaev, and F.A. Kondrashov, *Dobzhansky-Muller incompatibilities in protein evolution.* Proc Natl Acad Sci U S A, 2002. **99**(23): p. 14878-83.

[81] Wieland, T., *Next-Generation Sequencing Data Analysis*. 2015, Diss. Technische Universität München: http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20151111-1244253-1-6.

[82] MacArthur, D.G., T.A. Manolio, D.P. Dimmock, H.L. Rehm, J. Shendure, G.R. Abecasis, D.R. Adams, R.B. Altman, S.E. Antonarakis, E.A. Ashley, J.C. Barrett, L.G. Biesecker, D.F. Conrad, G.M. Cooper, N.J. Cox, M.J. Daly, M.B. Gerstein, D.B. Goldstein, J.N. Hirschhorn, S.M. Leal, L.A. Pennacchio, J.A. Stamatoyannopoulos, S.R. Sunyaev, D. Valle, B.F. Voight, W. Winckler, and C. Gunter, *Guidelines for investigating causality of sequence variants in human disease.* Nature, 2014. **508**(7497): p. 469-76.

[83] Xue, Y., Y. Chen, Q. Ayub, N. Huang, E.V. Ball, M. Mort, A.D. Phillips, K. Shaw, P.D. Stenson, D.N. Cooper, C. Tyler-Smith, and C. Genomes Project, *Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing.* Am J Hum Genet, 2012. **91**(6): p. 1022-32.

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

[85] Kruglyak, L. and D.A. Nickerson, *Variation is the spice of life.* Nat Genet, 2001. **27**(3): p. 234-6.

[86] O'Sullivan, B.P. and S.D. Freedman, *Cystic fibrosis.* Lancet, 2009. **373**(9678): p. 1891-904.

[87] Carroll, C.J., V. Brilhante, and A. Suomalainen, *Next-generation sequencing for mitochondrial disorders.* Br J Pharmacol, 2014. **171**(8): p. 1837-53.

[88] Chinnery, P.F., *Mitochondrial Disorders Overview*, in *GeneReviews(R)*, R.A. Pagon, et al., Editors. 1993: Seattle (WA).

[89] Alston, C.L., M.C. Rocha, N.Z. Lax, D.M. Turnbull, and R.W. Taylor, *The genetics and pathology of mitochondrial disease.* The Journal of Pathology, 2016: p. n/a-n/a.

[90] Thompson, K., H. Majd, C. Dallabona, K. Reinson, M.S. King, C.L. Alston, L. He, T. Lodi, S.A. Jones, A. Fattal-Valevski, N.D. Fraenkel, A. Saada, A. Haham, P. Isohanni, R. Vara, I.A. Barbosa, M.A. Simpson, C. Deshpande, S. Puusepp, P.E. Bonnen, R.J. Rodenburg, A. Suomalainen, K. Ounap, O. Elpeleg, I. Ferrero, R. McFarland, E.R. Kunji, and R.W. Taylor, *Recurrent De Novo Dominant Mutations in SLC25A4 Cause Severe Early-Onset Mitochondrial Disease and Loss of Mitochondrial DNA Copy Number.* Am J Hum Genet, 2016. **99**(6): p. 1405.

[91] Kong, A., M.L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S.A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, W.S. Wong, G. Sigurdsson, G.B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D.F. Gudbjartsson, A. Helgason, O.T. Magnusson, U. Thorsteinsdottir, and K. Stefansson, *Rate of de novo mutations and the importance of father's age to disease risk.* Nature, 2012. **488**(7412): p. 471-5.

[92] Stenson, P.D., E.V. Ball, M. Mort, A.D. Phillips, J.A. Shiel, N.S. Thomas, S. Abeysinghe, M. Krawczak, and D.N. Cooper, *Human Gene Mutation Database (HGMD): 2003 update.* Hum Mutat, 2003. **21**(6): p. 577-81.

[93] Landrum, M.J., J.M. Lee, G.R. Riley, W. Jang, W.S. Rubinstein, D.M. Church, and D.R. Maglott, *ClinVar: public archive of relationships among sequence variation and human phenotype.* Nucleic Acids Res, 2014. **42**(Database issue): p. D980-5.

[94] Oyston, J., *Online Mendelian Inheritance in Man.* Anesthesiology, 1998. **89**(3): p. 811-2.

[95] Collet, M., Z. Assouline, D. Bonnet, M. Rio, F. Iserin, D. Sidi, A. Goldenberg, C. Lardennois, M.D. Metodiev, B. Haberberger, T. Haack, A. Munnich, H. Prokisch, and A. Rotig, *High incidence and variable clinical outcome of cardiac hypertrophy due to ACAD9 mutations in childhood.* Eur J Hum Genet, 2016. **24**(8): p. 1112-6.

[96] O'Donnell-Luria, A.H. and D.T. Miller, *A Clinician's perspective on clinical exome sequencing.* Hum Genet, 2016. **135**(6): p. 643-54.

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

[98] Ohtake, A., K. Murayama, M. Mori, H. Harashima, T. Yamazaki, S. Tamaru, Y. Yamashita, Y. Kishita, Y. Nakachi, M. Kohda, Y. Tokuzawa, Y. Mizuno, Y. Moriyama, H. Kato, and Y. Okazaki, *Diagnosis and molecular basis of mitochondrial respiratory chain disorders: exome sequencing for disease gene identification.* Biochim Biophys Acta, 2014. **1840**(4): p. 1355-9.

[99] Distelmaier, F., T.B. Haack, S.B. Wortmann, J.A. Mayr, and H. Prokisch, *Treatable mitochondrial diseases: cofactor metabolism and beyond.* Brain, 2017. **140**(Pt 2): p. e11.

[100] Foley, A.R., M.P. Menezes, A. Pandraud, M.A. Gonzalez, A. Al-Odaib, A.J. Abrams, K. Sugano, A. Yonezawa, A.Y. Manzur, J. Burns, I. Hughes, B.G. McCullagh, H.

Jungbluth, M.J. Lim, J.P. Lin, A. Megarbane, J.A. Urtizberea, A.H. Shah, J. Antony, R. Webster, A. Broomfield, J. Ng, A.A. Mathew, J.J. O'Byrne, E. Forman, M. Scoto, M. Prasad, K. O'Brien, S. Olpin, M. Oppenheim, I. Hargreaves, J.M. Land, M.X. Wang, K. Carpenter, R. Horvath, V. Straub, M. Lek, W. Gold, M.O. Farrell, S. Brandner, R. Phadke, K. Matsubara, M.L. McGarvey, S.S. Scherer, P.S. Baxter, M.D. King, P. Clayton, S. Rahman, M.M. Reilly, R.A. Ouvrier, J. Christodoulou, S. Zuchner, F. Muntoni, and H. Houlden, *Treatable childhood neuronopathy caused by mutations in riboflavin transporter RFVT2.* Brain, 2014. **137**(Pt 1): p. 44-56.

[101] Haack, T.B., C.B. Jackson, K. Murayama, L.S. Kremer, A. Schaller, U. Kotzaeridou, M.C. de Vries, G. Schottmann, S. Santra, B. Buchner, T. Wieland, E. Graf, P. Freisinger, S. Eggimann, A. Ohtake, Y. Okazaki, M. Kohda, Y. Kishita, Y. Tokuzawa, S. Sauer, Y. Memari, A. Kolb-Kokocinski, R. Durbin, O. Hasselmann, K. Cremer, B. Albrecht, D. Wieczorek, H. Engels, D. Hahn, A.M. Zink, C.L. Alston, R.W. Taylor, R.J. Rodenburg, R. Trollmann, W. Sperl, T.M. Strom, G.F. Hoffmann, J.A. Mayr, T. Meitinger, R. Bolognini, M. Schuelke, J.M. Nuoffer, S. Kolker, H. Prokisch, and T. Klopstock, *Deficiency of ECHS1 causes mitochondrial encephalopathy with cardiac involvement.* Ann Clin Transl Neurol, 2015. **2**(5): p. 492-509.

[102] Soler-Alfonso, C., G.M. Enns, M.K. Koenig, H. Saavedra, E. Bonfante-Mejia, and H. Northrup, *Identification of HIBCH gene mutations causing autosomal recessive Leigh syndrome: a gene involved in valine metabolism.* Pediatr Neurol, 2015. **52**(3): p. 361-5.

[103] Pagliarini, D.J., S.E. Calvo, B. Chang, S.A. Sheth, S.B. Vafai, S.E. Ong, G.A. Walford, C. Sugiana, A. Boneh, W.K. Chen, D.E. Hill, M. Vidal, J.G. Evans, D.R. Thorburn, S.A. Carr, and V.K. Mootha, *A mitochondrial protein compendium elucidates complex I disease biology.* Cell, 2008. **134**(1): p. 112-23.

[104] Elstner, M., C. Andreoli, T. Klopstock, T. Meitinger, and H. Prokisch, *The mitochondrial proteome database: MitoP2.* Methods Enzymol, 2009. **457**: p. 3-20.

[105] Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer, and B. Wold, *Mapping and quantifying mammalian transcriptomes by RNA-Seq.* Nat Methods, 2008. **5**(7): p. 621-8.

[106] Tyanova, S., T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, and J. Cox, *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat Methods, 2016. **13**(9): p. 731-40.

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

[108] Kremer, L.S. and H. Prokisch, *Identification of Disease-Causing Mutations by Functional Complementation of Patient-Derived Fibroblast Cell Lines.* Methods Mol Biol, 2017. **1567**: p. 391-406.

[109] Danhauser, K., A. Iuso, T.B. Haack, P. Freisinger, K. Brockmann, J.A. Mayr, T. Meitinger, and H. Prokisch, *Cellular rescue-assay aids verification of causative DNA-variants in mitochondrial complex I deficiency.* Mol Genet Metab, 2011. **103**(2): p. 161-6.

[110] Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. **72**: p. 248-54.

[111] Calvo, S.E., A.G. Compton, S.G. Hershman, S.C. Lim, D.S. Lieber, E.J. Tucker, A. Laskowski, C. Garone, S. Liu, D.B. Jaffe, J. Christodoulou, J.M. Fletcher, D.L. Bruno, J. Goldblatt, S. Dimauro, D.R. Thorburn, and V.K. Mootha, *Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing.* Sci Transl Med, 2012. **4**(118): p. 118ra10.

[112] Kremer, L.S., C. L'Hermitte-Stead, P. Lesimple, M. Gilleron, S. Filaut, C. Jardel, T.B. Haack, T.M. Strom, T. Meitinger, H. Azzouz, N. Tebib, H. Ogier de Baulny, G. Touati, H. Prokisch, and A. Lombes, *Severe respiratory complex III defect prevents liver adaptation to prolonged fasting.* J Hepatol, 2016. **65**(2): p. 377-85.

[113] Invernizzi, F., M. Tigano, C. Dallabona, C. Donnini, I. Ferrero, M. Cremonte, D. Ghezzi, C. Lamperti, and M. Zeviani, *A homozygous mutation in LYRM7/MZM1L associated with early onset encephalopathy, lactic acidosis, and severe reduction of mitochondrial complex III activity.* Hum Mutat, 2013. **34**(12): p. 1619-22.

[114] Dallabona, C., T.E. Abbink, R. Carrozzo, A. Torraco, A. Legati, C.G. van Berkel, M. Niceta, T. Langella, D. Verrigni, T. Rizza, D. Diodato, F. Piemonte, E. Lamantea, M. Fang, J. Zhang, D. Martinelli, E. Bevivino, C. Dionisi-Vici, A. Vanderver, S.G. Philip, M.A. Kurian, I.C. Verma, S. Bijarnia-Mahay, S. Jacinto, F. Furtado, P. Accorsi, A. Ardissone, I. Moroni, I. Ferrero, M. Tartaglia, P. Goffrini, D. Ghezzi, M.S. van der Knaap, and E. Bertini, *LYRM7 mutations cause a multifocal cavitating leukoencephalopathy with distinct MRI appearance.* Brain, 2016. **139**(Pt 3): p. 782-94.

[115] Ghezzi, D., E. Baruffini, T.B. Haack, F. Invernizzi, L. Melchionda, C. Dallabona, T.M. Strom, R. Parini, A.B. Burlina, T. Meitinger, H. Prokisch, I. Ferrero, and M. Zeviani, *Mutations of the mitochondrial-tRNA modifier MTO1 cause hypertrophic cardiomyopathy and lactic acidosis.* Am J Hum Genet, 2012. **90**(6): p. 1079-87.

[116] Baruffini, E., C. Dallabona, F. Invernizzi, J.W. Yarham, L. Melchionda, E.L. Blakely, E. Lamantea, C. Donnini, S. Santra, S. Vijayaraghavan, H.P. Roper, A. Burlina, R. Kopajtich, A. Walther, T.M. Strom, T.B. Haack, H. Prokisch, R.W. Taylor, I. Ferrero, M. Zeviani, and D. Ghezzi, *MTO1 mutations are associated with hypertrophic cardiomyopathy and lactic acidosis and cause respiratory chain deficiency in humans and yeast.* Hum Mutat, 2013. **34**(11): p. 1501-9.

[117] Charif, M., S.M. Titah, A. Roubertie, V. Desquiret-Dumas, N. Gueguen, I. Meunier, J. Leid, F. Massal, X. Zanlonghi, J. Mercier, E. Raynaud de Mauverger, V. Procaccio, B. Mousson de Camaret, G. Lenaers, and C.P. Hamel, *Optic neuropathy, cardiomyopathy, cognitive disability in patients with a homozygous mutation in the nuclear MTO1 and a mitochondrial MT-TF variant.* Am J Med Genet A, 2015. **167A**(10): p. 2366-74.

[118] Martin, M.A., M.T. Garcia-Silva, G. Barcia, A. Delmiro, M.E. Rodriguez-Garcia, A. Blazquez, R. Francisco-Alvarez, E. Martin-Hernandez, P. Quijada-Fraile, P. Tejada-Palacios, J. Arenas, C. Santos, and F. Martinez-Azorin, *The homozygous R504C mutation in MTO1 gene is responsible for ONCE syndrome.* Clin Genet, 2017. **91**(1): p. 46-53.

[119] Tischner, C., A. Hofer, V. Wulff, J. Stepek, I. Dumitru, L. Becker, T. Haack, L. Kremer, A.N. Datta, W. Sperl, T. Floss, W. Wurst, Z. Chrzanowska-Lightowlers, M.H. De Angelis, T. Klopstock, H. Prokisch, and T. Wenz, *MTO1 mediates tissue specificity of OXPHOS defects via tRNA modification and translation optimization,*

which can be bypassed by dietary intervention. Hum Mol Genet, 2015. **24**(8): p. 2247-66.

[120] Becker, L., E. Kling, E. Schiller, R. Zeh, A. Schrewe, S.M. Holter, I. Mossbrugger, J. Calzada-Wack, V. Strecker, I. Wittig, I. Dumitru, T. Wenz, A. Bender, M. Aichler, D. Janik, F. Neff, A. Walch, L. Quintanilla-Fend, T. Floss, R. Bekeredjian, V. Gailus-Durner, H. Fuchs, W. Wurst, T. Meitinger, H. Prokisch, M.H. de Angelis, and T. Klopstock, *MTO1-deficient mouse model mirrors the human phenotype showing complex I defect and cardiomyopathy.* PLoS One, 2014. **9**(12): p. e114918.

[121] Posey, J.E., T. Harel, P. Liu, J.A. Rosenfeld, R.A. James, Z.H. Coban Akdemir, M. Walkiewicz, W. Bi, R. Xiao, Y. Ding, F. Xia, A.L. Beaudet, D.M. Muzny, R.A. Gibbs, E. Boerwinkle, C.M. Eng, V.R. Sutton, C.A. Shaw, S.E. Plon, Y. Yang, and J.R. Lupski, *Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation.* N Engl J Med, 2017. **376**(1): p. 21-31.

[122] Hildick-Smith, G.J., J.D. Cooney, C. Garone, L.S. Kremer, T.B. Haack, J.N. Thon, N. Miyata, D.S. Lieber, S.E. Calvo, H.O. Akman, Y.Y. Yien, N.C. Huston, D.S. Branco, D.I. Shah, M.L. Freedman, C.M. Koehler, J.E. Italiano, Jr., A. Merkenschlager, S. Beblo, T.M. Strom, T. Meitinger, P. Freisinger, M.A. Donati, H. Prokisch, V.K. Mootha, S. DiMauro, and B.H. Paw, *Macrocytic anemia and mitochondriopathy resulting from a defect in sideroflexin 4.* Am J Hum Genet, 2013. **93**(5): p. 906-14.

[123] Kremer, L.S., F. Distelmaier, B. Alhaddad, M. Hempel, A. Iuso, C. Kupper, C. Muhlhausen, R. Kovacs-Nagy, R. Satanovskij, E. Graf, R. Berutti, G. Eckstein, R. Durbin, S. Sauer, G.F. Hoffmann, T.M. Strom, R. Santer, T. Meitinger, T. Klopstock, H. Prokisch, and T.B. Haack, *Bi-allelic Truncating Mutations in TANGO2 Cause Infancy-Onset Recurrent Metabolic Crises with Encephalocardiomyopathy.* Am J Hum Genet, 2016. **98**(2): p. 358-62.

[124] Bard, F., L. Casano, A. Mallabiabarrena, E. Wallace, K. Saito, H. Kitayama, G. Guizzunti, Y. Hu, F. Wendler, R. Dasgupta, N. Perrimon, and V. Malhotra, *Functional genomics reveals genes involved in protein secretion and Golgi organization.* Nature, 2006. **439**(7076): p. 604-7.

[125] Maynard, T.M., D.W. Meechan, M.L. Dudevoir, D. Gopalakrishna, A.Z. Peters, C.C. Heindel, T.J. Sugimoto, Y. Wu, J.A. Lieberman, and A.S. Lamantia, *Mitochondrial localization and function of a subset of 22q11 deletion syndrome candidate genes.* Mol Cell Neurosci, 2008. **39**(3): p. 439-51.

[126] Lalani, S.R., P. Liu, J.A. Rosenfeld, L.B. Watkin, T. Chiang, M.S. Leduc, W. Zhu, Y. Ding, S. Pan, F. Vetrini, C.Y. Miyake, M. Shinawi, T. Gambin, M.K. Eldomery, Z.H. Akdemir, L. Emrick, Y. Wilnai, S. Schelley, M.K. Koenig, N. Memon, L.S. Farach, B.P. Coe, M. Azamian, P. Hernandez, G. Zapata, S.N. Jhangiani, D.M. Muzny, T. Lotze, G. Clark, A. Wilfong, H. Northrup, A. Adesina, C.A. Bacino, F. Scaglia, P.E. Bonnen, J. Crosson, J. Duis, G.H. Maegawa, D. Coman, A. Inwood, J. McGill, E. Boerwinkle, B. Graham, A. Beaudet, C.M. Eng, N.A. Hanchard, F. Xia, J.S. Orange, R.A. Gibbs, J.R. Lupski, and Y. Yang, *Recurrent Muscle Weakness with Rhabdomyolysis, Metabolic Crises, and Cardiac Arrhythmia Due to Bi-allelic TANGO2 Mutations.* Am J Hum Genet, 2016. **98**(2): p. 347-57.

[127] Haack, T.B., M. Gorza, K. Danhauser, J.A. Mayr, B. Haberberger, T. Wieland, L. Kremer, V. Strecker, E. Graf, Y. Memari, U. Ahting, R. Kopajtich, S.B. Wortmann, R.J. Rodenburg, U. Kotzaeridou, G.F. Hoffmann, W. Sperl, I. Wittig, E. Wilichowski, G. Schottmann, M. Schuelke, B. Plecko, U. Stephani, T.M. Strom, T. Meitinger, H. Prokisch, and P. Freisinger, *Phenotypic spectrum of eleven patients and five novel*

MTFMT mutations identified by exome sequencing and candidate gene screening. Mol Genet Metab, 2014. **111**(3): p. 342-52.

[128] Park, J.Y., P. Clark, E. Londin, M. Sponziello, L.J. Kricka, and P. Fortina, *Clinical exome performance for reporting secondary genetic findings.* Clin Chem, 2015. **61**(1): p. 213-20.

[129] Klein, C.J., S. Middha, X. Duan, Y. Wu, W.J. Litchy, W. Gu, P.J. Dyck, R.H. Gavrilova, D.I. Smith, J.P. Kocher, and P.J. Dyck, *Application of whole exome sequencing in undiagnosed inherited polyneuropathies.* J Neurol Neurosurg Psychiatry, 2014. **85**(11): p. 1265-72.

[130] Love, M.I., A. Mysickova, R. Sun, V. Kalscheuer, M. Vingron, and S.A. Haas, *Modeling read counts for CNV detection in exome sequencing data.* Stat Appl Genet Mol Biol, 2011. **10**(1).

[131] Madrigal, I., L. Rodriguez-Revenga, L. Armengol, E. Gonzalez, B. Rodriguez, C. Badenas, A. Sanchez, F. Martinez, M. Guitart, I. Fernandez, J.A. Arranz, M. Tejada, L.A. Perez-Jurado, X. Estivill, and M. Mila, *X-chromosome tiling path array detection of copy number variants in patients with chromosome X-linked mental retardation.* BMC Genomics, 2007. **8**: p. 443.

[132] Campbell, P.J., P.J. Stephens, E.D. Pleasance, S. O'Meara, H. Li, T. Santarius, L.A. Stebbings, C. Leroy, S. Edkins, C. Hardy, J.W. Teague, A. Menzies, I. Goodhead, D.J. Turner, C.M. Clee, M.A. Quail, A. Cox, C. Brown, R. Durbin, M.E. Hurles, P.A. Edwards, G.R. Bignell, M.R. Stratton, and P.A. Futreal, *Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing.* Nat Genet, 2008. **40**(6): p. 722-9.

[133] St Clair, D., *Copy number variation and schizophrenia.* Schizophr Bull, 2009. **35**(1): p. 9-12.

[134] Pang, A.W., J.R. MacDonald, D. Pinto, J. Wei, M.A. Rafiq, D.F. Conrad, H. Park, M.E. Hurles, C. Lee, J.C. Venter, E.F. Kirkness, S. Levy, L. Feuk, and S.W. Scherer, *Towards a comprehensive structural variation map of an individual human genome.* Genome Biol, 2010. **11**(5): p. R52.

[135] Conrad, D.F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T.D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C.H. Ihm, K. Kristiansson, D.G. Macarthur, J.R. Macdonald, I. Onyiah, A.W. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, C. Wellcome Trust Case Control, C. Tyler-Smith, N.P. Carter, C. Lee, S.W. Scherer, and M.E. Hurles, *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-12.

[136] Levy, S., G. Sutton, P.C. Ng, L. Feuk, A.L. Halpern, B.P. Walenz, N. Axelrod, J. Huang, E.F. Kirkness, G. Denisov, Y. Lin, J.R. MacDonald, A.W. Pang, M. Shago, T.B. Stockwell, A. Tsiamouri, V. Bafna, V. Bansal, S.A. Kravitz, D.A. Busam, K.Y. Beeson, T.C. McIntosh, K.A. Remington, J.F. Abril, J. Gill, J. Borman, Y.H. Rogers, M.E. Frazier, S.W. Scherer, R.L. Strausberg, and J.C. Venter, *The diploid genome sequence of an individual human.* PLoS Biol, 2007. **5**(10): p. e254.

[137] Veltman, J.A. and H.G. Brunner, *De novo mutations in human genetic disease.* Nat Rev Genet, 2012. **13**(8): p. 565-75.

[138] Acuna-Hidalgo, R., J.A. Veltman, and A. Hoischen, *New insights into the generation and role of de novo mutations in health and disease.* Genome Biol, 2016. **17**(1): p. 241.

[139] Hoischen, A., B.W. van Bon, C. Gilissen, P. Arts, B. van Lier, M. Steehouwer, P. de Vries, R. de Reuver, N. Wieskamp, G. Mortier, K. Devriendt, M.Z. Amorim, N.

Revencu, A. Kidd, M. Barbosa, A. Turner, J. Smith, C. Oley, A. Henderson, I.M. Hayes, E.M. Thompson, H.G. Brunner, B.B. de Vries, and J.A. Veltman, *De novo mutations of SETBP1 cause Schinzel-Giedion syndrome.* Nat Genet, 2010. **42**(6): p. 483-5.

[140] Ng, S.B., A.W. Bigham, K.J. Buckingham, M.C. Hannibal, M.J. McMillin, H.I. Gildersleeve, A.E. Beck, H.K. Tabor, G.M. Cooper, H.C. Mefford, C. Lee, E.H. Turner, J.D. Smith, M.J. Rieder, K. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D.A. Nickerson, M.J. Bamshad, and J. Shendure, *Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.* Nat Genet, 2010. **42**(9): p. 790-3.

[141] Hoischen, A., B.W. van Bon, B. Rodriguez-Santiago, C. Gilissen, L.E. Vissers, P. de Vries, I. Janssen, B. van Lier, R. Hastings, S.F. Smithson, R. Newbury-Ecob, S. Kjaergaard, J. Goodship, R. McGowan, D. Bartholdi, A. Rauch, M. Peippo, J.M. Cobben, D. Wieczorek, G. Gillessen-Kaesbach, J.A. Veltman, H.G. Brunner, and B.B. de Vries, *De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome.* Nat Genet, 2011. **43**(8): p. 729-31.

[142] Li, J., Y. Jiang, T. Wang, H. Chen, Q. Xie, Q. Shao, X. Ran, K. Xia, Z.S. Sun, and J. Wu, *mirTrios: an integrated pipeline for detection of de novo and rare inherited mutations from trios-based next-generation sequencing.* J Med Genet, 2015. **52**(4): p. 275-81.

[143] Lee, H., J.L. Deignan, N. Dorrani, S.P. Strom, S. Kantarci, F. Quintero-Rivera, K. Das, T. Toy, B. Harry, M. Yourshaw, M. Fox, B.L. Fogel, J.A. Martinez-Agosto, D.A. Wong, V.Y. Chang, P.B. Shieh, C.G. Palmer, K.M. Dipple, W.W. Grody, E. Vilain, and S.F. Nelson, *Clinical exome sequencing for genetic identification of rare Mendelian disorders.* JAMA, 2014. **312**(18): p. 1880-7.

[144] Yang, Y., D.M. Muzny, F. Xia, Z. Niu, R. Person, Y. Ding, P. Ward, A. Braxton, M. Wang, C. Buhay, N. Veeraraghavan, A. Hawes, T. Chiang, M. Leduc, J. Beuten, J. Zhang, W. He, J. Scull, A. Willis, M. Landsverk, W.J. Craigen, M.R. Bekheirnia, A. Stray-Pedersen, P. Liu, S. Wen, W. Alcaraz, H. Cui, M. Walkiewicz, J. Reid, M. Bainbridge, A. Patel, E. Boerwinkle, A.L. Beaudet, J.R. Lupski, S.E. Plon, R.A. Gibbs, and C.M. Eng, *Molecular findings among patients referred for clinical whole-exome sequencing.* JAMA, 2014. **312**(18): p. 1870-9.

[145] Posey, J.E., J.A. Rosenfeld, R.A. James, M. Bainbridge, Z. Niu, X. Wang, S. Dhar, W. Wiszniewski, Z.H. Akdemir, T. Gambin, F. Xia, R.E. Person, M. Walkiewicz, C.A. Shaw, V.R. Sutton, A.L. Beaudet, D. Muzny, C.M. Eng, Y. Yang, R.A. Gibbs, J.R. Lupski, E. Boerwinkle, and S.E. Plon, *Molecular diagnostic experience of whole-exome sequencing in adult patients.* Genet Med, 2016. **18**(7): p. 678-85.

[146] Meienberg, J., R. Bruggmann, K. Oexle, and G. Matyas, *Clinical sequencing: is WGS the better WES?* Hum Genet, 2016. **135**(3): p. 359-62.

[147] Stavropoulos, D.J., D. Merico, R. Jobling, S. Bowdin, N. Monfared, B. Thiruvahindrapuram, T. Nalpathamkalam, G. Pellecchia, R.K.C. Yuen, M.J. Szego, R.Z. Hayeems, R.Z. Shaul, M. Brudno, M. Girdea, B. Frey, B. Alipanahi, S. Ahmed, R. Babul-Hirji, R.B. Porras, M.T. Carter, L. Chad, A. Chaudhry, D. Chitayat, S.J. Doust, C. Cytrynbaum, L. Dupuis, R. Ejaz, L. Fishman, A. Guerin, B. Hashemi, M. Helal, S. Hewson, M. Inbar-Feigenberg, P. Kannu, N. Karp, R.H. Kim, J. Kronick, E. Liston, H. MacDonald, S. Mercimek-Mahmutoglu, R. Mendoza-Londono, E. Nasr, G. Nimmo, N. Parkinson, N. Quercia, J. Raiman, M. Roifman, A. Schulze, A. Shugar, C. Shuman, P. Sinajon, K. Siriwardena, R. Weksberg, G. Yoon, C. Carew, R. Erickson, R.A. Leach, R. Klein, P.N. Ray, M.S. Meyn, S.W. Scherer, R.D. Cohn, and C.R.

Marshall, *Whole-genome sequencing expands diagnostic utility and improves clinical management in paediatric medicine.* Npj Genomic Medicine, 2016. **1**: p. 15012.

[148] Genomes Project, C., G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, R.E. Handsaker, H.M. Kang, G.T. Marth, and G.A. McVean, *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.

[149] Li, X., Y. Kim, E.K. Tsang, J.R. Davis, F.N. Damani, C. Chiang, Z. Zappala, B.J. Strober, A.J. Scott, A. Ganna, J. Merker, I.M. Hall, A. Battle, and S.B. Montgomery, *The impact of rare variation on gene expression across tissues.* bioRxiv, 2016: p. 1-22.

[150] Zeng, Y., G. Wang, E. Yang, G. Ji, C.L. Brinkmeyer-Langford, and J.J. Cai, *Aberrant Gene Expression in Humans.* PLoS Genetics, 2015. **11**: p. 1-20.

[151] Guan, J., E. Yang, J. Yang, Y. Zeng, G. Ji, and J.J. Cai, *Exploiting aberrant mRNA expression in autism for gene discovery and diagnosis.* Human Genetics, 2016. **135**: p. 1-15.

[152] Zhao, J., I. Akinsanmi, D. Arafat, T.J. Cradick, C.M. Lee, S. Banskota, U.M. Marigorta, G. Bao, and G. Gibson, *A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood.* American Journal of Human Genetics, 2016. **98**: p. 299-309.

[153] McKean, D.M., J. Homsy, H. Wakimoto, N. Patel, J. Gorham, S.R. DePalma, J.S. Ware, S. Zaidi, W. Ma, N. Patel, R.P. Lifton, W.K. Chung, R. Kim, Y. Shen, M. Brueckner, E. Goldmuntz, A.J. Sharp, C.E. Seidman, B.D. Gelb, and J.G. Seidman, *Loss of RNA expression and allele-specific expression associated with congenital heart disease.* Nature Communications, 2016. **7**: p. 12824.

[154] Consortium, G.T., *Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-60.

[155] Albers, C.a., D.S. Paul, H. Schulze, K. Freson, J.C. Stephens, P.a. Smethurst, J.D. Jolley, A. Cvejic, M. Kostadima, P. Bertone, M.H. Breuning, N. Debili, P. Deloukas, R. Favier, J. Fiedler, C.M. Hobbs, N. Huang, M.E. Hurles, G. Kiddle, I. Krapels, P. Nurden, C.a.L. Ruivenkamp, J.G. Sambrook, K. Smith, D.L. Stemple, G. Strauss, C. Thys, C. van Geet, R. Newbury-Ecob, W.H. Ouwehand, and C. Ghevaert, *Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome.* Nature genetics, 2012. **44**: p. 435-9, S1-2.

[156] Chess, A., *Random and non-random monoallelic expression.* Neuropsychopharmacology, 2013. **38**(1): p. 55-61.

[157] Eckersley-Maslin, M.A. and D.L. Spector, *Random monoallelic expression: Regulating gene expression one allele at a time.* Trends in Genetics, 2014. **30**: p. 237-244.

[158] Reinius, B. and R. Sandberg, *Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation.* Nature Reviews Genetics, 2015. **16**: p. 653-664.

[159] Fahrner, J.A. and H.T. Bjornsson, *Mendelian disorders of the epigenetic machinery: tipping the balance of chromatin states.* Annu Rev Genomics Hum Genet, 2014. **15**: p. 269-93.

[160] Tazi, J., N. Bakkour, and S. Stamm, *Alternative splicing and disease*, in *Biochimica et Biophysica Acta - Molecular Basis of Disease*. 2009. p. 14-26.

[161] Singh, R.K. and T.A. Cooper, *Pre-mRNA splicing in disease and therapeutics.* Trends in Molecular Medicine, 2012. **18**: p. 472-482.
[162] Scotti, M.M. and M.S. Swanson, *RNA mis-splicing in disease.* Nature Reviews Genetics, 2015. **17**: p. 19-32.
[163] Xiong, H.Y., B. Alipanahi, L.J. Lee, H. Bretschneider, D. Merico, R.K.C. Yuen, Y. Hua, S. Gueroussov, H.S. Najafabadi, T.R. Hughes, Q. Morris, Y. Barash, A.R. Krainer, N. Jojic, S.W. Scherer, B.J. Blencowe, and B.J. Frey, *The human splicing code reveals new insights into the genetic determinants of disease.* Science, 2015. **347**: p. 1254806-1254806.
[164] Muntoni, F., S. Torelli, and A. Ferlini, *Dystrophin and mutations: one gene, several proteins, multiple phenotypes.* The Lancet. Neurology, 2003. **2**: p. 731-40.
[165] Gonorazky, H., M. Liang, B. Cummings, M. Lek, J. Micallef, C. Hawkins, R. Basran, R. Cohn, M.D. Wilson, D. MacArthur, C.R. Marshall, P.N. Ray, and J.J. Dowling, *RNAseq analysis for the diagnosis of muscular dystrophy.* Annals of clinical and translational neurology, 2016. **3**: p. 55-60.
[166] Morel, C.F., M.A. Thomas, H. Cao, C.H. O'Neil, J.G. Pickering, W.D. Foulkes, and R.A. Hegele, *A LMNA Splicing Mutation in Two Sisters with Severe Dunnigan-Type Familial Partial Lipodystrophy Type 2.* The Journal of Clinical Endocrinology & Metabolism, 2006. **91**: p. 2689-2695.
[167] Qu, Y.-j., J.-l. Bai, Y.-y. Cao, W.-h. Zhang, H. Wang, Y.-w. Jin, and F. Song, *A rare variant (c.863G>T) in exon 7 of SMN1 disrupts mRNA splicing and is responsible for spinal muscular atrophy.* European Journal of Human Genetics, 2016. **24**: p. 864-870.
[168] Sibley, C.R., L. Blazquez, and J. Ule, *Lessons from non-canonical splicing.* Nat Rev Genet, 2016. **17**(7): p. 407-21.
[169] Cummings, B.B., J.L. Marshall, T. Tukiainen, M. Lek, S. Donkervoort, A.R. Foley, V. Bolduc, L.B. Waddell, S.A. Sandaradura, G.L. O'Grady, E. Estrella, H.M. Reddy, F. Zhao, B. Weisburd, K.J. Karczewski, A.H. O'Donnell-Luria, D. Birnbaum, A. Sarkozy, Y. Hu, H. Gonorazky, K. Claeys, H. Joshi, A. Bournazos, E.C. Oates, R. Ghaoui, M.R. Davis, N.G. Laing, A. Topf, C. Genotype-Tissue Expression, P.B. Kang, A.H. Beggs, K.N. North, V. Straub, J.J. Dowling, F. Muntoni, N.F. Clarke, S.T. Cooper, C.G. Bonnemann, and D.G. MacArthur, *Improving genetic diagnosis in Mendelian disease with transcriptome sequencing.* Sci Transl Med, 2017. **9**(386).
[170] Li, Y.I., B. van de Geijn, A. Raj, D.A. Knowles, A.A. Petti, D. Golan, Y. Gilad, and J.K. Pritchard, *RNA splicing is a primary link between genetic variation and disease.* Science, 2016. **352**(6285): p. 600-4.
[171] Li, Y.I., D.A. Knowles, and J.K. Pritchard, *LeafCutter: Annotation-free quantification of RNA splicing.* bioRxiv, 2016: p. 044107.
[172] Vafai, S.B. and V.K. Mootha, *Mitochondrial disorders as windows into an ancient organelle.* Nature, 2012. **491**: p. 374-83.
[173] Holzerova, E., K. Danhauser, T.B. Haack, L.S. Kremer, M. Melcher, I. Ingold, S. Kobayashi, C. Terrile, P. Wolf, J. Schaper, E. Mayatepek, F. Baertling, J.P. Friedmann Angeli, M. Conrad, T.M. Strom, T. Meitinger, H. Prokisch, and F. Distelmaier, *Human thioredoxin 2 deficiency impairs mitochondrial redox homeostasis and causes early-onset neurodegeneration.* Brain : a journal of neurology, 2016. **139**: p. 346-54.
[174] Wortmann, S.B., 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.* Neuropediatrics, 2017.

[175] Abela, L., R. Spiegel, L.M. Crowther, A. Klein, K. Steindl, S.M. Papuc, P. Joset, Y. Zehavi, A. Rauch, B. Plecko, and T.L. Simmons, *Plasma metabolomics reveals a diagnostic metabolic fingerprint for mitochondrial aconitase (ACO2) deficiency*. PLoS One, 2017. **12**(5): p. e0176363.

Figures

Acknowledgements

I am incredibly thankful to Dr. Holger Prokisch for the possibility to work in his group, his strong support, brilliant advice, and encouragement during my PhD studies. I am deeply grateful to all the inspiring discussions and guidance along the way.

Furthermore I want to thank all the group members (Robert Kopajtich, Arcangela Iuso, Birgit Repp, Marieta Borzes, Katharina Schramm, Caterina Terrile, Caroline Biagosch, Eliska Konarikova, Mirjana Gusic, Michael Färberböck, Agnieszka Nadel, Elisa Mastantuono) and other members of the Institute of Human Genetics (Susanne Diener, Thomas Wieland, Thomas Schwarzmayr, Riccardo Berutti) who were always supporting me with advice, help, and comfort in struggling times and created such an unbelievably enjoyable working environment.

A special thanks to Martina Kuhnert for all the administrative help and Thomas Meitinger for financial support, especially for my extensive conference participations.

I thank all my collaboration partners and coauthors without whom this work would not have been possible.

I would like to show my greatest appreciation to Hauke Hillen for his unconditional understanding and care.

Most importantly, I want to express my deepest gratitude and heartfelt admiration to my family. You inspire me with your strength, love, and trust.

Anhang I

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die bei der promotionsführenden Einrichtung
Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

der TUM zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:
Discovery and validation of coding and non-coding pathogenic variants in mitochondrial disorders

in
Entwicklungsgenetik
Fakultät, Institut, Lehrstuhl, Klinik, Krankenhaus, Abteilung

unter der Anleitung und Betreuung durch: PD Dr. T. Floss ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Ab. 6 und 7 Satz 2 angebotenen Hilfsmittel benutzt habe.

☒ Ich habe keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht, oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

☒ Ich habe die Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.

☐ Die vollständige Dissertation wurde in _____ veröffentlicht.
Die promotionsführende Einrichtung

_____

hat der Veröffentlichung zugestimmt.

☒ Ich habe den angestrebten Doktorgrad noch nicht erworben und bin nicht in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.

☐ Ich habe bereits am _____ bei der Fakultät für _____ der Hochschule _____ unter Vorlage einer Dissertation mit dem Thema _____ die Zulassung zur Promotion beantragt mit dem Ergebnis: _____

Die öffentlich zugängliche Promotionsordnung der TUM ist mir bekannt, insbesondere habe ich die Bedeutung von § 28 (Nichtigkeit der Promotion) und § 29 (Entzug des Doktorgrades) zur Kenntnis genommen. Ich bin mir der Konsequenzen einer falschen Eidesstattlichen Erklärung bewusst.

Mit der Aufnahme meiner personenbezogenen Daten in die Alumni-Datei bei der TUM bin ich

☒ einverstanden, ☐ nicht einverstanden.

_____
Ort, Datum, Unterschrift