



TECHNISCHE UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR MEDIZIN

INSTITUT FÜR HUMANGENETIK

EXPERIMENTELLE MEDIZIN

Role of coding and non-coding variants in mitochondrial disease genes

Mirjana Gusic

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Radu Roland Rad

Prüfer der Dissertation: 1. Prof. Dr. Thomas Meitinger
2. Prof. Dr. Julien Gagneur

Die Dissertation wurde am 09.10.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 13.04.2021 angenommen.

“Никад се ништа не понавља у историји људских бића, све што се на први поглед чини да је исто једва да је слично; сваки је човек звезда за себе, све се догађа увек и никад, све се понавља бескрајно и непоновљиво.”

Данило Киш (1935-1989)

“Nothing ever repeats in the history of mankind, everything that at first glance seems to be the same is hardly similar; every man is a star for himself, everything happens always and never, everything is repeated endlessly and unrepeatable.”

Danilo Kiš (1935-1989)

Abstract

After whole exome sequencing, around half of mitochondrial disease patients remain without genetic diagnosis. Focusing on them, I described two novel disease genes, encoding a Krebs cycle enzyme and a subunit of mitochondrial complex III. In addition, I performed RNA-sequencing as a complementary tool to whole exome sequencing, increasing the diagnostic yield and improving variant annotation. This expanded the genetic spectrum of disease and encourages implementation of RNA-sequencing in future diagnostics.

Abstrakt

Nach der Exom-Sequenzierung bleibt die Hälfte der Patienten mit mitochondrialen Erkrankungen ohne genetische Diagnose. Ich konzentrierte mich auf sie und beschrieb zwei neue Krankheitsgene, die für ein Krebs-Zyklus-Enzym bzw. eine Untereinheit des Mitochondrien-Komplexes III kodieren. Zusätzlich führte ich eine RNA-Sequenzierung als komplementäres Werkzeug zur Exom-Sequenzierung durch, um die diagnostische Ausbeute zu erhöhen und die Annotation von Varianten zu verbessern. Dies erweiterte das genetische Krankheitsspektrum und förderte die Implementierung der RNA-Sequenzierung in der zukünftigen Diagnostik.

Summary

Mitochondrial diseases present a heterogeneous group of genetic disorders characterized by a defect in pyruvate oxidation, the Krebs cycle, and/or oxidative phosphorylation. Affecting at least 1 in 5000 people, they are the largest group of inherited metabolic disorders. They are characterized by a remarkable variability in clinical presentations, often multisystemic, with significant levels of morbidity and mortality, and lack of curative treatments. The genetics of mitochondrial diseases is characterized by the involvement of both the nuclear and mitochondrial genome, all modes of inheritance, and often poor genotype-phenotype correlation. So far, 338 genes have been associated with a mitochondrial disease, and this number continues to grow. Such phenotype and genotype heterogeneity makes mitochondrial diseases amendable for systematic, unbiased diagnostic approaches, best represented by the whole exome sequencing (WES). Despite WES capturing only protein-coding regions of the genome, it covers over 85% of reported Mendelian disease-causing variants and is still a useful first-tier. Although the implementation of WES has had a revolutionary impact on the discovery of mitochondrial disease genes and on diagnostics, it still leaves about half of the patients without a diagnosis. Such diagnostic yield reflects the main challenge of human genetics: variant prioritization and lack of knowledge in non-coding regions.

During my Ph.D., my focus was on the cases which remained inconclusive following WES, where I implemented two approaches in the hope to establish a genetic diagnosis. In the first approach, I contributed to the discovery of novel mitochondrial disease genes. This was achieved by reanalysis of WES data, and identification and functional validation of variants in six novel disease genes. In my thesis, I describe in detail my contributions to two of such studies, *MDH2* and *UQCRCF1*. I additionally functionally validated variants in two known disease genes. Soon after I started with the first approach, I turned to the second, more comprehensive approach, where I have performed RNA-sequencing (RNA-seq) on a cohort of patients clinically suspected to suffer from a mitochondrial disease. This approach, described as my third study, enabled a systematic overview of cellular transcriptome and detection of transcript perturbation, providing functional validation of variants and a genetic diagnosis to 16% of previously WES-inconclusive cases.

In my first study, WES reanalysis revealed two heterozygous predicted as deleterious variants in a gene encoding a mitochondrial Krebs cycle enzyme, *MDH2*, previously not associated with mitochondrial disease. This case was connected for a joined analysis to two additional unrelated patients with similar neurological presentations and biallelic *MDH2* variants. To evaluate the causality of discovered variants, I performed functional studies. First, a Western blot and enzymatic assay in patient's skin-derived fibroblast cell lines demonstrated the deleterious effect of both variants on the protein level and activity. Second, a measurement of cellular oxygen consumption revealed the normal activity of oxidation phosphorylation complexes. Still, measurements of metabolites revealed disrupted ratios of Krebs cycle metabolites. Finally, by Sanger sequencing I confirmed the variants segregated in the family. This work established *MDH2* as a novel mitochondrial disease-causing gene.

In my second study, WES reanalysis of patients with cardiac phenotypes revealed in one patient a homozygous splice site variant in a gene encoding a mitochondrial complex III enzyme, *UQCRFS1*, previously unassociated with mitochondrial disease. This case was connected to an additional unrelated patient with similar cardiac and hair abnormalities and biallelic *UQCRFS1* variants. Functional studies were performed on patient-derived skin fibroblasts to validate the discovered variants' pathogenicity. RNA analysis confirmed the predicted effect of the variant on splicing. Western blot and immunostainings confirmed the deleterious effect of splicing defect on the UQCRFS1 protein. Blue native polyacrylamide gel electrophoresis revealed that the reduction of UQCRFS1 abolishes mature complex III. Moreover, I measured cellular oxygen consumption and found defects in cellular respiration. These defects were subsequently rescued by overexpression of wild-type *UQCRFS1* in fibroblasts by lentiviral transduction. Altogether, these experiments, together with the genetic data, established *UQCRFS1* as a novel mitochondrial disease-causing gene.

As my second approach and third study, I took part in generating a large RNA-seq compendium, composed of 310 samples from skin-derived fibroblasts from patients clinically suspected to suffer from mitochondrial disease. As RNA-seq provides information on both transcript abundance and sequence, the aim was to implement RNA-seq as a complementary tool to WES in diagnostics. Analysis of data was performed in close collaboration with the Department of informatics at the Technical University of Munich, who developed bioinformatics tools to systematically search for transcript events that can prove to be disease-causative. Implementation of these tools revealed four aberrantly expressed genes, 23 aberrant splicing events, and one mono-allelically expressed rare variant per sample. Together with RNA-seq variant calling, this approach simplified manual inspection for diagnostic purposes and search for causative variants in WES-elusive regions. The approach provided a genetic diagnosis in 16% of WES-inconclusive cases (34 out of 214), and overall established RNA-seq as a tool for functional validation and (re)prioritization of variants in both coding and non-coding regions.

Zusammenfassung

Mitochondriale Erkrankungen stellen eine heterogene Gruppe genetischer Störungen dar, welche durch Defekte der Pyruvatoxidation, des Krebszyklus und / oder der oxidativen Phosphorylierung gekennzeichnet sind. Mit einer Prävalenz von 1:5000 stellen sie die größte Gruppe erblicher Stoffwechselstörungen dar. Sie zeichnen sich durch eine bemerkenswerte Variabilität der klinischen Erscheinungsformen aus, häufig multisystemisch, ein signifikantes Maß an Morbidität und Mortalität sowie kaum verfügbarer Therapiemöglichkeiten. Ihre Genetik ist gekennzeichnet durch die Beteiligung sowohl des nuklearen als auch des mitochondrialen Genoms, aller Vererbungsarten und häufig einer unzureichenden Genotyp-Phänotyp-Korrelation. Bislang wurden 338 Gene mit einer mitochondrialen Erkrankung assoziiert, und diese Zahl steigt stetig an. Eine solche Heterogenität von Phänotyp und Genotyp macht diese Erkrankungen für systematische, unvoreingenommene diagnostische Ansätze ergänzungsfähig, die am besten durch die genomweite Exomsequenzierung (WES) dargestellt durchgeführt werden. Obgleich durch WES nur proteinkodierende Regionen des Genoms abgedeckt werden, erfasst es über 85% der bekannten Varianten, die mit genetischen Erkrankungen assoziiert sind, und stellt somit immer noch die Methode der Wahl dar. Obwohl die Implementierung von WES einen revolutionären Beitrag zur Entdeckung mitochondrialer Krankheitsgene und zur Diagnostik geleistet hat, erhält etwa Hälfte aller Patienten keine Diagnose. Diese diagnostische Ausbeute spiegelt die primäre Herausforderung der Humangenetik: das Priorisieren von Varianten und die fehlende Annotation von nichtkodierenden Regionen.

Im Rahmen meiner Promotion habe ich mich auf die Fälle fokussiert, die nach der WES Analyse ungeklärt blieben, und habe zwei Ansätze verfolgt, um eine genetische Diagnose für diese Fälle zu erstellen. Im ersten Ansatz habe ich zur Entdeckung neuer mitochondrialer Krankheitsgene beigetragen. Dies wurde durch eine erneute Analyse der WES-Daten und der Identifizierung und funktionellen Validierung von Varianten in sechs neuen Krankheitsgenen erreicht. In meiner Arbeit beschreibe ich mein Beitrag zu zwei solcher Studien, *MDH2* und *UQCRFS1*. Des Weiteren habe ich Varianten in zwei bekannten Krankheitsgenen funktionell validiert. Kurz nachdem ich mit dem ersten Ansatz begonnen hatte, widmete ich mich zusätzlich einem zweiten, umfassenderen Ansatz, in dem ich bei unserer Patientenkohorte, mit bestehendem klinischen Verdacht auf eine mitochondriale Erkrankung, eine RNA-Sequenzierung (RNA-seq) durchgeführt habe. Dieser Ansatz, meine dritte Studie, ermöglichte einen systematischen Überblick über das zelluläre Transkriptom und den Nachweis von Transkriptstörungen und ermöglichte in 16% der ungeklärten WES Fälle eine funktionelle Validierung von Varianten und schließlich eine genetische Diagnose.

In meiner ersten Studie deckte die Reanalyse der WES Daten zwei heterozygote, schädlich vorhergesagte Varianten in einem Gen auf, das für ein mitochondriales Enzym des Krebszyklus, *MDH2*, kodiert und zuvor nicht mit einer mitochondrialen Erkrankung in Verbindung gebracht worden ist. Dieser Fall wurde für eine gemeinsame Analyse mit zwei weiteren nicht verwandten Patienten mit ähnlichen neurologischen Darstellungen und

biallelischen *MDH2*-Varianten verbunden. Um die Kausalität der gefundenen Varianten zu bewerten, wurden funktionelle Studien durchgeführt. Mit Hilfe von Western Blot und enzymatische Messungen in Haut Fibroblastenzelllinien der Patienten konnte die schädliche Wirkung beider Varianten in Bezug auf die Proteinstabilität und Enzymaktivität gezeigt werden. Durch die Messung des zellulären Sauerstoffverbrauchs wurde eine normale Aktivität der mitochondrialen Atmungskette identifiziert. Messungen von Metaboliten ergaben jedoch ein gestörtes Verhältniss von Krebszyklus-Metaboliten. Die Sanger-Sequenzierung bestätigte weiter, dass die Varianten in der Familie segregierten. Diese Arbeit etablierte *MDH2* als neuartiges mitochondriales Krankheitsgen.

In meiner zweiten Studie ergab die Reanalyse der WES Daten von Patienten mit Herzphänotypen bei einem Patienten eine homozygote Variante der Spleißstelle in dem Gen *UQCRFS1*, das für eine Untereinheit des mitochondrialen Atmungsketten Komplex-III kodiert und noch nicht mit einer mitochondrialen Erkrankung assoziiert worden ist. Dieser Fall wurde gemeinsam mit einem zusätzlichen nicht verwandten Patienten mit ähnlichen Herz- und Haaranomalien analysiert, bei welchem ebenfalls biallelische Varianten in *UQCRFS1* identifiziert wurden. Funktionelle Studien wurden an Hautfibroblasten der Patienten durchgeführt, um die Pathogenität der detektierten Varianten zu validieren. Die RNA-Analyse bestätigte den vorhergesagten Effekt der Variante auf das Spleißen. Western Blot- und Immunfärbungen bestätigten den pathogenen Effekt des Spleißdefekts auf das *UQCRFS1*-Protein. Die Blue Native-Polyacrylamid-Gelelektrophorese ergab, dass die Reduktion von *UQCRFS1* Protein zu einer verminderten Menge an assembliertem Komplex III führt. Der Sauerstoffverbrauch der Zellen wurde gemessen und Defekte in der zellulären Respiration festgestellt. Diese Defekte wurden durch die Überexpression von wildtyp-*UQCRFS1* in Fibroblasten durch lentivirale Transduktion behoben. Insgesamt haben diese Experimente zusammen mit genetischen Daten *UQCRFS1* als neuartiges Gen für mitochondriale Erkrankungen etabliert.

Als zweiten Ansatz und dritte Studie erstellte ich ein großes RNA-seq-Kompendium, das aus 310 Proben von Hautfibroblasten von Patienten bestand, bei denen der Verdacht auf eine mitochondriale Erkrankung besteht. Da RNA-seq sowohl Informationen zur Transkripthäufigkeit als auch zur Sequenz liefert, war das Ziel, RNA-seq als komplementäres Werkzeug zu WES in der Diagnostik zu implementieren. Die Datenanalyse wurde in enger Zusammenarbeit mit dem Institut für Informatik der Technischen Universität München durchgeführt, welches bioinformatische Tools entwickelte, die systematisch nach potentiell krankheitsverursachenden Transkriptionsereignissen suchen. Die Anwendung dieser Methoden ergab vier aberrant exprimierte Gene, 23 aberrante Spleißereignisse und eine monoallelisch exprimierte seltene Variante pro Probe. Zusammen mit der Identifizierung von Varianten in den RNA-seq Daten, vereinfachte es die manuelle Inspektion zu diagnostischen Zwecken und die Suche nach ursächlichen Varianten in Regionen welche durch WES schlecht/nicht abgedeckt sind. Ein solcher Ansatz lieferte eine genetische Diagnose in 16% der WES-nicht eindeutigen Fälle (34 von 218) und hat somit RNA-Sequenzierung als Werkzeug zur funktionellen Validierung und (Re-) Priorisierung von Varianten sowohl in kodierenden als auch in nicht kodierenden Regionen etabliert

Contents

Abstract.....	IV
Summary.....	V
Contents.....	IX
List of abbreviations.....	XII
List of figures.....	XVI
List of tables.....	XVIII
List of publications.....	XIX
Peer-reviewed publications included in the thesis.....	XIX
Peer-reviewed publications not included in the thesis.....	XIX
1. Introduction.....	1
1.1. Mitochondria.....	1
1.1.1. Mitochondrial functions.....	1
1.1.2. Mitochondrial genetics.....	2
1.2. Mitochondrial diseases.....	5
1.2.1. Prevalence.....	6
1.2.2. Clinical presentation and diagnosis.....	6
1.2.3. Metabolic diagnosis- biomarkers.....	8
1.2.4. Biochemical and histochemical diagnosis.....	9
1.2.5. Genetic diagnosis.....	10
1.3. WES analysis.....	16
1.3.1. Variant detection.....	16
1.3.2. Variant annotation.....	17
1.3.3. Variant prioritization and validation.....	18
1.3.4. Validation of VUS.....	21
1.3.5. Reasons behind inconclusive WES analysis.....	22
1.4. Krebs cycle and its deficiencies.....	23
1.5. OXPHOS, complex III and its deficiencies.....	27
1.6. RNA-sequencing.....	31
1.6.1. Principles and advances.....	31
1.6.2. RNA-seq as a complementary tool in molecular diagnostics.....	32
1.7. Treatment of mitochondrial diseases.....	38

1.7.1.	Symptomatic treatments.....	38
1.7.2.	Strategies for the development of specific therapies	39
2.	Aim	42
3.	Material and methods.....	44
3.1.	Material	44
3.1.1.	Nucleic acids.....	44
3.1.2.	Cell lines	44
3.1.3.	Antibodies and the protein ladder	45
3.1.4.	Chemicals and solutions	46
3.2.	Methods.....	46
3.2.1.	Cell culture.....	46
3.2.2.	Bacterial culture and techniques	47
3.2.3.	DNA and RNA analysis.....	49
3.2.4.	Protein analytics.....	50
3.2.5.	Biochemical measurements	52
3.2.6.	Sequencing methods	53
3.2.7.	Functional complementation by lentiviral transduction	55
3.2.8.	Statistical analysis.....	58
3.2.9.	Online resources.....	58
4.	Results.....	59
4.1.	<i>MDH2</i> variants cause severe neurological presentations.....	59
4.1.1.	Case reports.....	59
4.1.2.	WES reveals biallelic variants in <i>MDH2</i> in three unrelated families	61
4.1.3.	<i>MDH2</i> in detail	62
4.1.4.	Observed <i>MDH2</i> variants are predicted as deleterious.....	62
4.1.5.	<i>MDH2</i> variants causes the loss of <i>MDH2</i> and its enzymatic activity.....	63
4.1.6.	Loss of <i>MDH2</i> does not affect fibroblasts respiration.....	64
4.1.7.	<i>MDH2</i> variants in P1 cause disruption in Krebs cycle metabolites.....	65
4.2.	<i>UQCRFS1</i> variants cause complex III deficiency, cardiomyopathy and <i>alopecia totalis</i>	66
4.2.1.	Case reports.....	66
4.2.2.	WES reveals biallelic variants in <i>UQCRFS1</i> in two unrelated families.....	69

4.2.3.	<i>UQCRFS1</i> in detail	69
4.2.4.	Observed <i>UQCRFS1</i> variants are predicted as deleterious.....	70
4.2.5.	RNA analysis of P1 fibroblasts reveals a splice defect.....	71
4.2.6.	Patients' fibroblasts exhibit impaired mitochondrial respiration.....	72
4.2.7.	The overexpression of wild-type <i>UQCRFS1</i> in patients' fibroblasts restores mitochondrial respiration	73
4.2.8.	<i>UQCRFS1</i> variants lead to a loss of UQCRFS1	74
4.2.9.	UQCRFS1 depletion affects complex III.....	75
4.2.10.	UQCRFS1 depletion is confirmed by immunofluorescence.....	75
4.3.	RNA-seq as a complementary tool in diagnostics	77
4.3.1.	RNA-Seq compendium for rare disease diagnostics	77
4.3.2.	Systematic analysis of RNA-seq data	78
4.3.3.	Selected case studies	80
4.3.4.	Summary of RNA-seq findings in diagnostic setting	91
5.	Discussion.....	98
5.1.	Disease gene discovery and functional validation	98
5.1.1.	<i>MDH2</i> as a novel mitochondrial disease gene	102
5.1.2.	<i>UQCRFS1</i> as a novel mitochondrial disease gene.....	104
5.2.	Genetic diagnosis of mitochondrial diseases: state-of-the-art.....	106
5.2.1.	Whole exome sequencing- benefits and drawbacks	106
5.2.2.	Whole genome sequencing- future challenges	109
5.2.3.	Towards the OMICS era: complementary tools to DNA-analysis	109
6.	Future perspectives	115
6.1.	New sequencing technologies	115
6.1.1.	Long read sequencing	115
6.1.2.	Single-cell sequencing	115
6.2.	Emerging non-coding functionality	116
6.3.	Beyond Mendelian inheritance.....	117
7.	References.....	119
	Acknowledgements.....	157

List of abbreviations

AA	amino acid
AAV	adeno-associated virus
ACMG	American College of Medical Genetics
AD	autosomal dominant
AR	autosomal recessive
ATP	adenosine triphosphate
BN-PAGE	blue native polyacrylamide gel electrophoresis
Bp	base pair
BWA	Burrows-Wheeler Aligner
c.	sequence position of cDNA
C-terminus	carboxy-terminus
CI-V, RCCI-V	respiratory chain complex I-V
cDNA	complementary DNA
Chr	chromosome
CNS	central nervous system
CNV	copy number variant
CoA	coenzyme A
CoQ ₁₀	ubiquinone, coenzyme Q(10)
COX	cytochrome c oxidase
CS	citrate synthase
CSF	cerebrospinal fluid
cyt c	cytochrome c
Da	Dalton
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
DSB	double-strand break
<i>E. coli</i>	<i>Escherichia coli</i>
e.g.	exempli gratia
ECL	enhanced chemiluminiscence
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
et al.	et alii
ETC	electron transport chain
eQTL	expression quantitative trait loci
FASTQ	file format for representing sequencing reads
FBS	fetal bovine serum
FC	fold change

FCCP	carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone
Fe-S	iron-sulfur
FPKM	fragments per kilobase per millions of reads
gnomAD	Genome Aggregation Database
GTE _x	Genotype-Tissue Expression Project
HEK293	human embryonic kidney 293 cells
hg19	human genome assembly GRCh37, UCSC
HGMD	Human Gene Mutation Database
HIF-1 α	hypoxia inducible factor 1 alpha (HIF-1 α)
HPLC H ₂ O	high-performance liquid chromatography water
HRP	horseradish peroxidase
IGV	Integrative Genomics Viewer
IMM, IM	inner mitochondrial membrane
IMS	intermembrane space
indel	small insertion and deletion variation
iPSC	induced pluripotent stem cells
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
kb	kilobase
kDa	kilodalton
KO	knockout
l	litre
LHON	Leber hereditary optic neuropathy
lncRNA	long non-coding RNA
LoF	loss-of-function
LRS	long read sequencing
m.	mitochondrial DNA sequence position
MADD	multiple acyl-CoA dehydrogenation deficiency
MAE	mono-allelic expression
MAF	minor allele frequency
Mb	mega base
MNVs	multi-nucleotide variants
MNGIE	mitochondrial neurogastrointestinal encephalomyopathy
MRI	magnetic resonance imaging
mRNA	messenger RNA
mtDNA	mitochondrial DNA
MTS	mitochondrial targeting sequence
N-terminus	amino-terminus
NA	not available, not assessed
NAD	nicotine adenine dinucleotide
NCBI	National Center for Biotechnology Information
nDNA	nuclear, genomic DNA
NGS	next generation sequencing

NHDF	normal human dermal fibroblasts
NM	RefSeq mRNA sequence
NMD	nonsense mediated decay
NP	RefSeq protein sequence
nt	nucleotides
OMIM	Online Mendelian Inheritance in Man
OMM, OM	outer mitochondrial membrane
ON	overnight
ORF	open reading frame
OXPPOS	oxidative phosphorylation
p.	protein sequence position
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PDH	pyruvate dehydrogenase
PE	pair-end
POLG	DNA polymerase gamma
poly(A)	polyadenylation
PTC	premature stop codon
PTV	protein-truncating variant
PVDF	polyvinylidene fluoride
Q	ubiquinone
QH ₂	ubiquinol
RefSeq	NCBI Reference Sequence Database
RCC	respiratory chain complexes
RFLP	restriction fragment length polymorphisms
RIN	RNA integrity number
RMA, StMA	random or stochastic mono-allelic expression
RNA	ribonucleic acid
RNA-seq	RNA-sequencing, transcriptomics
ROS	reactive oxygen species
rRNA	ribosomal RNA
RT	room temperature
RT-PCR	reverse transcriptase PCR
rxn	reaction
SCS	succinyl-CoA synthetase
SD	standard deviation
SDH	succinate dehydrogenase
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	standard error of the mean
SNP	single nucleotide polymorphism
SNV	single nucleotide variation
SV	structural variation

TBS-T	Tris-buffered saline plus Tween 20
TCA	tricarboxylic acid
T _m	melting temperature
TMT	tandem mass tag
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
tRNA	transfer RNA
U	unit
UCSC	University of California, Santa Cruz
uORF	upstream open reading frame
UTR	untranslated region
v/v	volume per volume
VUS	variant of unknown significance
w/v	weight per volume
WES	whole exome sequencing
WGS	whole genome sequencing
wt	wild-type
α-KGDH	alpha-ketoglutarate dehydrogenase

List of figures

Figure 1.	Graphical representation of a mitochondrion and its major functions.....	2
Figure 2.	Graphical representation of a human mtDNA.....	3
Figure 3.	Changes in heteroplasmy and its contributors.....	4
Figure 4.	Phenotypic spectrum of mitochondrial diseases.....	7
Figure 5.	Described mitochondrial disease genes categorized by their functional roles.....	12
Figure 6.	Diagnostic rate of WES in suspected mitochondrial disease across nine cohorts.....	15
Figure 7.	ACMG's proposed evidence framework for evaluating variant classification.....	19
Figure 8.	Summary of reported pathogenic variants in ClinVar.....	20
Figure 9.	WES analysis workflow.....	21
Figure 10.	Krebs cycle.....	24
Figure 11.	The mammalian OXPHOS-ETC.....	28
Figure 12.	Complex III assembly model.....	29
Figure 13.	Aberrant splicing as a cause of disease.....	36
Figure 14.	P3 inherited variants from each parent.....	61
Figure 15.	MDH2.....	62
Figure 16.	Effect of <i>MDH2</i> variants on gene and protein.....	63
Figure 17.	<i>MDH2</i> variants affect protein levels.....	64
Figure 18.	<i>MDH2</i> variants affect MDH2 enzymatic activity.....	64
Figure 19.	<i>MDH2</i> variants do not impact mitochondrial respiration.....	65
Figure 20.	<i>MDH2</i> variants disrupt Krebs cycle metabolites.....	65
Figure 21.	<i>UQCERS1</i> variants segregate in two unrelated families.....	69
Figure 22.	<i>UQCERS1</i>	70
Figure 23.	Effect of <i>UQCERS1</i> variants on gene and protein.....	71
Figure 24.	RNA analysis shows aberrant splicing of <i>UQCERS1</i> in P1.....	72
Figure 25.	Microscale respirometry analysis reveals defect in mitochondrial respiration in patients' fibroblasts.....	73
Figure 26.	Microscale respirometry analysis reveals rescue of mitochondrial respiration in patients' fibroblasts upon lentiviral transduction.....	74
Figure 27.	Depletion of <i>UQCERS1</i> in patients' fibroblasts and its restoration upon lentiviral transduction in P2.....	75
Figure 28.	Disruption of CIII in patients' fibroblasts and its restoration upon lentiviral transduction in P2.....	75
Figure 29.	Immunofluorescent stainings confirm depletion of <i>UQCERS1</i> in patients' fibroblasts and its restoration upon lentiviral transduction.....	76
Figure 30.	Median of outliers per sample per transcript event.....	78
Figure 31.	RNA-seq variant calling.....	80
Figure 32.	RNA-seq prioritizes homozygous <i>UFMI</i> variant.....	81
Figure 33.	Synonymous <i>TWNK</i> variant affects expression and splicing.....	82

Figure 34.	Mono-allelic expression of a deleterious <i>NFUI</i> variant.....	83
Figure 35.	WGS of case 3 identifies cause of depletion of the second allele.....	84
Figure 36.	Deep intronic variant creates a cryptic exon in <i>NDUFAF5</i>	85
Figure 37.	RNA-seq allows quantification of different transcript isoforms.....	87
Figure 38.	Defect in <i>MRPL44</i> affects mitochondrial ribosome and OXPHOS.....	88
Figure 39.	MNVs cause aberrant splicing in <i>DARS2</i>	89
Figure 40.	RNA-seq prevents false diagnosis.	90
Figure 41.	OUTRIDER calls dominant disease genes as outliers.....	90
Figure 42.	Summary of the RNA-seq diagnostic pipeline and diagnostic yield of the cohort.....	91
Figure 43.	Disease-causing and candidate RNA-defects across the RNA-seq cohort..	96
Figure 44.	RNA-seq enables validation, (re)prioritization and discovery of pathogenic variants.....	97
Figure 45.	Number of mitochondrial disease genes over years.....	99
Figure 46.	Yearly discovery rate of the mitochondrial-disease genes with the method of identification.....	100
Figure 47.	Proposed model for a functional splitting of the Krebs cycle into two complementary mini-cycles.	102
Figure 48.	A model depicting CIII as the O ₂ sensor underlying the HPV response....	106
Figure 49.	Summary of reasons behind inconclusive WES, with complementary approaches.....	107
Figure 50.	Possible effect of variants assessed by RNA-seq.....	111
Figure 51.	Proposed OMICS approach for genetic diagnostics based on central dogma of molecular biology.	114

List of tables

Table 1.	Clinical features of reported patients with Krebs cycle deficiencies.....	26
Table 2.	Clinical features of reported patients with complex III deficiencies.....	30
Table 3.	Overview of published RNA-seq-based diagnostic studies.....	33
Table 4.	List of used primers.....	44
Table 5.	List of used antibodies.....	45
Table 6.	Genotype and clinical presentations of patients with biallelic <i>UQCRFS1</i> variants.....	60
Table 7.	Genotype and clinical presentations of patients with biallelic <i>UQCRFS1</i> variants.....	68
Table 8.	Number of expressed genes stratified by tissue and gene class from our in-house RNA-seq cohort.....	77
Table 9.	Summary of cases diagnosed via RNA-seq.....	92
Table 10.	Summary of candidate genes discovered via RNA-seq.....	94
Table 11.	Summary of WES-diagnosed cases with an RNA-defect.....	95

List of publications

Peer-reviewed publications included in the thesis

Ait-El-Mkadem S, Dayem-Quere M, **Gusic M**, Chausseot A, Bannwarth S, François B, Genin EC, Fragaki K, Volker-Touw CLM, Vasnier C, Serre V, van Gassen KLI, Lespinasse F, Richter S, Eisenhofer G, Rouzier C, Mochel F, De Saint-Martin A, Abi Warde MT, de Sain-van der Velde MGM, Jans JJM, Amiel J, Avsec Z, Mertes C, Haack TB, Strom T, Meitinger T, Bonnen PE, Taylor RW, Gagneur J, van Hasselt PM, Rötig A, Delahodde A, Prokisch H, Fuchs SA, Paquis-Flucklinger V. Mutations in MDH2, Encoding a Krebs Cycle Enzyme, Cause Early-Onset Severe Encephalopathy. *Am J Hum Genet.* 2017 Jan 5;100(1):151-159. doi: 10.1016/j.ajhg.2016.11.014. Epub 2016 Dec 15. PMID: 27989324

Gusic M, Schottmann G, Feichtinger RG, Du C, Scholz C, Wagner M, Mayr JA, Lee CY, Yépez VA, Lorenz N, Morales-Gonzalez S, Panneman DM, Rötig A, Rodenburg RJT, Wortmann SB, Prokisch H, Schuelke M. Bi-Allelic UQCRFS1 Variants Are Associated with Mitochondrial Complex III Deficiency, Cardiomyopathy, and Alopecia Totalis. *Am J Hum Genet.* 2020 Jan 2;106(1):102-111. doi: 10.1016/j.ajhg.2019.12.005. Epub 2019 Dec 26. PMID: 31883641

Yépez VA, **Gusic M**, Kopajtich R, Mertes C, Smith N, Alston CL, Ban R, Beblo S, Berutti R, Blessing H, Ciara E, Distelmaier F, Freisinger P, Häberle J, Hayflick SJ, Hempel M, Itkis YS, Kishita Y, Klopstock T, Krylova TD, Lamperti C, Lenz D, Makowski CC, Mosegaard S, Müller MF, Muñoz-Pujol G, Nadel A, Ohtake A, Okazaki Y, Procopio E, Schwarzmayr T, Smet J, Staufner C, Stenton SL, Strom TM, Terrile C, Tort F, Van Coster R, Vanlander A, Wagner M, Xu M, Fang F, Ghezzi D, Mayr JA, Piekutowska-Abramczuk D, Ribes A, Rötig A, Taylor RW, Wortmann SB, Murayama K, Meitinger T, Gagneur J, Prokisch H. Clinical implementation of RNA sequencing for Mendelian disease diagnostics. In revision.

Peer-reviewed publications not included in the thesis

Catarino CB, Ahting U, **Gusic M**, Iuso A, Repp B, Peters K, Biskup S, von Livonius B, Prokisch H, Klopstock T. Characterization of a Leber's hereditary optic neuropathy (LHON) family harboring two primary LHON mutations m.11778G>A and m.14484T>C of the mitochondrial DNA. *Mitochondrion.* 2017 Sep;36:15-20. doi:10.1016/j.mito.2016.10.002. Epub 2016 Oct 6. PMID: 27721048

Wagner M, **Gusic M**, Günthner R, Alhaddad B, Kovacs-Nagy R, Makowski C, Baumeister F, Strom T, Meitinger T, Prokisch H, Wortmann SB. Biallelic Mutations in SLC1A2; an Additional Mode of Inheritance for SLC1A2-Related Epilepsy. *Neuropediatrics.* 2018 Feb;49(1):59-62. doi: 10.1055/s-0037-1606370. Epub 2017 Sep 15. PMID: 28915517

- Piekutowska-Abramczuk D, Assouline Z, Mataković L, Feichtinger RG, Koňáriková E, Jurkiewicz E, Stawiński P, **Gusic M**, Koller A, Pollak A, Gasperowicz P, Trubicka J, Ciara E, Iwanicka-Pronicka K, Rokicki D, Hanein S, Wortmann SB, Sperl W, Rötig A, Prokisch H, Pronicka E, Płoski R, Barcia G, Mayr JA. NDUFB8 Mutations Cause Mitochondrial Complex I Deficiency in Individuals with Leigh-like Encephalomyopathy. *Am J Hum Genet.* 2018 Mar 1;102(3):460-467. doi: 10.1016/j.ajhg.2018.01.008. Epub 2018 Feb 8. PMID: 29429571
- Yépez VA, Kremer LS, Iuso A, **Gusic M**, Kopajtich R, Koňáriková E, Nadel A, Wachutka L, Prokisch H, Gagneur J. OCR-Stats: Robust estimation and statistical testing of mitochondrial respiration activities using Seahorse XF Analyzer. *PLoS One.* 2018 Jul 11;13(7):e0199938. doi: 10.1371/journal.pone.0199938. eCollection 2018. PMID: 29995917
- Danhauser K, Alhaddad B, Makowski C, Piekutowska-Abramczuk D, Syrbe S, Gomez-Ospina N, Manning MA, Kostera-Pruszczyk A, Krahn-Peper C, Berutti R, Kovács-Nagy R, **Gusic M**, Graf E, Laugwitz L, Röblitz M, Wroblewski A, Hartmann H, Das AM, Bültmann E, Fang F, Xu M, Schatz UA, Karall D, Zellner H, Haberlandt E, Feichtinger RG, Mayr JA, Meitinger T, Prokisch H, Strom TM, Płoski R, Hoffmann GF, Pronicki M, Bonnen PE, Morlot S, Haack TB. Bi-allelic ADPRHL2 Mutations Cause Neurodegeneration with Developmental Delay, Ataxia, and Axonal Neuropathy. *Am J Hum Genet.* 2018 Nov 1;103(5):817-825. doi: 10.1016/j.ajhg.2018.10.005. Epub 2018 Oct 25. PMID: 30401461
- Eraslan B, Wang D, **Gusic M**, Prokisch H, Hallström BM, Uhlén M, Asplund A, Pontén F, Wieland T, Hopf T, Hahne H, Kuster B, Gagneur J. Quantification and discovery of sequence determinants of protein-per-mRNA amount in 29 human tissues. *Mol Syst Biol.* 2019 Feb 18;15(2):e8513. doi: 10.15252/msb.20188513. PMID: 30777893
- Suleiman J, Riedhammer KM, Jicinsky T, Mundt M, Werner L, **Gusic M**, Burgemeister AL, Alsaif HS, Abdulrahim M, Moghrabi NN, Nicolas-Jilwan M, AlSayed M, Bi W, Sampath S, Alkuraya FS, El-Hattab AW. Homozygous loss-of-function variants of TASP1, a gene encoding an activator of the histone methyltransferases KMT2A and KMT2D, cause a syndrome of developmental delay, happy demeanor, distinctive facial features, and congenital anomalies. *Hum Mutat.* 2019 Nov;40(11):1985-1992. doi: 10.1002/humu.23844. Epub 2019 Jul 22. PMID: 31209944
- Del Dotto V, Ullah F, Di Meo I, Magini P, **Gusic M**, Maresca A, Caporali L, Palombo F, Tagliavini F, Baugh EH, Macao B, Szilagyi Z, Peron C, Gustafson MA, Khan K, La Morgia C, Barboni P, Carbonelli M, Valentino ML, Liguori R, Shashi V, Sullivan J, Nagaraj S, El-Dairi M, Iannaccone A, Cutcutache I, Bertini E, Carrozzo R, Emma F, Diomedi-Camassei F, Zanna C, Armstrong M, Page M, Stong N, Boesch S, Kopajtich R, Wortmann S, Sperl W, Davis EE, Copeland WC, Seri M, Falkenberg M, Prokisch H, Katsanis N, Tiranti V, Pippucci T, Carelli V. SSBP1 mutations cause mtDNA depletion

underlying a complex optic atrophy disorder. *J Clin Invest.* 2020 Jan 2;130(1):108-125. doi: 10.1172/JCI128514. PMID: 31550240

Gusic M, Prokisch H. ncRNAs: New Players in Mitochondrial Health and Disease? *Front Genet.* 2020 Feb 28;11:95. doi: 10.3389/fgene.2020.00095. eCollection 2020. PMID: 32180794

Yepez VA, Mertes C, Müller MF, Klaproth-Andrade D, Wachutka L, Frésard L, **Gusic M**, Scheller IF, Goldberg PF, Prokisch H, Gagneur J. Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc.* 2021 Feb;16(2):1276-1296. doi: 10.1038/s41596-020-00462-5. Epub 2021 Jan 18. PMID: 33462443

Mertes C, Scheller IF, Yépez VA, Çelik MH, Liang Y, Kremer LS, **Gusic M**, Prokisch H, Gagneur J. Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun.* 2021 Jan 22;12(1):529. doi: 10.1038/s41467-020-20573-7. PMID: 33483494

Gusic M, Prokisch H. Genetic basis of mitochondrial diseases. *FEBS Lett.* 2021 Apr;595(8):1132-1158. doi: 10.1002/1873-3468.14068. Epub 2021 Mar 21. PMID: 33655490

Invernizzi F, Legati A, Nasca A, Lamantea E, Garavaglia B, **Gusic M**, Kopajtich R, Prokisch H, Zeviani M, Lamperti C, Ghezzi D. Myopathic mitochondrial DNA depletion syndrome associated with biallelic variants in *LIG3*. *Brain.* 2021 Jun 24:awab238. doi: 10.1093/brain/awab238. Online ahead of print. PMID: 34165507

1. Introduction

1.1. Mitochondria

Mitochondria (the Greek *mitos* (thread-like) and *khondros* (grain or granule)) are double-membrane bound subcellular compartments found in most eukaryotic organisms, essential for normal physiology (Henze and Martin, 2003). According to the endosymbiotic theory, mitochondria originate from prokaryotic organisms related to alphaproteobacteria- the hypothetical ancestor termed protomitochondrion (Sagan, 1967). During early development, the transition to a highly oxidizing atmosphere created a selective pressure that favored the organisms with respiratory capacity (Castresana and Saraste, 1995; Lyons et al., 2014), including heterotrophic anaerobes- precursors of the modern eukaryotic cells, which engulfed the protomitochondrion. This endosymbiosis occurred around two billion years ago, upon which with time mitochondrial signals and homeostasis have been synchronized with the eukaryotic cell, finally resulting in the compartmentalization of cellular metabolism (Timmis et al., 2004). Although with time mitochondria have lost many of their features and became dependent on communication with other cellular compartments (most notably the nucleus), they have still kept some unique bacteria-like features, such as their own circular genome (the mitochondrial DNA- mtDNA), and composition of double membranes and ribosomes. As mentioned, mitochondria have two membranes—a smooth outer and an inner one that encloses the matrix space, both forming invaginations named cristae. Mitochondria typically form a network, which occupies a substantial part of the cytosol (e.g. up to 20% of the cell volume in human liver cells; Alberts et al., 1994). Their surface area usually ranges between 0.75 and 3 μm^2 (Wiemerslage and Lee, 2016). Their number and shape can differ widely across cell types. For example, tissues with intensive oxidative metabolism, such as the heart muscle, have mitochondria with particularly large numbers of cristae. Moreover, they are highly dynamic organelles undergoing coordinated cycles of fission and fusion, referred to as “mitochondrial dynamics”, in order to maintain their shape, distribution, and size, and adapt to energy demands (Tilokani et al., 2018). In animals, mtDNA inheritance is almost exclusively maternal. This is achieved by the elimination of paternal mitochondria and mtDNA via different mechanisms such as nuclease- or ubiquitin-dependent degradation, and autophagy (reviewed in Sato and Sato, 2013).

1.1.1. Mitochondrial functions

Mitochondria are best known as the “powerhouses” of the cell, as they generate more than 90% of ATP, the “energy currency”, via oxidative phosphorylation (OXPHOS). This is a final step that integrates the catabolism of carbohydrates, fatty acids, and amino acids. These functions, amongst many others, make mitochondria deeply integrated into the cellular metabolism and signalling pathways (reviewed in Spinelli and Haigis, 2018). They present biosynthetic hubs, providing necessary building blocks for the biosynthesis of glucose, fatty acids, nucleotides, cholesterol, amino acids, and heme. They support the balance of reducing equivalents NAD^+/NADH and $\text{NADP}^+/\text{NADPH}$ via different redox shuttles (malate-aspartate

shuttle, folate shuttle (or the One carbon metabolism), α -Glycerophosphate shuttle) that favor NAD^+ synthesis in the cytosol and mitochondrial NADH synthesis in mitochondria. They repurpose the metabolic by-products (often regarded as “waste”), such as lactate, ammonia, reactive oxygen species (ROS), and H_2S . In fact, mitochondria are a major site of ROS production themselves (Sena and Chandel, 2012). Also, mitochondrial protein import and maturation (Wiedemann and Pfanner, 2012) as well as the iron–sulfur (Fe–S) clusters synthesis (Lill, 2009) are essential cellular processes. Finally, mitochondrial activity is integrated into many cytosolic signaling pathways, stress responses, and programmed cell death (apoptosis) (summarized in Figure 1; reviewed in Galluzzi et al., 2012; Bohovych and Khalimonchuk, 2016; Spinelli and Haigis, 2018; Eisner et al., 2018; Galluzzi et al., 2018; Pfanner et al., 2019).

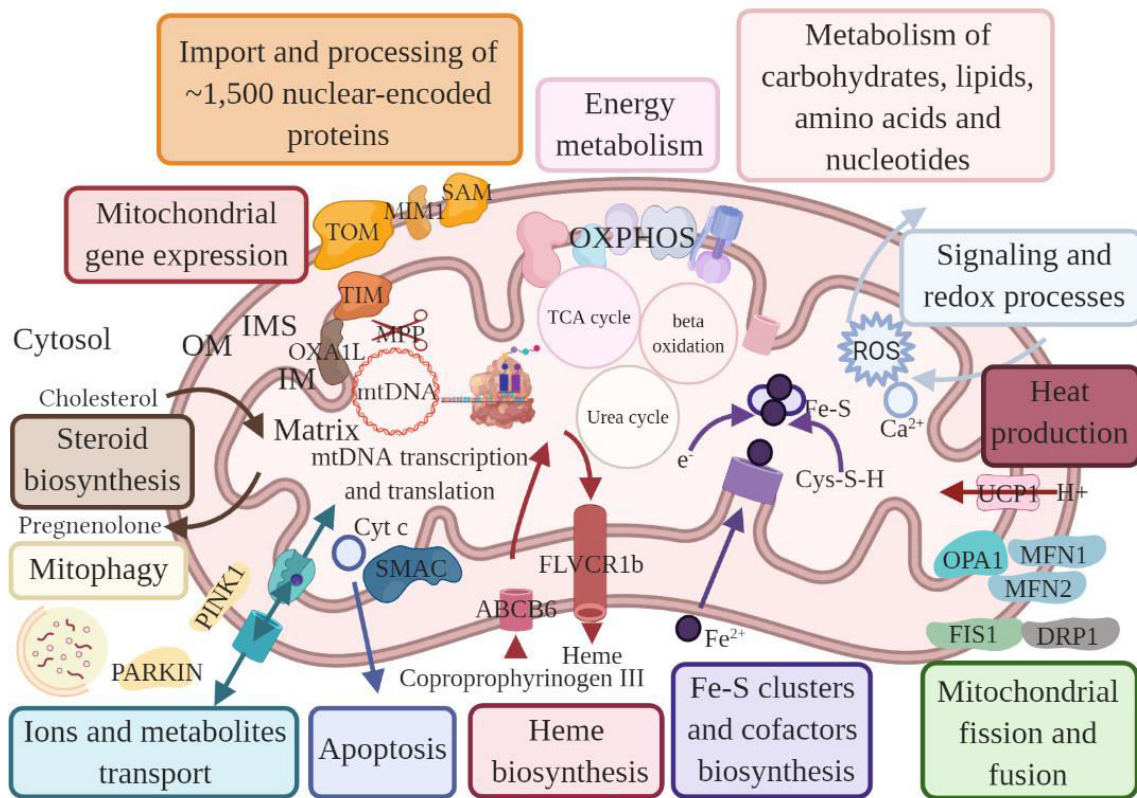


Figure 1. Graphical representation of a mitochondrion and its major functions. Figures shows main proteins and molecules involved. Abbreviations: OM, outer membrane; IMS, intermembrane space; IM, inner membrane; TCA, tricarboxylic acid.

1.1.2. Mitochondrial genetics

The control of mitochondrial gene expression is unique in that its components have origins in both mitochondria and nucleus. During evolution, most of the ancestral mitochondrial genomic material was lost or transferred to the nuclear genome (Gabaldon and Huynen, 2004). The remnant is a compact, circular genome, present in cells in a vast excess of copies relative to the nuclear genome (nuclear DNA, nDNA). Human mtDNA is 16,569 base pairs (bp) in length, encoding 2 rRNAs, 22 tRNAs, and mRNAs for 13 proteins- all

subunits of the OXPHOS system (Anderson et al., 1981; Figure 2). The mt-rRNA coding sequences and most of the protein coding sequences are separated by the mt-tRNAs. The entire mitochondrial genome is transcribed by the mitochondrial RNA polymerase (POLRMT) from both strands, named heavy (H) or light (L) based on their buoyancy in cesium chloride density gradients. According to the “tRNA punctuation” model (Anderson et al., 1981; Ojala, et al., 1981), the resulting polycistronic transcripts are subsequently processed by excision of mt-tRNAs, generating individual mt-mRNAs and mt-rRNAs. Upon release, the transcripts undergo different post-transcriptional modifications (reviewed in D’Souza and Minczuk, 2018). Finally, the mature mt-mRNAs, mt-tRNAs, and the assembled mitoribosome come together in the translation apparatus, resulting in the synthesis of 13 proteins. Nevertheless, as their coding capacity is very limited, mitochondria are heavily dependent on the import of over 1,500 nuclear-encoded proteins. Approximately 99% of mitochondrial proteins are encoded by the nuclear genes and depend on the specific targeting signals that direct them from the cytosol to the mitochondrial surface receptors, and then into the appropriate mitochondrial subcompartments (Schmidt et al., 2010; Calvo et al., 2016).

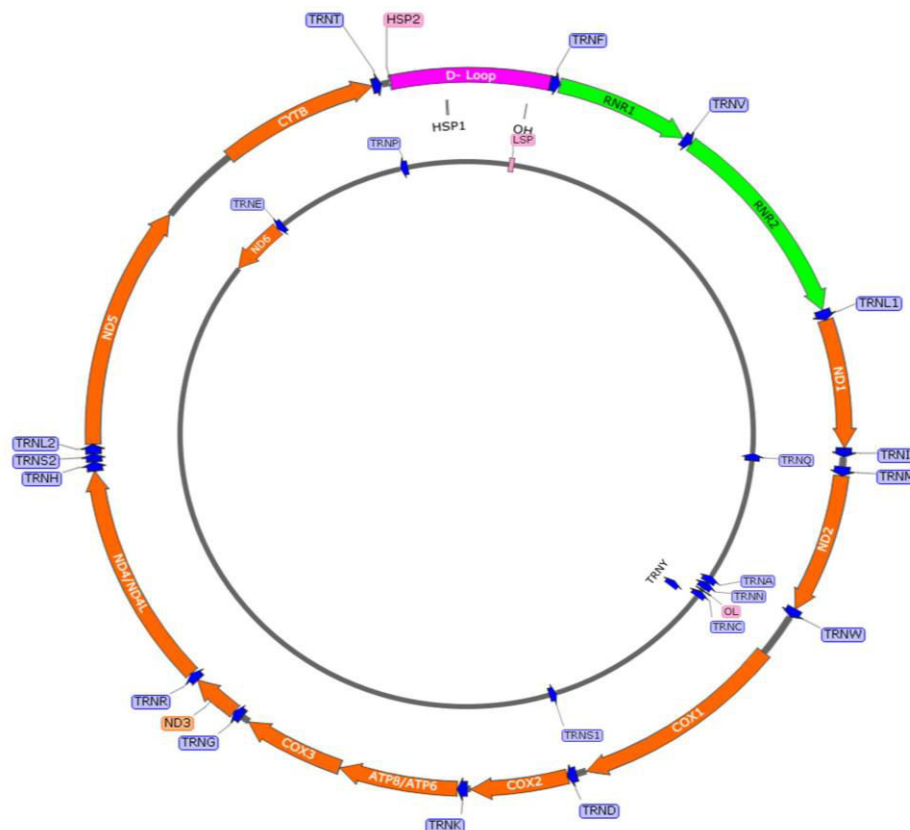


Figure 2. Graphical representation of a human mtDNA. Double stranded circular mtDNA consists of 37 genes, encoding 13 proteins of the OXPHOS (in orange), two rRNAs (in green), and 22 tRNAs (in blue), with D-loop (in pink) responsible for the regulation of replication, transcription and translation, and containing L-strand (LSP) and H-strand (HSP1/2) promoters, and origins of H-strand and L-strand replication (OH and OL, respectively).

The mtDNA has some distinct properties that make it an attractive source of information for both population and evolutionary genetics studies. First, the high copy number (average of 100–1000 copies per cell) enables inspection of mtDNA from a wide range of tissue/cellular sources (Cavelier et al., 2000). Second, its maternal inheritance enables researchers to historically trace lineages without the complexity of recombination (Pakendorf and Stoneking, 2005). Third, its *de novo* substitution rate of $\sim 10^{-8}$ substitutions per base pair per year and lack of recombination make it a good tool for inspecting events in human prehistory (Soares et al., 2009). Finally, each mitochondrion contains up to 10 copies of the mtDNA (Johns, 1996). When these mtDNA sequences are identical within a cell, it is referred to as homoplasmy. In contrast, any situation resulting in distinct mtDNA sequences, whether from variants occurring *de novo* or inherited are referred to as heteroplasmy (Ramos et al., 2013; Figure 3). Heteroplasmy in more than 1% of mtDNA molecules is present in 45% of individuals (Wei et al., 2019). Some factors (e.g. time, number of replication, ROS, low fidelity of the DNA polymerase γ (POLG)) can result in the accumulation of novel and potentially pathogenic variants (Payne et al., 2013; Li et al., 2019). Significant changes in heteroplasmy levels can also occur due to cell proliferation over time (Lehtinen et al., 2000). More dramatic and unpredictable change in the heteroplasmy frequencies, so-called “heteroplasmy shift”, can occur between mothers and children, as the result of a genetic “bottleneck”, shaped by selection acting in the female germline (Wei et al., 2019). In the case of damaging mtDNA variants, cells usually tolerate a certain level of mutated mtDNA, but after a threshold is exceeded, it can lead to cellular dysfunction and disease (Figure 3). Interestingly, there have been controversial reports of paternal contribution to the mtDNA heterogeneity (Luo et al., 2018; Lutz-Bonengel and Parson, 2019). The debate of “heteroplasmic haplotype” has come to an end when Wei et al. (2020) revealed that the “paternal haplotype” can be a false positive finding due to the large rare or unique nuclear-mitochondrial DNA segments (mega-NUMTs), transmitted from the father in an autosomal manner and disguised as a mitochondrial haplotype in sequencing data.

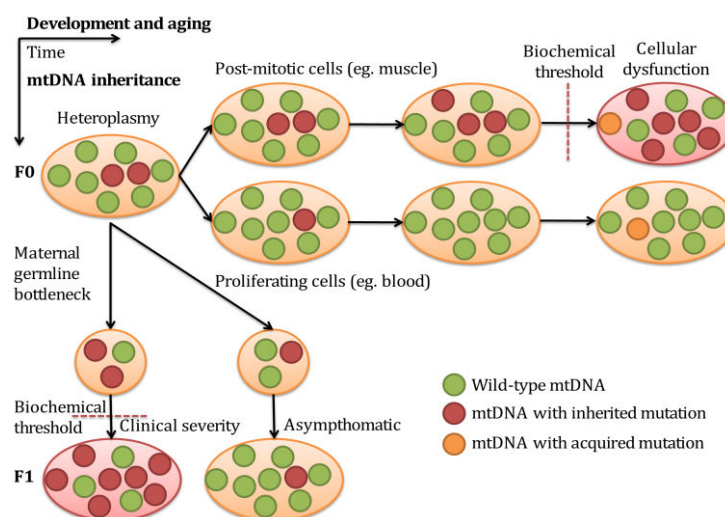


Figure 3. Changes in heteroplasmy and its contributors.

1.2. Mitochondrial diseases

Rare diseases are defined by the European Union as „life-threatening or chronically debilitating conditions whose prevalence is so low (less than 5 per 10,000 people) that special combined efforts are needed to prevent significant morbidity or perinatal and early mortality and address the considerable reduction in an individual’s quality of life or socioeconomic potential“ (Baldovino et al., 2016). Approximately 10,000 rare diseases are clinically described, but only 39% have a resolved genetic etiology (Ferreira, 2019). These numbers are on constant flux, as new gene-disease associations are described regularly. Although very versatile, what is common for all rare diseases are major clinical challenges due to the lack of scientific knowledge- often delayed or wrong diagnosis and seldom achieved optimal treatment and care. Rare diseases with an established genetic cause are named Mendelian diseases, and are caused by single-gene (monogenic) or single-locus changes that follow Mendel’s laws of inheritance (Eilbeck et al., 2017). Mitochondrial diseases stand out as a textbook example of all challenges and enigmas of rare diseases. They present the largest group of inherited metabolic disorders and are among the most common forms of inherited neurological disorders (Ferreira, 2019). Although there are different definitions of mitochondrial diseases, here they are defined by primary defects in the entire route of the pyruvate oxidation process, including the pyruvate dehydrogenase (PDH) complex, the Krebs cycle and the OXPHOS system. These defects can be caused by pathogenic variants in genes encoding subunits or assembly factors of primarily respiratory chain complexes (RCCs), but also by pathogenic variants in genes encoding proteins required for the mtDNA replication, transcription, and translation, generation or transport of substrates in reactions upstream of the OXPHOS or cofactors of OXPHOS, and, finally, genes which encode proteins important for the homeostasis of mitochondria (Mayr et al., 2015). Physiological consequences include decreased ATP production, imbalance in the NAD^+/NADH pools, increased ROS production, disruption of Ca^{2+} homeostasis, and impairment of pathways feeding into the OXPHOS such as the Krebs cycle and fatty acid β -oxidation (Munnich and Rustin, 2001; Reinecke et al., 2009), which can lead to increased production of lactate and ketone bodies, respectively.

In 1959, Luft reported the first case of a patient suffering from abnormal OXPHOS, presenting with generalized weakness, an inability to gain weight despite polyphagia, and extensive perspiration (Luft et al., 1962). Mitochondria from the skeletal muscle taken during the muscle biopsy were chosen for the biochemical investigations as they required a large amount of tissue, which remains until today one of the approaches to biochemically diagnose mitochondriopathies (Ernster et al., 1959). As more reports followed with time, it became evident that the phenotypic spectrum of mitochondrial disease is extremely heterogeneous. Such complexity is well described by Munnich and Rustin, who famously stated that mitochondrial diseases may cause “any symptom, in any organ or tissue, at any age, with any mode of inheritance” (Munnich and Rustin, 2001). They are frequently multisystemic in nature and cause significant morbidity and mortality (McFarland et al., 2010; DiMauro et al., 2013). The overall disease burden is extensive, resulting in substantial direct and indirect health care costs to the patient, their families, and society as a whole. Notably, mitochondrial

dysfunction not isolated only to the mitochondrial disorders. It can be observed as a secondary defect in other genetic disorders, such as ethylmalonic aciduria (Tiranti et al., 2009) and Wilson disease (Lutsenko and Cooper, 1998). It is also often observed in common diseases (such as Parkinson's, Alzheimer's, and Huntington's disease) (Lane et al., 2015), and is a mark of the aging process (Payne and Chinnery, 2015).

1.2.1. Prevalence

The prevalence of mitochondrial disease has proven difficult to establish, as a result of the clinical and genetic heterogeneity potentially leading to a false diagnosis or misdiagnosis. Therefore, its reported prevalence of 5-20 in 100,000 (Gorman et al., 2016) is likely an underestimate. Childhood-onset mitochondrial disease is more often caused by the recessive nDNA mutations, with an estimated prevalence of 5 to 15 in 100,000 individuals. Conversely, adult-onset diseases mainly arise due to mtDNA mutations with a prevalence of 10 in 100,000, while the prevalence due to nuclear mutations is estimated to account for just 3 in 100,000 (Gorman et al., 2015). Population-based studies revealed that 236 in 100,000 people carry a known pathogenic mtDNA variant (m.3243A>G) (Manwaring et al., 2007). It is important to state that due to founder mutations and consanguinity in a specific population, studies can be biased and need not reflect the global distribution. Also, prevalence estimations and cohorts are based on clinical diagnoses, which can lead to ascertainment bias due to overlooking of patients with atypical presentations of mitochondrial disease. Based on purely population genotypes, Tan et al. (2020) provided an estimation of the lifetime risk of almost 1 in 2000 for an autosomal recessive mitochondrial disorder. To conclude, the true epidemiological burden of the mitochondrial disease still remains unclear.

1.2.2. Clinical presentation and diagnosis

Though the OXPHOS defects are a typical outcome of mitochondrial disorders, their clinical presentation shows a remarkable variability (Figure 4). Such pleiotropy suggests that the mere disruption in the ATP production across tissues cannot wholly explain the observed clinical range of symptoms (Suomalainen and Battersby, 2018). Still, the most commonly and severely affected tissues in mitochondriopathies are the ones demanding the highest amount of energy and producing high levels of lactate and ROS, such as the brain, retina, kidney, liver, and skeletal or cardiac muscle. Hence, patients usually present with neurodegeneration, in many cases in combination with muscle weakness, cardiomyopathy, optic atrophy, or liver failure, thereby frequently making mitochondriopathy a multisystemic disorder. However, this is not a rule and the impairment of other tissues/organs might originate from tissue-specific isoforms, energy demands, or regulations of the electron flux, as well as heteroplasmy level. Most mitochondrial diseases are progressive, although they may remain stable over decades, or even ameliorate in some cases (Koene and Smeitink, 2011).

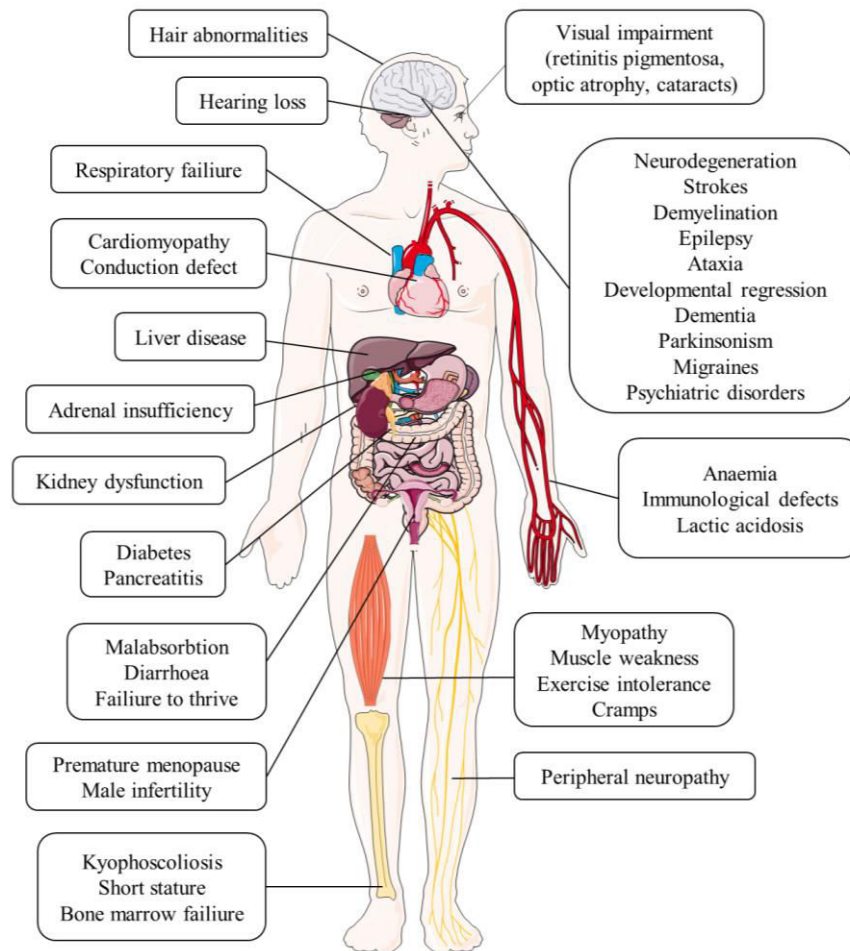


Figure 4. Phenotypic spectrum of mitochondrial diseases.

Mitochondrial disease patients may display a cluster of clinical features that fall into a discrete clinical syndrome (Chinnery, 2014). However, more often affected individuals display a variety of symptoms that make the conclusive diagnosis challenging. Therefore it has been suggested to suspect a mitochondrial disorder upon the occurrence of the combined impairment of seemingly unrelated organs (Wallace, 1992; Munnich et al., 1996). For further elucidation, metabolic or biochemical analyses are required. Clinical rating scales, such as the Mitochondrial Disease Criteria (MDC) score, can also indicate the likelihood of the disease (Witters et al., 2018).

Clinical diagnosis of mitochondrial diseases is complicated by the variety of clinical symptoms on the one side and the symptomatic overlap with other diseases on the other. The poor correlation between genotype and phenotype adds further complexity. In some genes the pathogenic variants, and moreover even the same ones, can give rise to different clinical symptoms. For example, the most prevalent pathogenic mtDNA variant m.3243A>G (Gorman et al., 2015) is associated with mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS). However, this variant also causes maternally inherited deafness and diabetes, progressive external ophthalmoplegia either isolated or in any combination of phenotypes. More perplexing, about a third of variant carriers present with

clinical manifestations that are currently not recognized as a classical mitochondrial disease. To conclude the enigma, 9% of variant carriers are asymptomatic (Nesbitt et al., 2013). In contrast, the same phenotypes can have diverse genetic causes. For example, Leigh syndrome can be caused by pathogenic variants in over 90 genes (Rahman et al., 2017).

Until very recently, the classical diagnostic approach for mitochondrial disorders consisted of an initial deep clinical phenotyping and the evaluation of the OXPHOS enzymes in tissues derived from an invasive muscle biopsy. Based on these findings a targeted Sanger sequencing of appropriate candidate genes followed (Wortmann et al., 2017). This approach is slowly but surely being replaced with the “genetics first” approach, thanks to the development and reduced costs of next generation sequencing (NGS) technologies.

1.2.3. Metabolic diagnosis- biomarkers

The conventional biomarkers used to support a diagnosis of mitochondrial disease in clinical practice include metabolic intermediates, enzymes, and end products of anaerobic glucose metabolism, resulting from the OXPHOS impairment (Shaham et al., 2010). However, these biomarkers are most often not disease-specific, exhibit poor sensitivity and specificity, or do not cover all mitochondrial diseases, so their interpretation is not straightforward (Haas et al., 2007). Besides, some biomarkers are only useful for the diagnosis of specific syndromes, such as measurements of plasma thymidine and deoxyuridine levels in mitochondrial neurogastrointestinal encephalomyopathy (MNGIE) (Nishino et al., 1999). The most frequently recognized laboratory abnormality in patients with mitochondrial diseases is lactic acidosis. Lactate is elevated due to dysfunction of OXPHOS, which forces cells to rely on glycolysis as a source of ATP production. The stimulation of glycolysis leads to an overproduction of pyruvate, which is then either transaminated to alanine or reduced to generate lactate (Koenig, 2008). Still, the largest study to examine lactic acidosis exclusively in children with OXPHOS deficiencies reported elevated venous lactate in only 30% of patients (Munnich, 1996). In our cohort of over 2000 patients suspected to suffer from a mitochondrial disease, almost half presented with lactic acidosis. One must be aware that elevated venous lactate is not specific just for mitochondrial diseases, but can also be observed in various other conditions (Koenig, 2008) and can be an artefact due to difficulty in acquiring the blood sample from a struggling child. Similarly, ketone bodies (β -hydroxybutyrate, acetoacetate, and acetone) can be employed (Smeitnik et al., 2006). These are produced in the liver from acetyl-CoA under conditions where acetyl-CoA cannot feed into the Krebs cycle as the latter is halted, and get secreted into the bloodstream to be used by other tissues for terminal oxidation (Puchalska and Crawford, 2017). This occurs during certain physiological states, including fasting and starvation, but in the case of mitochondrial disease ketone body levels in the blood might rise despite feeding (Munnich et al., 1996). Other biomarkers frequently analysed include creatine kinase, amino acids (e.g. alanine), acylcarnitines, and further organic acids, and are often accessed in the cerebrospinal fluid (CSF) and urine (Mitochondrial Medicine Society's Committee on Diagnosis et al., 2008). For example, increased urinary excretion of 3-methylglutaconic acid (3-methylglutaconic aciduria, 3-MGA) is indicative of gene defects in *AUH*, *CLPB*,

DNAJC19, *HTRA2*, *OPA1*, *OPA3*, *TIMM50*, *SERAC1*, *TAZ*, *TMEM70*, and *TIMM50* (Wortmann et al., 2013; OMIM). Finally, fibroblast growth factor 21 (FGF21) and growth/differentiation factor 15 (GDF15) have been shown to be increased in the serum of patients with mitochondrial diseases, both exhibiting high sensitivity and specificity (Suomalainen et al., 2011; Yatsuga et al., 2015). However, they were tested in smaller sample size and the following studies questioned their utility (Koene et al., 2014). Thus, their value still needs to be determined in larger, well-characterized mitochondrial disease cohorts.

1.2.4. Biochemical and histochemical diagnosis

Establishing the biochemical phenotype, although losing diagnostic significance in favor of NGS, is still an important part of the diagnostic work-ups. It may be helpful in candidate disease-causing gene selection, provide an additional level of interpretation to molecular diagnosis and narrow down defects detected by other methods. Mitochondrial diseases can be caused by deficiencies in the OXPHOS complexes. OXPHOS deficiencies may be isolated, affecting only one of the complexes, or they may be combined, affecting several in combination (Kirby et al., 2007). Additionally, they may be systemic and affect all tissues, or show tissue(s)-specificity. Biochemical assays can measure the activity of each of the OXPHOS enzyme complex, but also combined complex I+III and II+III activities. They are typically based on spectrophotometric measurements of complexes activity and citrate synthase (CS) activity (used for data normalization) of the mitochondrial fraction derived from tissue/cellular homogenate (Kirby et al., 2007). As skeletal muscle is frequently affected and muscle biopsy may be available (Taylor et al., 2004; McFarland et al., 2010), it presented a gold standard for diagnosis of mitochondrial disorders before genetic testing in Western Europe (Wortmann et al., 2017). Complex I deficiency in muscle presents the most frequent biochemical signature of mitochondrial disorders (Mayr et al., 2015). Although very useful for the detection of wide-spread mitochondrial defects, muscle biopsies have some drawbacks that must be considered. If not affected, the standardly used skeletal muscle may not necessarily express a biochemical defect (Smeitnik, 2003). On the other side, heart-, liver- or CNS-biopsies, although possible, are much more invasive and harmful for the patient. As an alternative, a skin biopsy to isolate fibroblasts is becoming more and more favorable as it is far less invasive. Cultured skin fibroblasts are a well-described *in vitro* model for assigning pathogenicity of identified variants and can be used to recapitulate the observed defect from other tissues. It is highly recommended to perform experiments on fresh material, as freezing of the sample can damage the tissue or impact its metabolic state, which is in practice rarely possible. Finally, up to the present day, there is no universal protocol for spectrophotometric quantification and each lab uses its own reference ranges, making the results comparison between labs very difficult (Rodenburg, 2011). Other assays that determine the amounts of OXPHOS complexes can be performed as an alternative or in addition, such as the blue native polyacrylamide gel electrophoresis (BN-PAGE) followed by Western blot (Schagger and von Jagow, 1991) and colorimetric in-gel measurements of enzyme activities following BN-PAGE (Wittig et al., 2007). Additionally, OXPHOS function can be analysed polarographically, via a technique that measures oxygen consumption (Rustin et al., 1991).

Polarographic oxygen sensors have been developed as a commercially available Oroboros Oxygraph Instrument (Haller et al., 1994) and Seahorse XF Analyzer (Rogers et al., 2011).

To assess the cell to cell variability due to heteroplasmy, histochemical stainings of cytochrome c oxidase (COX) of complex IV and/or succinate dehydrogenase (SDH) of complex II are standardly used to investigate OXPHOS in tissue cryosections (Sciocco and Bonilla, 1996). Although useful for detecting mosaic OXPHOS deficiency, this method is qualitative and susceptible to inter-observer variability. Recently, an alternative, more sensitive and quantitative assay of quadruple immunofluorescence was developed. Using fluorescently labeled antibodies against NDUFB8, COX, porin and laminin, it enables assessment of complex I and IV abundance relative to the mitochondrial mass in individual myofibres (Rocha et al., 2015).

To summarize, biochemical and histochemical approaches in the diagnostics of mitochondrial disorders are very helpful, but far from definitive. Negative findings do not always exclude, and positive findings do not always confirm a mitochondrial disease, so their results should be considered with caution (Smeitink, 2003).

1.2.5. Genetic diagnosis

Deciphering the precise molecular cause (genotype) that explains the clinical features presents the cornerstone of safe medical practice (Wright et al., 2018a). For patients with a rare genetic disorder, the genetic diagnosis brings, although perhaps not necessarily the cure, certain comfort and benefits. It can improve disease management and therapy, open access to disorder-specific support groups, and most importantly- enable accurate determination of risk to existing and future family members. Genetic etiology allows identification of ~10% of mitochondrial diseases potentially amenable to specific treatment strategies (Distelmaier et al., 2017). However, finding a diagnosis for each mitochondrial disease patient remains a considerable challenge because of the vast genetic and phenotypic variability and our incomplete knowledge regarding both the disease and genetics.

To understand the genetics of mitochondrial disorders, it is crucial to be aware of the unique properties of mitochondria: (1) dual genetic origin of mitochondrial proteins, encoded by mtDNA and, more often, nDNA, resulting in every possible pattern of inheritance—maternal, X-linked, autosomal recessive (AR), autosomal dominant (AD), and *de novo*. It is estimated that pathogenic variants in the mtDNA are responsible for up to 15-30% of childhood-onset and up to ~80% of adult-onset cases (Chinnery, 2014; Gorman et al., 2016). (2) maternal inheritance of mtDNA (3) heteroplasmy. Concerning heteroplasmy, the severity of the phenotype can depend on the percentage of the pathogenic variant present in the cell, which is termed the threshold effect (Rossignol et al., 2003). The threshold level can vary between different tissues and variants, but usually a pathogenic variant is required to reach >70% of heteroplasmy before a deleterious phenotype is observed in a cell. In addition, the transmission of heteroplasmy levels from mother to offspring is often random and unpredictable (Mishra and Chan, 2014). This explains the heterogeneity in heteroplasmy level, clinical phenotype, and severity frequently observed within the same family (Ng et al., 2016). To add, multiple mtDNA deletions and the mtDNA depletion (reduced copy number

of mtDNA) can occur as secondary changes due to mutations in the mtDNA replication and/or maintenance genes such as *POLG*, *PEO1*, *ANTI*, *DGUOK*, *TYMP*, *TWINK*, *SSBP1*, etc. (Schon et al., 2012; Stenton and Prokisch, 2020).

1.2.5.1. Historical overview

Until the human nuclear DNA was fully sequenced in 2003, the molecular era in mitochondrial disorders was dominated by the identification of mutations in the mtDNA. The first pathogenic mutation in the mtDNA was described in 1988, when Holt et al. identified large-scale single deletions in the mtDNA of patients with mitochondrial myopathies by linkage analysis using restriction fragment length polymorphisms (RFLPs), in addition proving the existence of heteroplasmy in humans. Soon thereafter, Wallace et al. (1988) described the first mtDNA point mutation (m.11778G>A) in *MT-ND4* in a family with Leber's hereditary optic neuropathy (LHON). In the next decade, new pathogenic mutations of the mtDNA were described at the pace of about ten per year, so by 2001 115 point mutations were listed (Servidei, 2001). Clinical syndromes were attributed to the mtDNA point mutations or rearrangements of the mtDNA. In other cases, however, the inheritance pattern of the disease indicated that it had to reside in the nDNA. In 1989 and 1991 two mitochondrial diseases were described by Zeviani et al. and Moraes et al., following autosomal dominant and recessive modes of inheritance, respectively. However, the disease-causing genes remained elusive. In 1989, pathogenic variants in the *PDHA1* were discovered as a cause of lactic acidosis, making it the first mitochondrial disease-associated nuclear gene (Endo et al., 1989). Next, in 1995, *SDHA* was identified in two sisters with complex II deficiency (Bourgeron et al., 1995). Many others subsequently followed this discovery. In over 16 years, 94 further nuclear-encoded gene defects were identified by candidate gene sequencing, often in combination with linkage studies or homozygosity mapping (Frazier et al., 2019). Assuming a clear genotype-phenotype correlation, the growing list of identified genes in combination with the respective clinical presentation prompted screening for candidate genes on a small scale. This small-scale analysis was expanded to disease gene panels as sequencing costs were reduced. Finally, the advent of the NGS technologies has offered an unbiased, high-throughput approach in the identification of disease-causing variants and caused a revolution in genetic diagnostics. This technology enabled the discovery of more than 120 novel mitochondrial disease genes in a period of just 10 years. It has also shifted the primary challenge of human genetics from variant discovery to variant interpretation (Goldstein et al., 2013; MacArthur et al., 2014). To date, pathogenic variants in 338 genes have been described to cause a mitochondrial disease (Stenton and Prokisch, 2020). The mode of inheritance is autosomal recessive (AR) in the majority- 262 genes, maternal in 36, autosomal dominant (AD) in 8, X-linked dominant in 6, X-linked recessive in 4, and finally 22 genes are inherited by a combination of AR and AD. The affected genes are implicated in a variety of mitochondrial processes (Figure 5). Focusing on the mtDNA, hundreds of its variants are implicated in a variety of human diseases (Lott et al., 2013), but only 94 have a confirmed status (MITOMAP, <https://www.mitomap.org>).

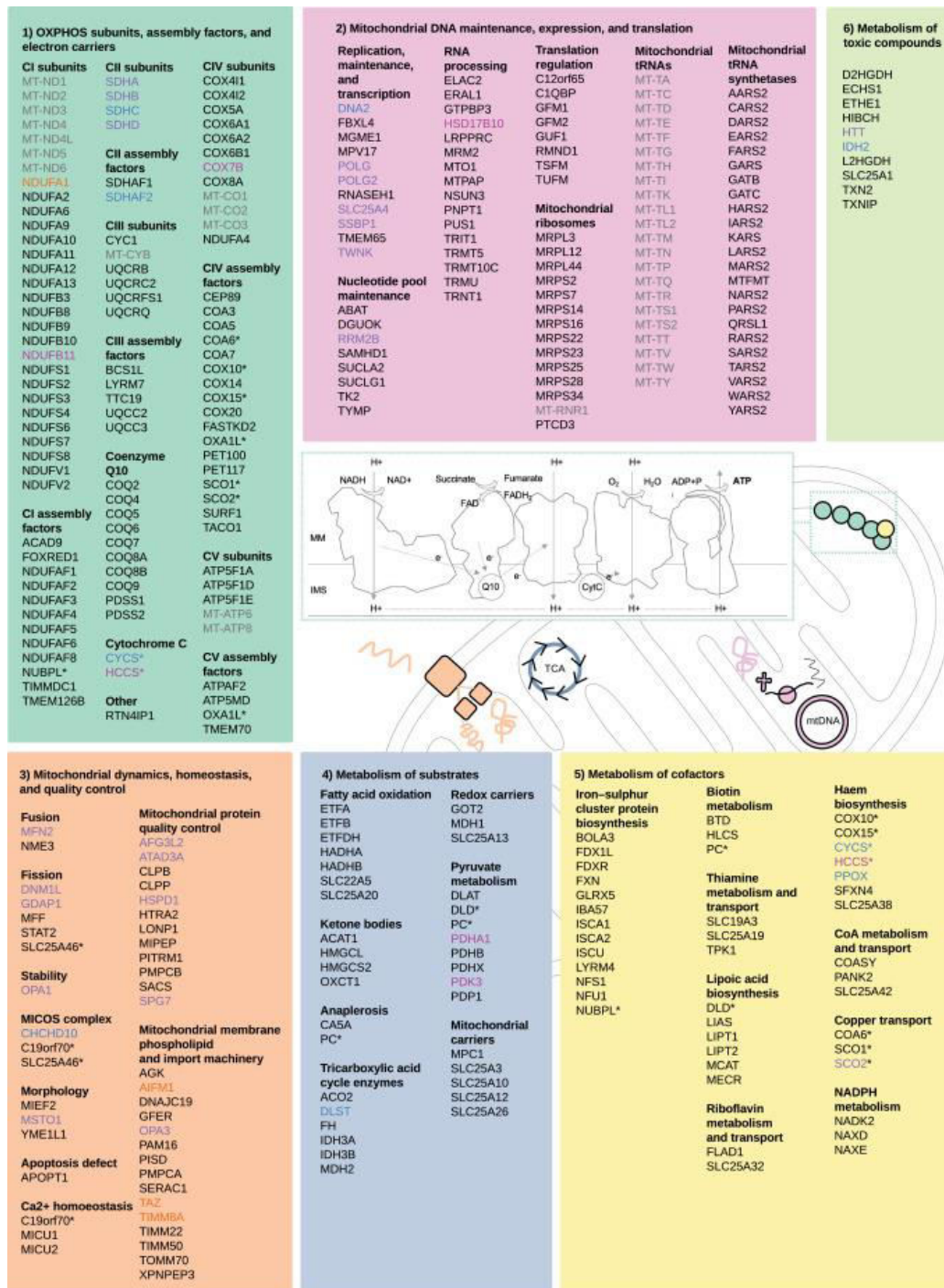


Figure 5. Described mitochondrial disease genes categorized by their functional roles. Asterisk indicates genes with dual roles. Figure taken from Stenton and Prokisch (2020).

1.2.5.2. Sanger sequencing

Sanger sequencing is a “first-generation” DNA sequencing method. Developed in 1977, it is based on the selective incorporation of chain-terminating dideoxynucleotides by the DNA polymerase during the *in vitro* DNA replication (Sanger et al., 1977). This method allows the analysis of up to 300 nucleotide (nt) sequence. Its use marked the start of the molecular era in clinical diagnostics as it allowed location of a pathogenic mutation with a single base resolution. In single-gene testing, a particular gene is selected for Sanger sequencing according to the clinical presentation of the patient. The probability of a correct diagnosis is dependent on the clinician correctly identifying the condition and selecting the appropriate genetic test. Thus, this method is best suited for diagnosing highly distinct clinical conditions that are caused by just one or a few genes. This can be useful in case of, for example, LHON, caused by three distinct mtDNA point mutations in more than 95% of cases (Yu-Wai-Man et al., 2002). However, the more frequently observed weak phenotype-genotype correlation and sometimes incorrect interpretation of previous biochemical findings can enlarge the choice of candidate genes beyond feasibility, and even mislead it (Haack et al., 2012b). Indeed, the disadvantage of implementing Sanger sequencing for mitochondrial disorders is best seen in its low diagnostic yield of 11% (Neveling et al., 2013). Despite this shortcoming, it maintains an essential place in clinical genomics for at least two specific purposes. First, it is used to confirm NGS findings and inspect variant segregation in families, as data can be obtained within hours or a day and are easy to interpret. Second, it provides a means to “patch” the coverage of regions that are poorly covered by NGS.

1.2.5.3. Next generation sequencing

Since the completion of the Human Genome Project, the cost of NGS has decreased at a dramatic rate. At the same time, its accuracy, robustness, and handling have improved, making it a widely used alternative approach to direct Sanger sequencing. In addition, it is a non-invasive approach in a clinical setting, requiring just a few millilitres of EDTA blood or even a bloodspot card (Hollegaard et al., 2013). Its introduction to clinical diagnostics about 10 years ago has had a revolutionary impact, increasing diagnostic yield and moreover novel disease gene discovery. It has enabled a systematic overview of the full range of the potentially causal genetic variation in a single assay, resulting in faster genetic diagnosis.

Prosperities of the NGS technologies for human genetics could not have occurred without the systematic approaches to data sharing. They enabled and/or improved both the characterization of new disorders and the clinical interpretation of potential disease-causing variants (Claussnitzer et al., 2020). Global collaborative networks have been developed to share and match genetic and phenotypic data (for example, DECIPHER and GeneMatcher) (Firth et al., 2009; Sobreira et al., 2015). Open databases associate genes with the rare disorders (OMIM and ORPHANET) (Amberger et al., 2015; <http://www.orpha.net>), contain a clinical interpretation of submitted variants (HGMD, ClinVar and ClinGen) (Stenson et al., 2003; Landrum et al., 2018; Rehm et al., 2015) and offer patient records (DECIPHER and MyGene2) (Firth et al., 2009; <https://mygene2.org/MyGene2/>). In addition, comprehensive

catalogues of genetic variation across populations (such as gnomAD) enable the confident estimate of the variant frequency (Lek et al., 2016; Karczewski et al. 2020).

Whole genome sequencing (WGS) gives an overview of the entire genome. An alternative approach is to sequence only the exon, called whole exome sequencing (WES). It enables the sequencing of the protein-coding regions at a greater depth for a lower cost. Sequencing depth can become even greater for a lower cost by using a targeted sequencing panel, though here the genes to be sequenced need to be predefined.

1.2.5.4. Whole mtDNA sequencing

Application of NGS for whole mtDNA sequencing remains a common practice in many mitochondrial disease diagnostic centers as a first step before performing WES or WGS, especially in the adult-onset cases, where mtDNA etiology is much more common. It allows identification of the mtDNA variants with an accurate assessment of heteroplasmy levels (Tang et al., 2013). While doing so, one must be cautious of a choice of tissue to investigate, as many pathogenic mtDNA variants are restricted to the affected-tissues. It must also be considered that within easily accessible tissues such as the blood, a preferential selection process for wild-type mtDNA over mutated mtDNA over time is reported, diluting the diagnostic value of blood for the detection of mtDNA variants with age (Sue et al., 1998).

1.2.5.5. NGS gene panel

Focused gene panels contain a selected set of genes or gene regions that have known or suspected associations with the disease or phenotype under genetic investigation. They can be purchased with preselected content (predesigned) or custom-designed to include genomic regions of interest. Targeting offers sequencing of selected genes at high depth (500-1000x or higher), enabling identification of variants at low allele frequencies (down to 5%). However, panel sequencing applied in patients with suspected mitochondrial disorders is unable to achieve a high diagnostic yield, reaching only 8.8% when using a 447-gene panel (DaRe et al., 2013) or 15.2% with 132 genes (Legati et al., 2016). However, both groups analysed only about 130 previously undiagnosed patients. The application of an extended panel, such as the MitoExome, could identify known disease gene in 24% of the cases and potential novel disease gene in 31% of the patients (Calvo et al., 2012). However, given the limited shelf-life of panels, even this one lacks a growing number of disease genes that have been discovered since its design (Stenton and Prokisch, 2018).

1.2.5.6. Whole exome sequencing

WES is the technique of targeted sequencing, where the protein coding (exonic) regions of the genome are enriched. Although covering only about 2% of the human genome (Bamshad et al., 2011), this region harbors approximately 400 protein modifying rare variants per individual (Van Hout et al., 2019). In a diagnostic setting, WES has proven to be an efficient alternative to WGS due to lower costs and easier data storage and variant interpretation (Wang et al., 2013). Indeed, more than 85% of variants reported in ClinVar (Landrum et al., 2014) reside in the protein-coding regions and this percentage increases to 99% when intronic regions in their immediate proximity are included, making WES, although

restricted, a cost-effective sequencing in large-scale (van Dijk et al., 2014; Petersen et al., 2017). It covers reliably (at least 20 times) up to 97% of protein-coding regions and is in good concordance with Sanger sequencing (Sulonen et al., 2011). The advent of WES has accelerated the identification of pathogenic variants, but also novel disease genes, with a current diagnostic yield of 25–50% across patients with suspected Mendelian disorders (Yang et al., 2014; Taylor et al., 2015; Chong et al., 2015; Wright et al., 2018b). Mitochondrial diseases sit at the upper end of this diagnostic rate, so WES has become the first tier approach for their genetic diagnosis (Wortmann et al., 2017). Finally, WES can also be used to analyse the mtDNA, whose reads are captured “off-target” with high recall, precision, and comparable estimation of heteroplasmy levels (Wagner et al., 2019). However, WES is typically performed on blood where heteroplasmy levels can be low, so its negative result does not exclude the presence of mtDNA variants in other (affected) tissues.

The first example of WES application to the diagnosis of mitochondrial disease was published in 2010 when compound heterozygous variants in *ACAD9* were identified and subsequently functionally validated as a cause of severe complex I deficiency (Haack et al., 2010). In the following years, implementation of WES has led to diagnostic yield from 35 to 70% across different mitochondrial disease cohorts (Haack et al., 2012a; Taylor et al., 2014; Ohtake et al., 2014; Wortmann et al., 2015; Kohda et al., 2016; Pronicka et al., 2016; Legati et al., 2016; Puusepp et al., 2018; Theunissen et al., 2018) (Figure 6). The differences in the mean diagnostic yield are likely caused by different criteria when selecting a patient cohort. Larger cohorts are usually less clearly clinically defined. Accordingly, with bigger size cohorts, lower diagnostic yield is observed (Figure 6). Despite such variation, diagnostic yield in mitochondrial diseases is still one of the highest among the rare diseases (Clark et al., 2019). Overall, the observed impact of WES comes as a result of not only advances in sequencing technology, but also advances of bioinformatics tools that enable sequence alignment, annotation, variant calling, and filtering (Li and Durbin, 2009; Li et al., 2009; McKenna et al., 2010; Li, 2011), and shared databases of pathogenic variants (HGMD, ClinVar), and population sequences (gnomAD).

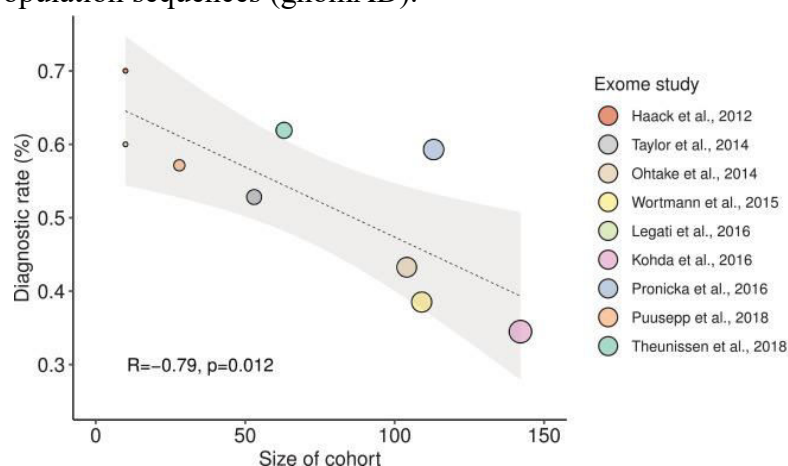


Figure 6. Diagnostic rate of WES in suspected mitochondrial disease across nine cohorts. Figure taken from Stenton and Prokisch (2020).

1.2.5.7. Whole genome sequencing (WGS)

WGS provides sequence information for almost the entire genome (98%) (Alfares et al., 2018). One main diagnostic advantage of WGS over WES lays in the identification of structural variants (SV), defined as DNA rearrangements involving a minimum of 50 nucleotides: copy number variants (CNVs), translocations and inversions; and deep intronic variants (Lightowers et al., 2015; Fang et al., 2014; Belkadi et al., 2015; Meienberg et al., 2016). The workflow is less labor intensive as it is PCR free and requires less starting material. Moreover, the coverage is more even and consequently less biased. Nevertheless, long repetitive elements or complex regions remain a challenge to sequence (Dewey et al., 2014; Goodwin et al., 2016). A systematic analysis undertaken by Ebbert et al. (2019) identified 2855 “dark regions” across 748 protein-coding genes in WGS data. WGS analyses across paediatric patients cohorts have reached a diagnostic yield of 21-63% (Jiang et al., 2013; Gilissen et al., 2014; Taylor et al., 2015; Stavropoulos et al., 2016; Scocchia et al., 2019, Thiffault et al., 2019; Liu et al., 2019). Within the UK Biobank, diagnostic yield of 16.1% was reported across the cohort of 9,802 patients with rare diseases (Turro et al., 2020). Although WGS’s potential to reveal entire genetic variation looks promising in a diagnostic setting, two main aspects make its result’s clinical interpretation a formidable challenge (Biesecker and Green, 2014). First, WGS detects 3.4 to 5 million variants per individual, significantly hampering the variant filtering and interpretation (Natarajan et al., 2018). Even if focused on a variant detected just once, one would meet daunting 5,000 variants (Natarajan et al., 2018). Second, it detects variants within poorly structurally and functionally annotated regions. Other challenges include costs, speed of delivery, and data storage (Taylor et al., 2015). Only a few studies compared the utility of WES and WGS in diagnostics of genetic diseases. Although Lionel et al. (2018) reported a WGS diagnostic yield 26% higher than WES in 103 patients with a suspected genetic disorder, another report with a more modest hit of 7% in 108 patients suggests reanalysing WES before opting for WGS (Alfares et al., 2018). A single reported application of WGS to mitochondrial disease cohort of 40 patients yielded a diagnosis in 55% of the cases (Riley et al., 2020). To summarize, the potential value of WGS in clinical diagnostics is unquestionable, however, many outstanding challenges are still complicating its routine application in diagnostics and limiting it to the research setting.

1.3. WES analysis

Broad implementation of NGS in the last decade has led to the generation of detailed catalogues of genetic variation in both the general population and disease cohorts. With the increased amount of data, identified variants must be subjected to rigorous analysis to separate disease-causing or -associated variants from the broader number of variants that are rare, potentially functional, but not pathogenic (MacArthur et al., 2014).

1.3.1. Variant detection

After the massively parallel sequencing has been performed, the raw WES data in the form of short sequence reads are stored as a vcf file. Its computational processing includes

sequence alignment to a reference genome, variant calling based on differences between the input and reference sequence, filtering out low-quality variants, and annotation of the remaining high-quality variants.

Sequence alignment encompasses the mapping of each read pair to the hg19 reference genome, a synthetic single-stranded representation of common human genome sequence. For this thesis, the mapping was performed with Burrows-Wheeler-Alignment tool (Li and Durbin, 2009). It is a software package based on a backward search with Burrows-Wheeler Transform which aligns short reads to the reference genome, allowing mismatches and gaps. The output alignment is in the SAM or BAM format.

Variant callers used in this study were SAMtools and GATK Haplotype caller. SAMtools employs Bayesian models and detects all single nucleotide variations (SNVs) and short indel variants (Li et al., 2009). GATK HaplotypeCaller calls simultaneously SNVs and indels by discarding existing mapping information and performing a local *de novo* assembly of haplotypes whenever it encounters a region that shows variations (active region) (McKenna et al., 2010). This allows more accurate calling of indels, SNVs, and regions that contain different variants close to each other (Hwang et al., 2015).

Calling SVs, especially CNVs, defined as deletions or duplications larger than 1 kb, is a well-known weakness of all short-read sequencing technologies (Tan et al., 2014). Several strategies- read depth, split-read, paired-end, assembly, and a combination of listed- have been developed to improve CNVs detection in paired-end NGS data (Medvedev et al., 2009; Tan et al., 2014). For this study, ExomDepth was used, which is based on a hidden Markov model, comparing the read depth of an exon in a sample to the read depth of the respective exon in ~10 reference sets (Plagnol et al., 2012). It showed the highest sensitivity compared to other commonly used tools, but also has a high false positive rate (Tan et al., 2014). Thus, its results should be in-depth inspected.

1.3.2. Variant annotation

During the standard computational processing, detected variants are filtered for the following quality criteria: coverage ≥ 8 , minimum 3 reads indicating a variation in $\geq 5\%$ on both strands, and a Phred-like consensus quality of ≥ 20 (Haack et al., 2010; Ng et al., 2010). Variants that have passed the filter are annotated by ANNOVAR (Wang et al., 2010), SnpEff (Cingolani et al., 2012), or customized in-house tools as in this study. While the consequences of most of the frameshift and nonsense variants are quite self-evident, missense variants are much more difficult to interpret as the amino acid substitution can affect the structure or function of a protein in different ways, or may not affect them at all (Thusberg and Vihinen, 2009). As most of the variation in the human genome arises as a single nucleotide polymorphism (SNP), many computational methods have been developed to predict the effect of a variant (Thusberg et al., 2011). Based on measuring the effect on protein structure/function (MutationTaster, Schwarz et al., 2010; PolyPhen-2, Adzhubei et al., 2010), sequence conservation (SIFT, Kumar et al., 2009), and overall pathogenic potential by a combination of these sources of information (CADD, Rentzsch et al., 2019), these tools provide scores that estimate the deleteriousness of the variant. However, while useful, one

should always cautiously approach their results for several reasons. Even if applied to the same variant, their scores can correlate poorly (Thusberg et al., 2011).

1.3.3. Variant prioritization and validation

With the advent of WES and WGS, variant prioritization and interpretation has become a critical step in establishing a molecular diagnosis. Due to a large number of variants identified per individual, finding the one or two variants responsible for a Mendelian disease can be referred to as a “needle in the haystack” problem (Cooper and Shendure, 2011). On average, WES reveals 85,000 single nucleotide variants (SNVs) per individual (Bamshad et al., 2011; Wieland, 2015). Focusing on their effect on proteins, ~400–500 are classified as protein-modifying rare variants (Choi et al., 2009; Ng et al., 2009), ~100 as loss-of-function (LoF) variants (Robinson et al., 2014), and ~60 as novel protein-altering variants, still not reported in databases (Volk and Kubisch, 2017). Most of them are common polymorphisms and some even sequencing errors (Bamshad et al 2011; Wieland, 2015). Considering such an amount of genetic variation, it is clear that variants must go through an intensive evaluation through comprehensive filtering steps and prediction tools in order to point out the true disease-causing variant(s).

To help, improve, and standardize variant classification, guidelines have been created. MacArthur et al. (2014) proposed a two-step process during the investigation of the causality of a specific disease. First, disease-implication of the gene in which the variants fall is considered using statistical support from genetic analyses, supported by informatics sources and functional analysis. Second, the variant’s causality is assessed through a combination of genetic, experimental and informatics supportive evidence. As the re-evaluation of published cases revealed false assignments or lack of direct evidence for pathogenicity (Bell et al., 2011; Xue et al., 2012; Norton et al., 2012) these evaluations are recommended even if the discovered variant has been previously reported. In 2015, the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) published widely implemented standards and recommendations for the variant interpretation of sequence variants (Richards et al., 2015). This report recommends the classification of variants into five categories: pathogenic, likely pathogenic, uncertain significance (VUS), likely benign, and benign. Moreover, it describes the process behind the classification of variants based on different sets of evidence for variant interpretation. Two sets of criteria are proposed- benign and pathogenic, each organized by a type and strength (Figure 7). If a variant does not fulfil a sufficient number of the criteria or the evidence is conflicting, it is categorized as a VUS by default.

	Benign			Pathogenic		
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population Data	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>			Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
Computational And Predictive Data		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>
Functional Data	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>	
Segregation Data	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
De novo Data				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
Allelic Data		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>		
Other Database		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
Other Data		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>			

Figure 7. ACMG's proposed evidence framework for evaluating variant classification. Each criterion is organized by the type of evidence and the strength of the criteria for a classification of variant as benign or pathogenic. Abbreviations: BS, benign strong; BP, benign supporting; FH, family history; LOF, loss-of-function; MAF, minor allele frequency; path., pathogenic; PM, pathogenic moderate; PP, pathogenic supporting; PS, pathogenic strong; PVS, pathogenic very strong. Evidence code descriptions can be found in Richards et al. (2015), where the figure was taken from.

The frequency of a variant in the general population is one of the key criteria for its clinical interpretation as its rarity is a prerequisite for pathogenicity in Mendelian disorders (Kobayashi et al., 2017). The scientific community defines variant as rare when its minor allele frequency (MAF) is less than 1% in a control population (MacArthur et al., 2014). A variant with a frequency higher than 1% is defined as a polymorphism. However, depending on the investigated disease, more stringent or more relaxed filtering can be used. Without careful consideration, too stringent thresholds may increase the risk of incorrectly classifying pathogenic variants as benign. Well-known examples include founder mutations *BRCA1* c.68_69delAG and *BRCA2* c.5946_5949delTGGGA with MAFs of 1% among Ashkenazi Jews (Levy-Lahad et al., 1997) and *CFTR* p.Phe508del with MAF of 1% in the European population (Sosnay et al., 2014). However, it is important to note that pathogenic variants with such MAFs are typically well documented. gnomAD, the latest version containing 125,748 exomes and 71,702 genomes, currently presents the most comprehensive database for assessing the variant frequency (Karczewski et al., 2020). It is commonly used for control population together with the in-house databases. In-house cohorts, although limited concerning the statistical power, enable correction for systematic errors of the WES pipeline

and evaluation of high frequency variants present in isolated populations due to founder effect. By filtering for a MAF <0.01 , the number of variants can be reduced to ~ 250 . The variants can be further filtered based on the assumed pattern of inheritance. As most mitochondrial diseases have an early-onset and follow a recessive mode of inheritance, a search for only potential biallelic- homozygous or compound heterozygous- variants can be employed. Homozygosity is more probable to occur in consanguineous families. With this filter, 5-25 genes with potentially biallelic variants are typically presented. Familial “trio” sequencing (index, mother and father), although not of routine practice, facilitates phasing of haplotypes, the detection of *de novo* and autosomal dominant variants, and reduces the number of candidates by ten-fold when compared with sequencing only the proband (McRae et al., 2017). An autosomal dominant pattern of inheritance should be suspected in certain mitochondrial diseases, such as progressive external ophthalmoplegia (PEO, Chinnery, 1993). Generally, very few pathogenic *de novo* mutations in nDNA have been described for mitochondrial diseases (Thompson et al., 2016; Alston et al., 2017). On the other hand, 25% of mtDNA point mutations and most of the large-scale mtDNA deletions occur sporadically (Alston et al., 2017; Sallevelt et al., 2017). Next, variants can be further prioritized based on their presence and disease association(s) in public databases such as HGMD and ClinVar. There are currently 88641 variants enlisted as pathogenic in ClinVar (summarized in Figure 8). Previously unreported variants in known disease-associated genes reported in OMIM can be pinpointed as currently $\sim 5,800$ genes are associated with a particular Mendelian disease.

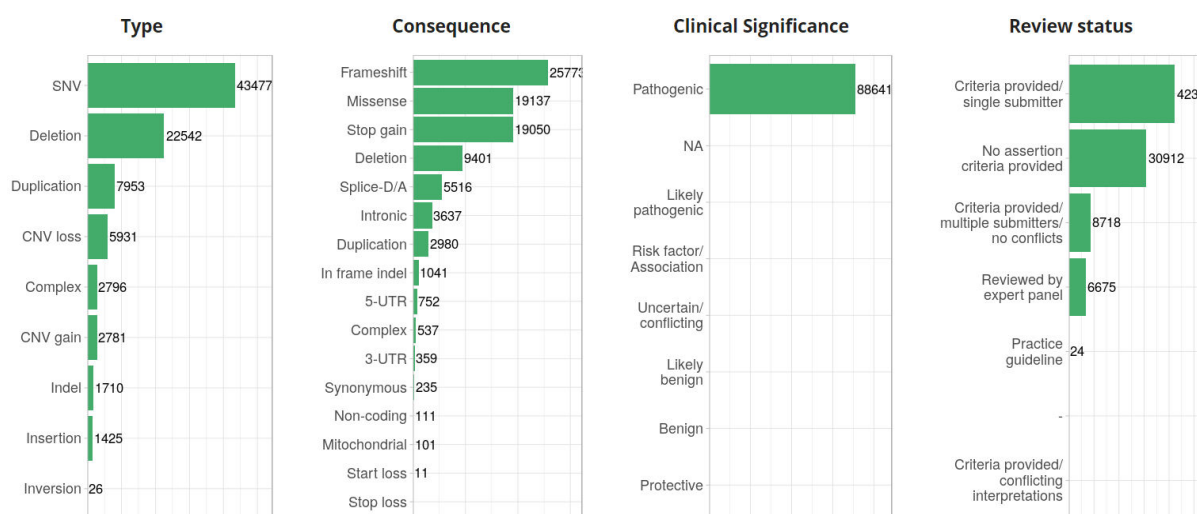


Figure 8. Summary of reported pathogenic variants in ClinVar (<http://simple-clinvar.broadinstitute.org/>).

For the mitochondrial diseases, one can prioritize for the genes encoding proteins with predicted mitochondrial localization according to the MitoP2 or MitoCarta databases (Eltner et al., 2009; Calvo et al., 2016). This approach is promising in revealing novel disease genes. Finally, an unbiased WES approach enables the detection of disease-causing non-mitochondrial genes that mimic mitochondriopathy clinical presentations (Panneman et al., 2018). To summarize, identifying disease-causing variants is a complex, multistep process, highly automated at early steps, but left to human interpretation in the final, critical steps. An overview of the analysis workflow is depicted in Figure 9.

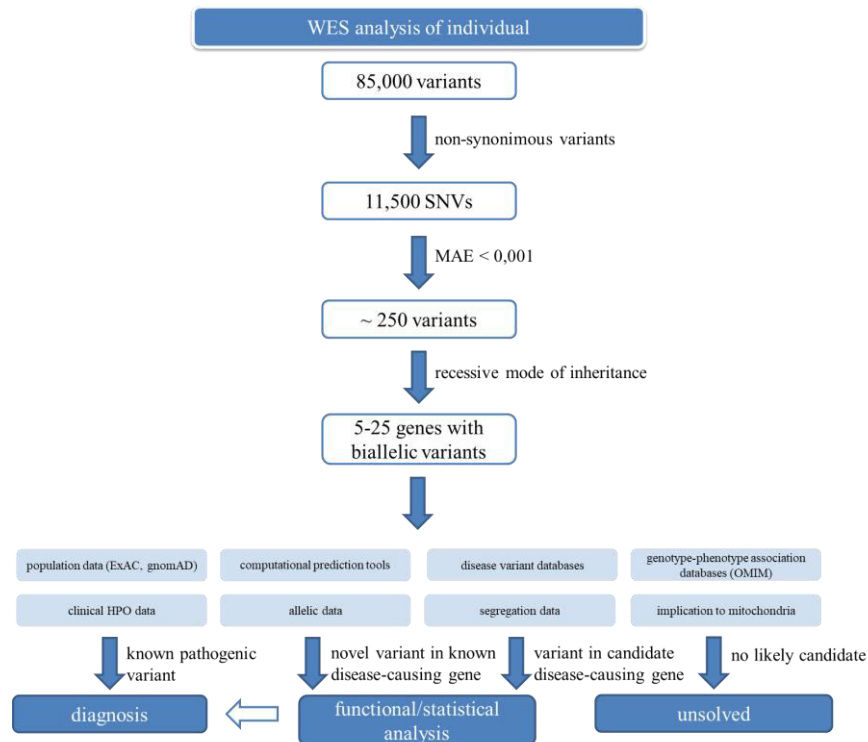


Figure 9. WES analysis workflow.

1.3.4. Validation of VUS

In the case of VUS, WES findings should be validated by additional functional experiments. Although computational tools can strongly indicate variant's pathogenicity, they are still only an indicator and conclusive evidence can only be provided by functional studies. To start with, the identified variant(s) should be verified by Sanger sequencing, to ensure that the WES finding is indeed present in the patient, and not a result of sequencing errors or sample mixing/swap (Haack et al., 2010), and done together with segregation analysis in family members to confirm the expected mode of inheritance. The type of variant may hint which mechanism lies behind its deleterious effect. Variants at direct splice sites most likely affect splicing, so the transcript analysis by reverse transcriptase PCR (RT-PCR) or RNA-sequencing (RNA-seq) is suggested. Stop-gain variants likely result in a truncated protein or no protein at all due to transcript degradation. A non-synonymous (missense) variant in the functional domain can dramatically alter the protein structure or function. These predictions

can be assessed by Western blot, BN-PAGE, enzymatic activity assays, quantitative proteomics, and measurements of the oxygen consumption rate, amongst others (Haack et al, 2010; Wittig et al., 2006; Spinazzi et al., 2012; Brand and Nicholls, 2011; Yopez et al., 2018). Experimental proof of the variant's role in biochemical defect can be provided via the rescue of the observed phenotype in the patient-derived cells by over-expression of a wild-type copy of the gene (Kremer and Prokisch, 2017). Especially in cases of a novel genotype-phenotype association, the collection of unrelated patients with a common phenotype and likely pathogenic variants may provide more convincing accumulating evidence.

1.3.5. Reasons behind inconclusive WES analysis

Although WES presents a recognizable tool in a routine diagnostic setting, 30–65% of mitochondrial disease patients are left without a genetic diagnosis. There are a number of possible reasons/explanations for this.

- 1) **Pleiotropy:** a phenomenon in which a single gene contributes to multiple phenotypic manifestations. Gene associated with a certain phenotype may cause a different, yet unreported one, in the inspected patient.
- 2) **Genetic heterogeneity:** in contrast to pleiotropy, this is a phenomenon in which a specific phenotype or a genetic disorder may be caused by multiple variants within the same gene (“allelic”) or by variants in the different gene loci (“locus” heterogeneity). The phenotype associated with some gene(s), may be also caused by other, already reported disease genes. A good example of genetic heterogeneity in mitochondrial disease is Leigh syndrome.
- 3) **Incomplete penetrance:** penetrance is defined as the percentage of individuals carrying a specific variant or genotype that exhibit the phenotype of the associated disorder or genotype (Cooper et al., 2013). Incomplete or reduced penetrance indicates that some individuals do not present with the phenotype, even though they carry the phenotype-causing variant or genotype. Although reduced penetrance is usually considered just for the diseases caused by a defect in mtDNA, where reported penetrance is just 10-50% (Bianco et al., 2016; Grady et al., 2018), this might apply also to the other cases. It implies that variant(s) present in both the patient and healthy control(s) may still be disease-causing for the patient. Detection of such pathogenic variants is particularly difficult due to lack of segregation with affected individuals in the pedigree, and higher than expected allele frequency in population.
- 4) **Variable expressivity:** defined as the extent to which a genotype is phenotypically expressed in individuals. Individuals with the same variants may exhibit differences in disease severity, even within the same family, and are thus difficult to prioritize.
- 5) **Other more complex pathomechanisms:** the cause of the disease may be due to differences in imprinting or be polygenic instead of monogenic. Besides, the severity of the disease can be influenced by epistasis- a phenomenon where the effect of one gene is dependent on the presence of one or more “modifier genes”, which can ameliorate or reinforce the expression of their target (Nagel, 2005). Such mechanisms

are usually not considered and their detection is hampered by the reduced power of usual genetic analysis.

- 6) Targeted sequencing: the disease-causing variants may be located in non-coding regions to which WES is blind (Wortmann et al., 2015).
- 7) Technical limitations: WES targets around 97% of exons of which ~10% do not reach the coverage at a sufficient level for clinical interpretation (Yao et al., 2017). Overall, WES displays limitations in sequencing and variant calling the following types of genetic variations: large rearrangements, large CNVs, variants in repetitive or high GC rich regions, variants in genes with corresponding pseudogenes, and other highly homologous sequences.
- 8) Variant prioritization and interpretation: WES analysis relies on filtering approaches which rely on automated assumptions about variants and their potential role in disease. Although most variants can be quickly classified as benign or likely benign based on the MAF, each WES analysis yields several VUS. They demand a follow-up analysis in order to understand their role and prove their causality. Although reporting of the VUS depends on the local policies, an astonishing 41% of all variants reported in ClinVar are asserted in this criterion.

1.4. Krebs cycle and its deficiencies

Since its elucidation in the 1937 (Krebs and Johnson, 1980), the Krebs cycle (also known as the citric acid cycle (CAC) and the TCA cycle (tricarboxylic acid cycle)) has been considered as a key integrator of carbohydrate, fat, and protein metabolism. In short, the cycle starts with a product of catabolism- acetyl-CoA. Although mostly produced by the activity of pyruvate dehydrogenase from pyruvate arising from glycolysis, acetyl-CoA can also come from fatty acid degradation or several amino acids (glutamate, alanine, etc.). Through a series of reactions carried out by nine enzymes in the mitochondrial matrix, acetyl-CoA is completely oxidized and the final products of a single cycle are one GTP (or ATP), three NADH, one FADH₂, and two CO₂ (Figure 10). The generated NADH and FADH₂ are fed into the OXPHOS. Krebs cycle flux is regulated on different levels. Most of its regulation comes from the activities of its enzymes citrate synthase, isocitrate, and α -KG dehydrogenase (α -KGDH). However, substrate supply, NADH/NAD⁺ ratio, coenzyme A availability, matrix phosphorylation potential, and Ca²⁺ concentration are also limiting factors regulating several steps (Briere et al., 2006). Krebs cycle enzymes consist of 15 proteins, all encoded by nuclear genes. Four of these enzymes, including malate dehydrogenase (MDH), possess both a mitochondrial and a cytosolic form. Interestingly, Krebs cycle metabolites have been shown in recent years to have additional, non-metabolic, roles as signaling molecules (reviewed in Martinez-Reyes and Chandel, 2020).

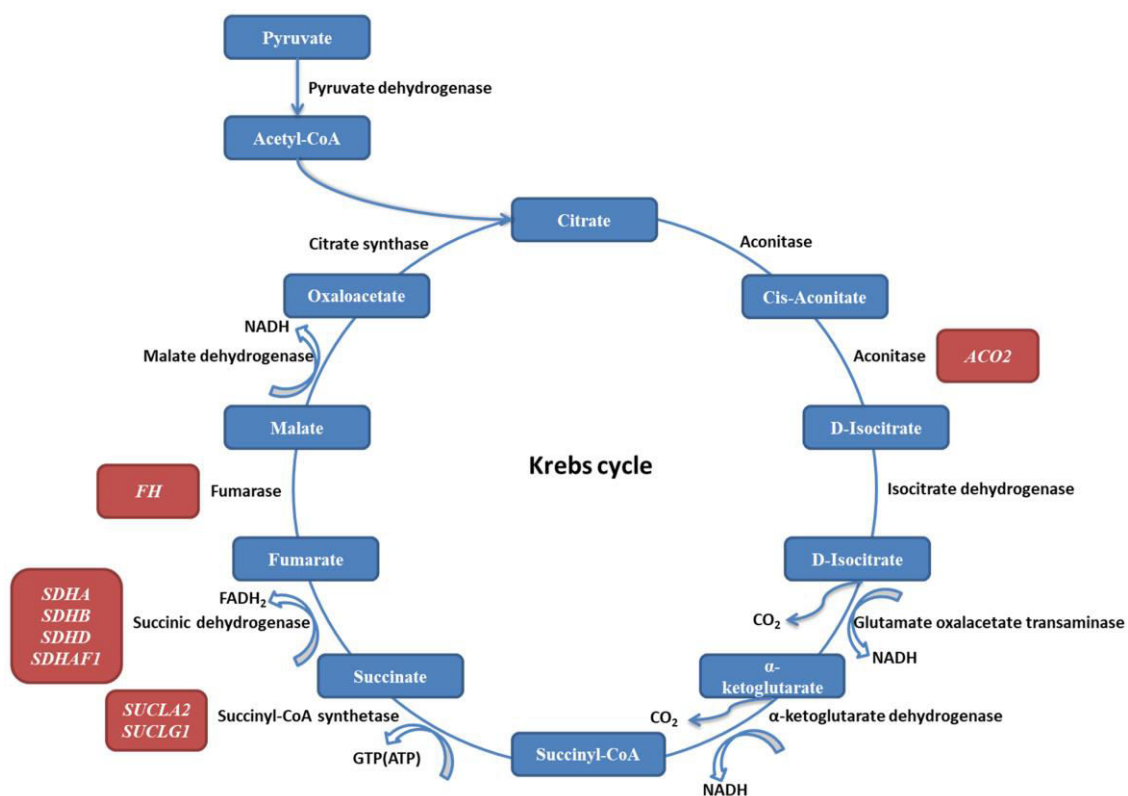


Figure 10. Krebs cycle. Diagram showing the substrates and enzymes involved in the cycle, with reported disease genes in red.

Due to the crucial role in cellular metabolism, it is not surprising that Krebs cycle's enzyme deficiencies have been seldom reported. So far, succinyl-CoA synthetase (SCS) deficiencies, succinate dehydrogenase (SDH) deficiency, fumarase deficiency, and alpha-ketoglutarate dehydrogenase (α -KGDH) deficiency (also known as 2-oxoglutaric aciduria) (summarized in Table 1) have been reported. They are caused by pathogenic variants in *ACO2* (aconitase 2, MIM: 100850); the subunits of the succinyl-CoA synthetase (*SUCLG1* (MIM: 611224) and *SUCLA2* (MIM: 603921)); SDH subunits (*SDHA* (MIM: 600857) and *SDHB* (MIM: 185470)), additional structural component *SDHD* (MIM: 602690), and assembly factor *SDHAF1* (MIM: 612848); and *FH* (fumarate hydratase, MIM: 136850), all of which follow an autosomal recessive mode of inheritance (Figure 10).

Variants in *ACO2* have been described to cause cerebellar-retinal degeneration in childhood (Spiegel et al., 2012). SCS deficiency is unique for its involvement in both the Krebs cycle, and mtDNA maintenance. Ostergaard et al. (2007) hypothesized that the observed mtDNA depletion occurs as decreased mitochondrial nucleoside diphosphate kinase (NDPK) activity is unable to form a complex with SUCL. Pathogenic *SUCLG1* variants lead to early-onset encephalomyopathy with methylmalonic aciduria and mtDNA depletion (Ostergaard et al., 2007). *SUCLA2* variants have been described in patients with encephalopathy, Leigh syndrome, with or without methylmalonic aciduria and mtDNA depletion (Elpeleg et al., 2005). Also, severe sensorineural hearing impairment is particularly associated with the *SUCLA2* defect (Carrozzo et al., 2007). When it comes to SDH, apart

from catalyzing the conversion of succinate to fumarate, it is also part of the OXPHOS as complex II (succinate-ubiquinone oxidoreductase). Therefore, pathogenic variants in its protein-coding genes lead to isolated complex II deficiencies. Pathogenic *SDHA* variants are mostly associated with Leigh syndrome (Bourgeron et al., 1995), and *SDHB*, *SDHD*, and *SDHAF1* variants are involved in myopathy, encephalopathy, leukodystrophy, and cardiomyopathy (Ghezzi et al., 2009; Alston et al., 2012; Jackson et al., 2014). *FH* variants have been described as a cause of fumarase deficiency, characterized by severe and early-onset encephalopathy with seizures and muscular weakness leading to growth and developmental delay, and often death in the first years of life (Bourgeron et al., 1994; Ottolenghi et al., 2011). Finally, α -KGDH deficiencies have been described since the 1990s, but without a clear genetic diagnosis. There are two distinct phenotypes. The first one is an encephalopathic form related to the E2 subunit deficiency. Bonnefont et al. (1992) reported three siblings with hypotonia, metabolic acidosis, and hyperlactatemia, dying at about 30 months of age due to neurological deterioration. Guffon et al. (1993) reported two siblings with α -KGDH deficiency presenting a progressive, severe encephalopathy with axial hypotonia, psychotic behavior, pyramidal symptoms, failure to thrive, and permanent lactic acidosis. The second form exhibits clinical features of the DOOR syndrome (deafness, onychoosteodystrophy, thumbs and sensorineural deafness), and is due to a defect of the E1 subunit (Surendan et al., 2002).

Regarding the diagnosis of these disorders, the most important diagnostic clues reside in organic analysis of the urine, as an abnormal urinary excretion of organic acids frequently occurs. This is exemplified by the excretion of α -KG, suggesting that the activity of α -KGDH is decreased, potentially due to the sequestration of CoA between various CoA esters in the case of SDH and fumarase deficiency (Briere et al., 2006). In addition, fumarase deficiency can occur with increased excretion of fumarate associated with succinate and lactate excretion. In SCS defects mild methylmalonic aciduria with abnormal urine carnitine ester profile is reported, with only mild abnormalities of Krebs cycle intermediates (Morava and Carrozzo, 2014).

Krebs cycle and its deficiencies

Table 1. Clinical features of reported patients with Krebs cycle deficiencies.

Krebs cycle deficiencies									
Gene	<i>ACO2</i>	<i>SDHA</i>	<i>SDHB</i>	<i>SDHD</i>	<i>SDHAF1</i>	<i>SUCLA2</i>	<i>SUCLG1</i>	<i>FH</i>	?
Diagnosis	ACO2 defect	SDH defect	SDH defect	SDH defect	SDH defect	SCS deficiency	SCS deficiency	fumarase deficiency	α-KGDH deficiency
Number of patients	10	20	4	2	5	23	4	5	9
Onset	neontal, infancy, childhood	infancy	infancy	infancy	infancy	infancy	infancy	neontal, infancy	neontal, infancy
Choreoathetosis	+					±	±		±
Dystonia		±	±	+	+	±	+	+	±
Encephalopathy	±	±	+	+	+	+	+	±	
Leukodystrophy	+	±	+		+				
Hypotonia , axial	+	±	+	+	+	±	+	+	+
Neurological symptoms	+	±	±	+	+	+	+	+	+
Seizures	+	±	-	+	+	±	±	+	
Pyramidal signs		±	+		+	±	+	+	+
Retardation, psychomotor	±	±	+	+	+	+	+	+	+
Muscle weakness		±	±	+	+	+		±	
Failure to thrive	±	±	+	+	+	+	+	+	+
Cardiomyopathy		±	±	±			±		
Liver dysfunction						±	±	+	+
Optic atrophy	+	±						±	
Impaired vision	+	±		+				±	
Speech delay					+			+	
Deafness, sensorineural						±	±		±
Dystrophic thumbs									±
Osteodystrophy	-	+	+			+	+	+	+
Lactic acidosis/ elevated lactate	-	+	±	+	+	+	+	+	+
Special laboratory findings		CII defect	CII defect	CII defect	CII defect	methylmalonic aciduria ±	methylmalonic aciduria +	2-Oxoglutaric acid (U) ±; Fumaric acid (U); Succinic acid (U) ±	2-Oxoglutaric acid (U) +
Other features	cerebral and cerebellar degeneration	Leigh syndrome, cardiomyopathy	reduced penetrance			mtDNA depletion, Leigh syndrome ±	mtDNA depletion	dysmorphic features, metabolic acidosis	

*Abbreviations: CII, complex II; U, urine.

1.5. OXPHOS, complex III and its deficiencies

The OXPHOS system presents the final biochemical pathway involved in ATP production. Located in the inner mitochondrial membrane, it is composed of five multiprotein enzyme complexes (I–V) and two electron carriers: coenzyme Q (CoQ) and cytochrome c (cyt c) (Hatefi, 1985; Figure 11). In short, metabolism of sugars, proteins and fats results in the formation of NADH and FADH₂ during catabolic processes such as glycolysis, fatty acid oxidation, and Krebs cycle. These molecules are energy-rich as they contain a pair of electrons with a high transfer potential. Within OXPHOS, these electrons are passed from NADH or FADH₂ by four protein complexes within the electron transport chain (ETC) and finally used to reduce O₂ to water. The flow of electrons is accompanied by the pumping of protons out of the mitochondrial matrix. This results in an uneven distribution of protons, generating an electrochemical proton gradient (Δp) across the inner mitochondrial membrane. The H⁺-pumping ETC complexes are: complex I (CI, NADH-ubiquinone oxidoreductase), complex III (CIII, ubiquinol-cytochrome c oxidoreductase, cytochrome bc₁ complex) and complex IV (CIV, cytochrome c oxidase). These transmembrane complexes contain multiple oxidation-reduction centers, such as quinones, flavins, iron-sulfur clusters, hemes, and copper ions. Complex II (CII, succinate-ubiquinone oxidoreductase) indirectly contributes to the proton gradient via the reduction of the ubiquinone (Q) pool. The final phase of OXPHOS is carried out by complex V (CV, ATP synthase), an ion channel that enables a proton flux to go back into the mitochondrial matrix while synthesizing ATP. This presents a “chemiosmotic theory” (Mitchell, 1961). Within the ETC, CI, CIII and CIV tend to physically interact and form the supercomplexes together. The most abundant of these supercomplexes, consisting of CI₁III₂IV₁ together with CoQ and cyt c, has been termed the “respirasome” (Gu et al., 2016; Letts et al., 2016; Wu et al., 2016). In addition, CI can also be found with CIII₂ alone (CI₁III₂) (Schagger and Pfeiffer, 2001), and CIII₂ with CIV (CIII₂IV₁) (Schagger and Pfeiffer, 2000; Iwata et al., 1998). Finally, CIV and CV can also form dimers together (CIV₂V₂) (Allegretti et al., 2015). The ratio of supercomplexes varies between tissues and species (Schagger and Pfeiffer, 2000) and alters in response to the cellular metabolic demands (Greggio et al., 2017). Acin-Perez et al (2008) proposed a “plasticity model”, suggesting that the biogenesis of supercomplexes occurs by incorporation of previously fully assembled individual complexes.

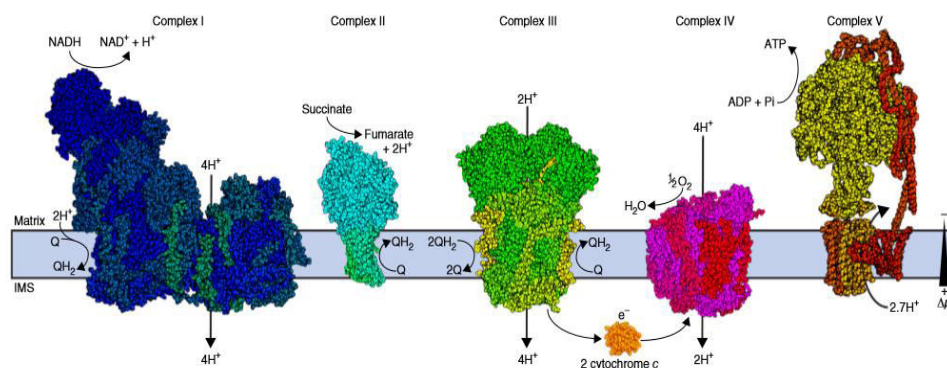


Figure 11. The mammalian OXPHOS-ETC. The five complexes and cytochrome *c* are shown as experimentally determined atomic structure models. Abbreviations: Q, ubiquinone; QH₂, ubiquinol. Figure taken from Letts and Sazanov (2017).

CIII has a central position in the ETC, as it receives electrons from CI and CII through CoQ and transfers them to CIV via cyt *c*. This redox reaction is coupled to the proton translocation from the mitochondrial matrix to the intermembrane space via the “Q-cycle,” adding to the proton gradient (Crofts et al., 2008). In addition, it is also at the crossroads of additional electron transfer pathways, such as glycerol-3-phosphate dehydrogenase, electron transfer flavoprotein (ETF), sulfide–quinone reductase (SQR), and dihydroorotate dehydrogenase (DHODH), that all converge onto CoQ (Lenaz et al., 2007). Mammalian CIII is an enzymatic complex with a symmetrical dimeric structure (CIII₂) where each monomer consists of 10 subunits: three catalytic core (CYB, CYC1 and UQCRFS1) and seven supernumerary subunits (Iwata et al., 1998; Smith et al., 2012; Xia et al., 2013). CYB contains two b-type haems with different redox potential and two CoQ binding sites. CYC1 binds a hem c1 group that transfers electrons to cytochrome *c*. The intermembrane-space-facing C-terminus of UQCRFS1 contains a Fe–S cluster. Assembly of CIII is well described in yeast. The initial and the final stages are well conserved between humans and yeast, so it is believed that they are also similar in intermediate steps that are still unknown in humans (Fernandez-Vizarra and Zeviani, 2015) (Figure 12). It starts with the synthesis of cytochrome b/CYB in humans in mitochondria and its insertion into the inner membrane, mediated by assembly factors Cbp3/UQCC1 and Cbp6/UQCC2 that stay bound to CYB after it is completely synthesized. Cbp4/UQCC3 joins after the first haem-b is incorporated (Gruschke et al., 2011; 2012; Hildenbeutel et al., 2014). Once the structural subunits UQCRB and UQCRQ get incorporated, UQCC1 and UQCC2 leave the site of CIII synthesis (Gruschke et al., 2011; 2012). Soon the dimerization occurs and the pre-CIII₂ complex is formed. CIII maturation occurs with the addition of the Rieske Fe–S protein (Rip1/UQCRFS1). UQCRFS1 is synthesized as a pre-protein in the cytosol and imported into the mitochondrial matrix via its cleavable N-terminus mitochondrial targeting sequence (MTS). After the import, UQCRFS1 is stabilized by the chaperone LYRM7 (Sanchez et al., 2013). This LYRM7-UQCRFS1 intermediate then recruits a Fe-S transfer complex, consisting of co-chaperone HSC20, HSPA9, and ISCU, through the direct binding of HSC20 to LYRM7, to enable the Fe-S cluster incorporation into UQCRFS1 (Maio et al., 2017). After Fe-S cluster acquisition,

BCS1L translocates and incorporates UQCRFS1 into the pre-CIII₂ complex, making it mature and catalytically active (Fernandez-Vizarra and Zeviani, 2018). Interestingly, only then does the cleavage of the UQCRFS1 MTS occur (Maio et al., 2017). Although crystal structures from bovine heart CIII revealed the presence of peptides derived from UQCRFS1 N-terminus localized between the UQCRC1 and UQCRC2 (Brandt et al., 1993; Iwata et al., 1998; Fernandez-Vizarra and Zeviani, 2018), recent studies showed that their marked accumulation in the absence of TTC19 becomes detrimental to CIII function (Bottani et al., 2017). The current assumption is that the UQCRFS1 MTS is processed *in situ* upon its incorporation into CIII₂ and that this cleavage produces several peptides that remain bound to CIII₂. Later, these peptides need to be removed to keep CIII₂ structural integrity and function in a still not well-defined process favored by TTC19 (Bottani et al., 2017).

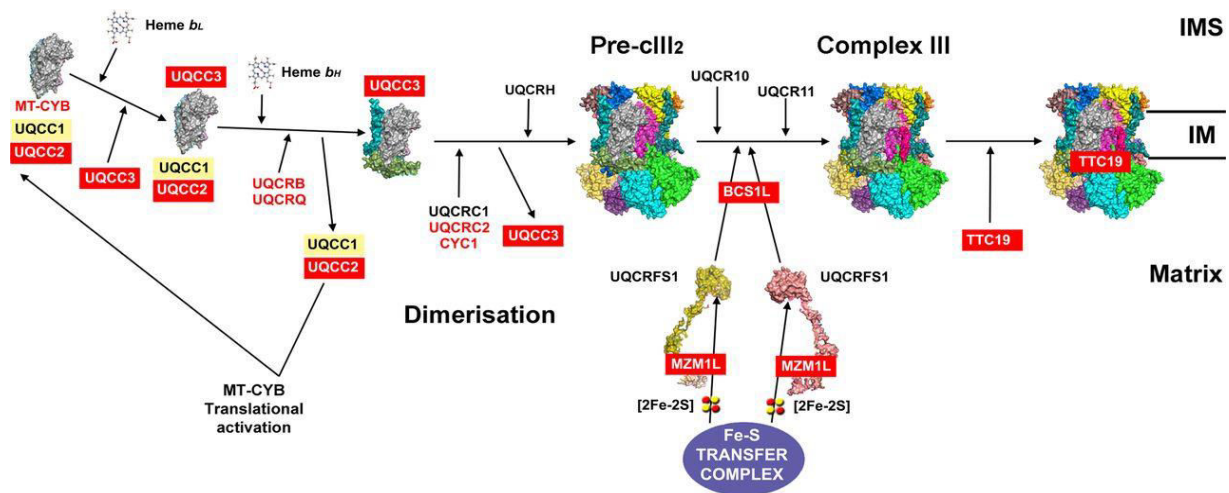


Figure 12. Complex III assembly model. Red color indicates proteins with reported pathogenic variants in their encoding genes. Abbreviations: IMS, inner membrane space, IM, inner membrane. Figure taken from Signes and Fernandez-Vizarra (2018).

Isolated CIII deficiencies belong to the least frequently diagnosed mitochondrial disorders. So far, pathogenic variants in genes encoding five subunits (*MT-CYB* (MIM: 516020; Andreu et al., 1999; Schuelke et al., 2002), *CYC1* (MIM: 123970; Gaignard et al., 2013), *UQCRC2* (MIM: 191329; Miyake et al., 2013), *UQCRB* (MIM: 191330; Haut et al., 2003), and *UQCRQ* (MIM: 612080; Barel et al., 2008)) and five assembly factors (*UQCC2* (MIM: 614461; Tucker et al., 2013; Feichtinger et al., 2017), *UQCC3* (MIM: 616097; Wanschers et al., 2014), *LYRM7* (MIM: 615831; Invernizzi et al., 2013), *BCS1L* (MIM: 603647; de Lonlay et al., 2001; Visapaa et al., 2002; Hinson et al., 2007) and *TTC19* (MIM: 613814; Ghezzi et al., 2011) have been reported across more than 140 patients with heterogeneous clinical presentations (Benit et al., 2009; Fernandez-Vizarra and Zeviani, 2015; Feichtinger et al., 2017) (summarized in Table 2; Figure 12). It is impossible to differentiate subunit defects from assembly factor defects based on phenotype solely. All reported individuals follow an autosomal recessive mode of inheritance, with the exception of pathogenic variants in the mtDNA-encoded *MT-CYB*. It is noteworthy that CIII defects, except for *TTC19*, often occur as combined respiratory chain deficiencies together with CI

and, occasionally, CIV deficiencies (Moran et al., 2010; Tucker et al., 2013; Carossa et al., 2014), possibly due to the disruption of supercomplexes. The most frequently reported causes of CIII deficiency are variants in *MT-CYB*, with more than 50 patients reported that typically presented myopathy and exercise intolerance. Next, more than 30 cases with *BCSIL* variants have been described with either Björnstad syndrome or GRACILE syndrome, severe neurologic and multi-systemic diseases with neonatal onset (Fernandez-Vizarra and Zeviani, 2015; Feichtinger et al., 2017). Individuals carrying *UQCRB*, *UQCRC2*, and *CYCI* pathogenic variants have similar phenotypes, composed of neonatal or early infantile onset, recurrent metabolic crises with elevated lactate levels, and hypoglycemia, treatable with intravenous glucose. All but one patient presented with normal development and intellect in later life (Feichtinger et al., 2017).

Table 2. Clinical features of reported patients with complex III deficiencies.

Complex III deficiencies										
Gene	<i>MT-CYB</i>	<i>CYC</i>	<i>UQCRB</i>	<i>UQCRC2</i>	<i>UQCRC1</i>	<i>UQCRC2</i>	<i>UQCRC3</i>	<i>TTC19</i>	<i>LYRM7</i>	<i>BCSIL</i>
Number of patients	>50	1	1	4	25, 1 kindered	2	1	~15	9	>30
Onset	Childhood, adulthood	Infancy, early childhood	Late infancy	Neonatal	First months of life	Intrauterine	Birth	Late infancy, adulthood	Infancy, 14 years	First years, infancy
Intrauterine growth retardation						Yes				
Hearing impairment						Yes/n.a.	No			Yes
Hypotonia	Yes					Yes	Yes		Yes	Yes
Seizures						Yes	No			Yes
Abnormal EEG						Yes				Yes
Metabolic crisis			Yes				Yes		Yes	
Lactic acidosis	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes
Increased CSF lactate	Yes					Yes				
Hypoglycaemia		Yes	Yes	Yes			Yes			
Developmental disability		No	No	No		No/n.a.	Yes	Yes	Yes	Yes
Intellectual disability		No	No	Yes (1)/no (1)	Yes	n.a.	Yes		Yes	Yes
Other features		Hyperammonemic liver failure in one case			Extrapyr amidal movement disorder, survival into thirties	Renal tubular acidosis; status epilepticus	Muscular weakness	Later regression with spasticity and movement disorder	Deterioration after metabolic crises, specific MRI pattern	Hepatopathy, renal involvement, often early death

1.6. RNA-sequencing

DNA sequencing has become widely adopted and accepted for the molecular diagnosis of genetic diseases (Smith et al., 2019). However, a significant proportion of patients are left without a diagnosis following genetic investigation. As mentioned previously, 85% of the annotated pathogenic variants reside in the exonic regions. This makes WES a cost-effective approach, but also indicates that a significant number of pathogenic variants resides in the non-coding regions. They can be detected by WGS, but the daunting number of non-coding variants and their difficult interpretation makes recognition of their contribution to disease challenging and neglected. Following the central dogma of molecular biology, RNA serves as a highly dynamic, transient intermediate between DNA and protein. Profiling the complete set of transcripts in a cell- transcriptome is therefore useful for interpreting the variation in the genome and also for understanding cellular physiology, development, and disease. Keeping in mind that pathogenic non-coding variants more likely disrupt gene expression, RNA-based measurements have the potential for clinical application (Byron et al., 2016). Among RNA-based measurements, RNA-sequencing (RNA-seq) stands out as it offers an in-depth, unbiased, systematic overview of the transcriptome and detection of transcript variation. Unlike other RNA-based techniques, RNA-seq is an open platform technology, allowing detection and quantification of both known/pre-defined, but also rare and novel transcripts within a sample (Mortazavi et al., 2008).

1.6.1. Principles and advances

RNA-seq combines high-throughput NGS methodology with computational methods to capture and quantify transcripts (Ozsolak and Milos, 2011). It leverages deep sampling with many short fragments to allow computational reconstruction of the original RNA transcript by aligning reads to a reference genome (Wang et al., 2009; Conesa et al., 2016). Its average dynamic range is of five orders of magnitude (Lowe et al., 2017), and with a very low background signal for 100 bp, theoretically, there is no upper limit of quantification (Ozsolak and Milos, 2011).

As is true for WES, the most frequently used RNA-seq platform is provided by Illumina (Goodwin et al 2016). The commonly employed Illumina TruSeq RNA Library preparation protocol is optimized for input of 0.1 to 1 μg of total high-quality RNA derived from various human tissues. Overall, although the same by principle, RNA-seq methods differ in the use of transcript enrichment, fragmentation, amplification, single or paired-end sequencing, and whether they preserve strand information. The sensitivity of an RNA-seq experiment can be increased by enriching classes of RNA that are of interest and depleting known abundant RNAs. As rRNA is the most highly abundant RNA class in animal and human samples (>80% to 90% of total RNA), it is usually removed by the selection of polyadenylated (poly(A)) RNA transcripts using the oligo-dT beads, or rRNAs depletion through hybridization capture followed by magnetic bead separation (Zhao et al., 2018). The oligo-dT approach hinders detection of transcripts lacking a poly(A) tail, such as ribosomal RNAs (rRNAs), small RNAs generated by the RNA polymerase III, replication-dependent

histone mRNAs and some long non-coding RNAs (lncRNAs); as well as transcripts shorter than 100 nt such as the microRNAs (miRNAs) (Yang et al., 2011). Additional protocols have been established for the investigation of such classes of RNAs. After, RNA is fragmented and converted into cDNA by the reverse transcription reaction. This is a key aspect of library construction (Knierim et al., 2011). It can include chemical hydrolysis, nebulization, or sonication, or implement simultaneous fragmentation, and tagging of cDNA by transposase enzymes (Lowe et al., 2017). Strand-specificity can be achieved by marking one strand by chemical modification, either on the RNA by bisulfite treatment or dUTP incorporation during second-strand cDNA synthesis followed by degradation of the unmarked strand (Levin et al., 2010). The subsequent platforms are the same as for the genomic data.

1.6.2. RNA-seq as a complementary tool in molecular diagnostics

In a usual WES diagnostic setting, the main focus is on protein-altering variants. However, it leaves at least half of the investigated cases unsolved and also underestimates the complexity of genome regulation. Non-coding variants could disrupt transcription factors binding, local chromatin structure, and/or co-factors recruitment, ultimately changing the expression of the target genes (Ma et al., 2015). Comparing variants against regulatory elements suggested that up to 30% of disease-causing variants impact RNA and fall within the non-coding regions (Ma et al., 2015; Stenson et al., 2017). However, such variant effect is often difficult to predict and association with a particular disease usually demands functional validation, starting from the transcript level. Besides, knowing the primary effect of a variant is not only crucial for the understanding the pathogenic mechanism of a disease, but also the basic genetic mechanisms. For example, it makes a difference if a variant results in loss of gene expression due to aberrant splicing and triggered nonsense-mediated decay (NMD), rather than normal expression of a protein carrying a missense variant. This is particularly important as it raises the possibility of individually tailored splicing-targeted therapies (Havens et al., 2013). Starting in 2017 with the pioneer works of Kremer et al. and Cummings et al., RNA-seq has emerged as a complementary tool to WES, providing functional data for variant interpretation and discovery of aberrant transcript events. These and proceeding studies have obtained increased diagnostic rates over WES/WGS for a variety of Mendelian disorders, ranging between 8% and 36% (Kremer et al., 2017; Cummings et al., 2017; Fresard et al., 2019; Gonorazky et al., 2019; summarized in Table 3). They focused on the detection of transcript outliers exhibiting at least one of the three types of aberrant events that could lead to disease: aberrant expression, aberrant splicing and mono-allelic expression (MAE). In addition, Lee et al. (2019b) used RNA-seq to validate rare variants from WGS, and Wai et al. (2020) focused only on the VUSs predicted to affect splicing (Table 3).

Table 3. Overview of published RNA-seq-based diagnostic studies

Reference	Disease cohort	Tissue of investigation	Number of samples	Diagnostic yield
Kremer et al. (2017)	Mitochondrial disease	Fibroblasts	105	10%
Cummings et al. (2017)	Muscle disorders	Muscle	63	35%
Gonorazky et al. (2019)	Neuromuscular disorders	Muscle, fibroblasts, t-myotubes	70	36%
Fresard et al. (2019)	Rare Mendelian diseases	Blood	143	8%
Lee et al. (2019b)	Rare Mendelian diseases	Whole blood, fibroblast, muscle, bone marrow	113	9%
Wai et al. (2020)	Patients with VUSs	Blood	17	29%

1.6.2.1. Aberrant expression

Individual gene expression is a result of both genetic and non-genetic factors, as well as their combined action. Variants termed expression quantitative trait loci (eQTLs) influence the expression level of genes and can play an essential role in human diseases and other complex phenotypes (Albert and Kruglyak, 2015). Studies of eQTLs have contributed to a better understanding of gene expression regulation. However, these studies are designed for assessing the effects of common variants and are based on population-level test statistics (Zeng et al., 2015). What is more, recent studies suggest that rare SNPs contribute to the large effects on gene expression much more so than the common ones (Montgomery et al., 2011; Zeng et al., 2015; Zhao et al., 2016). Focusing on these large effects, aberrantly expressed genes, or expression outliers, are defined as genes whose expression level is aberrantly higher or lower in a sample compared to the other samples within a cohort. Identification of expression outliers is based on using a stringent cut-off on expression variation, usually based on the Z-score (Li et al., 2017; Montgomery et al., 2011) or statistics at the level of the whole gene set (Zeng et al., 2015; Guan et al., 2016). Elucidating the genetic cause of observed aberrant expression can help single out rare variants with phenotypic effects, especially variants in the often overlooked promoter, enhancer and intronic regions (Li et al., 2017; Zhao et al., 2016). For example, Li et al. (2017) reported that 58% of underexpression and 28% of overexpression outliers from the GTEx dataset (GTEx Consortium, 2017) contained nearby conserved rare variants. These outliers were enriched for SNVs, near splice site variants, frameshifts, start or stop codons, variants near transcription start sites (TSSs), and in conserved regions. Calling expression outliers has been successfully used to identify and/or confirm disease-causing genes variants in rare disorders patients (Kremer et al., 2017; Cummings et al., 2017; Fresard et al., 2019; Gonorazky et al., 2019). In a rare-disease diagnostic setting, the cohort consists of dozens of samples, usually without replicates, which are inspected in search for a clear expression outlier. This approach is vastly different from the usual treatment versus control experimental design with an abundance of replicates, where the interests lay in detecting a subtle change between two populations and for which differential-expression algorithms, such as DESeq (Love et al., 2014) and EdgeR (Robinson et al., 2010) have been developed. Recently, the OTRIDER algorithm (Brechtmann et al., 2018) has been developed for rare disease diagnostic platforms. It combines an autoencoder, which controls for covariations among the gene read counts, and includes a statistical test for

outlier detection. Cummings et al. (2017) calculated Z-scores on the log-transformed gene-length-normalized read counts by subtracting the mean count and dividing by the standard deviation, calling a gene aberrantly expressed if its absolute Z-score is greater than 3. Kremer et al. (2017) also used the threshold of absolute Z-score greater than 3, but on read-counts, and statistical significance according to DESeq2. Gonorazky et al. (2019) used EdgeR and OUTRIDER to identify outliers whose expression changed by 2-fold when compared to their mean expression in the GTEx control cohort. Fresard et al. (2019) searched for outlier genes by comparing individuals with the disease against controls from the GTEx cohort, after correcting the data for hidden confounders by SVA and using absolute Z-score greater than 2.

1.6.2.2. Aberrant splicing

Alternative splicing is a ubiquitous mechanism of posttranscriptional gene expression regulation that allows the production of multiple distinct transcripts from a single gene. It is considered to have a key role in the functional complexity of higher eukaryotes (Chen et al., 2012). Up to 95% of human genes undergo some level of alternative splicing (Pan et al., 2008; Wang et al., 2008), and human genes with multiple exons produce on average at least three distinct RNA isoforms (Lee and Rio, 2015). In addition, mass spectrometry analyses revealed that ~37% of protein-coding genes generate multiple protein isoforms (Kim et al., 2014). Generated transcripts isoforms can differ in their untranslated regions (UTRs) or coding regions through four basic splicing events: exon skipping, intron retention and use of alternative 5' or 3' splicing sites. These differences might ultimately affect transcript stability, localization, and/or translation (Baralle and Giudice, 2017).

The key player in splicing is the spliceosome, a complex of small nuclear RNAs (snRNAs) and tens of associated proteins (Matera and Wang, 2014). The spliceosome recognizes conserved sequences within the pre-mRNA: the 5'-splice site (CAG/GUAAGU sequence, splice donor), the branch point sequence, the 3'-splice site (NYAG/G sequence; splice acceptor), and the polypyrimidine tract (Cartegni et al., 2002). Almost always, the 5' and 3' ends of introns include GU and AG sequences, respectively. Apart from these core consensus sequences, additional *cis* elements- pre-mRNA sequences within a gene (such as exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE), and intronic splicing silencer (ISS) elements), and *trans* factors-proteins and ribonucleoproteins, are also required for splicing regulation (Blencowe, 2006). Alterations in splicing have been described to cause Mendelian diseases, modify the disease phenotype, or be linked with disease susceptibility (Wang and Cooper, 2007; Scotti and Swanson, 2016). The mechanisms behind these defective splicing events include disruptions of *cis* elements, which directly affect the expression of the gene containing the disrupted sequence, or *trans* factors, which could potentially affect the expression of multiple genes by interfering with the spliceosome recruitment (Singh and Cooper, 2012). As an example, ~15% of cases of retinitis pigmentosa, the most frequent inherited form of retinal blindness, are autosomal-dominant forms caused by pathogenic variants in the pre-mRNA processing factors (Buskin et al., 2018). Focusing on the *cis* elements, variants can disrupt canonical splice sites or enhancers and silencers that can affect splice site usage. However, the prediction of splicing defects

from the sequence is difficult as *cis*-regulatory elements are not yet fully understood. Currently, 9% of all variants reported in HGMD affect the splicing pattern of genes. Some authors suggested that approximately a third of all disease-causing variants alter splicing, including 22% of variants originally classified as missense (Lim et al., 2011). A computational model by Xiong et al. (2015) predicted that among deep intronic variants, known disease variants alter splicing nine times more often than common ones, and among missense exonic disease variants, those that minimally impact protein function are over five times more likely to disrupt splicing. Using MaPSy, a dual parallel splicing system, Soemedi et al. (2017) reported that ~10% of exonic disease alleles disrupt splicing. In addition, around half of the synonymous positions in codons of conserved alternatively spliced mRNAs are under selection pressure, suggesting that conserved alternative exons and flanking introns are enriched in the splicing regulatory elements (Blencowe, 2006). In this respect, up to 25% of synonymous substitutions have been estimated to disrupt normal splicing using the CFTR exon 12 as a splicing model (Pagani et al., 2005). This collectively suggests that defective splicing is a more frequent pathomechanism than assumed and that more variants, even synonymous and deep intronic, should be examined for pathogenicity. There are several tools developed to predict the splicing effect, such as NNSplice, Splice Site Finder, MaxEntScan (MES), Human Splicing Finder (HSF), SpliceAI (Shapiro et al., 1987; Reese et al., 1997; Yeo et al., 2004; Desmet et al., 2009; Jaganathan et al., 2019). However, in a diagnostic setting their abundant predictions still necessitate functional validation. For example, Wai et al. (2020) assessed the effect of 257 variants predicted to affect splicing via RT-PCR and RNA-seq, functionally associating 33% of them with aberrant splicing, of which 13% resided outside of the splice region. It is important to know that the splice defect does not necessarily cause loss of function. If the resulting aberrant splicing event is in-frame, the shorter protein will be produced that may still be functional. In the case of an out-of-frame splicing events, a premature stop codon (PTC) will usually be introduced. The presence of PTC can lead to the synthesis of a shorter protein, but more typically it will initiate mRNA degradation through NMD (McGlinchy and Smith, 2008). Consequences of splicing-altering variants can be roughly divided into five categories and have been described as disease-causative across literature (Kremer et al., 2018; Anna and Monika, 2018; Figure 13):

- 1) exon skipping and inclusion of novel cryptic exons caused by deep intronic variants creating new splice sites (as in *TIMMDC1* in Kremer et al. (2017))
- 2) exon skipping caused by variants within the canonical splice sites or exonic variants usually leading to ESE disruption (as in *DMD* in Habara et al. (2009))
- 3) exon truncation caused by variants within the canonical splice sites or exonic variants creating new splice sites (as in *GLA* in Chang et al. (2017))
- 4) exon extension caused by variants in the canonical splice sites (as in *ATP7A* in Skjorringe et al. (2011))
- 5) intron retention caused by variants in the canonical splice sites (as in *SERAC1* in Wortmann et al. (2012))

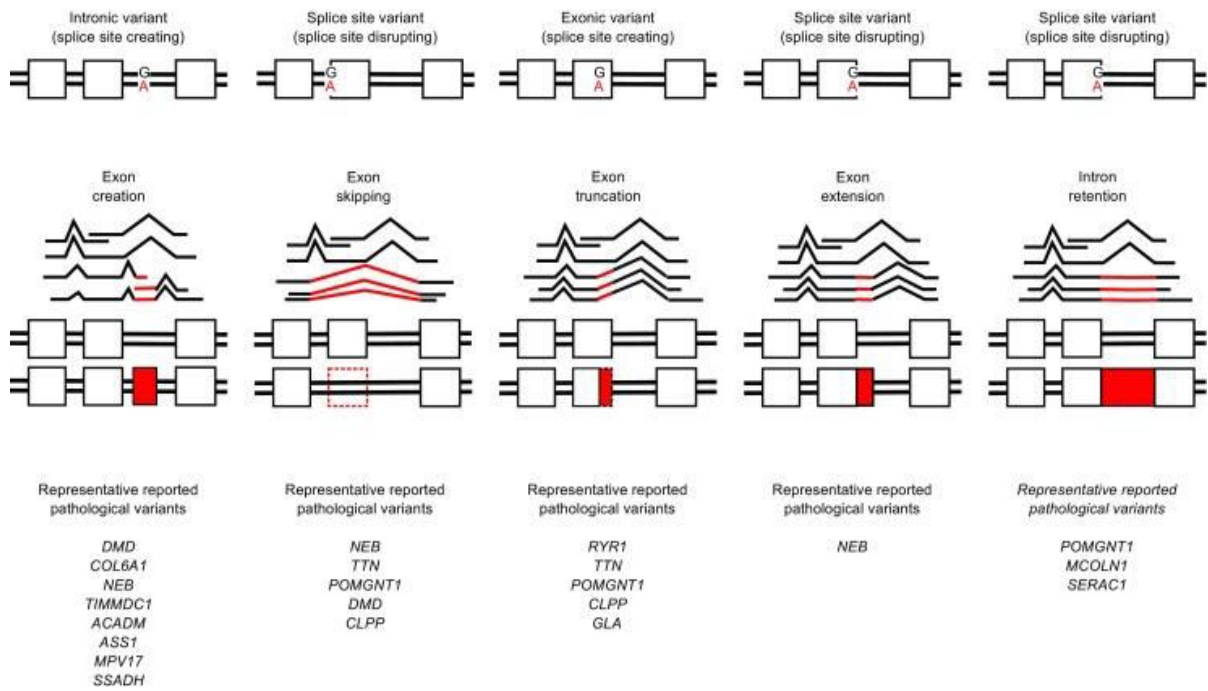


Figure 13. Aberrant splicing as a cause of disease. Splicing aberrations caused by coding and non-coding variants with published examples for each category. Figure taken from Kremer et al. (2018).

Different approaches have been developed to call aberrant splicing. Kremer et al. (2017) used an adjustment of the differential splicing test LeafCutter with an adjusted P value < 0.05 (Li et al., 2018), which forms intron clusters and tests for differential usage between one sample and all others. However, LeafCutter is restricted to the split reads, so it is difficult to detect intron retention events. Cummings et al. (2017) and Gonorazky et al. (2019) used a cut-off based approach, where splice junctions in genes were filtered in terms of the number of samples containing a splice junction and the number of reads, and the normalized value supporting that junction. However, their cut-offs are arbitrary and may fail to recognize aberrant events in weak splice sites (Kapustin et al., 2011). Neither of these two approaches control for sample covariation. Fresard et al. (2019) used a method based on Z-score, which does not control for false discovery rate and can be inaccurate when dealing with low read counts. To address these limitations, the FRASER algorithm has been developed. It uses a denoising autoencoder to control for latent confounders and detects aberrant splicing according to a count-based statistical test. Unlike other methods, it can also capture intron retention by considering non-split reads that overlap the splice sites (Mertes et al., 2021).

1.6.2.3. Mono-allelic expression

Mono-allelic expression (MAE) is a phenomenon where in the diploid organisms only one allele is expressed. The other allele is transcriptionally silenced or post-transcriptionally degraded based on genetic or epigenetic grounds. Genetic grounds can include heterozygous regulatory, splicing or nonsense variants, or deletions, which usually provoke transcript degradation via NMD (Lykke-Andersen and Jensen, 2015). Epigenetic grounds are

implicated in both constitutive and random MAE. A constitutive MAE expression occurs for the same allele throughout the whole organism or tissue as a result of genomic imprinting-expression based on parent-of-origin effects (reviewed in Ferguson-Smith, 2011). In a random or stochastic mono-allelic expression (RMA, StMA), the expressed allele is randomly selected early in development and this choice is kept in the subsequent cellular progeny (Gimelbrant et al., 2007; Reinius and Sandberg, 2015). By this, RMA contributes to the heterogeneity among cells and probably to the variance of phenotypes among individuals of identical genotype (Chess, 2012; Reinius and Sandberg, 2015). The best-studied example of RMA is the X-chromosome inactivation (reviewed in Wutz, 2011).

When assuming a recessive mode of inheritance during WES/WGS analyses, genes with a single heterozygous rare coding variant are not prioritized. However, MAE of such variants mimics the impact of a homozygous variant on the RNA level, thereby fitting the recessive mode of inheritance assumption and helping to diagnose the case (Kremer et al., 2017). Detection of MAE can shed light on the variants that result in allele silencing and that are either not detected by WES or not prioritized. For example, in a pilot study Albers et al. (2012) used RNA-seq to unravel the compound heterozygous inheritance mechanism involving *de novo* deletion in combination with the regulatory 5'UTR or intronic SNP as a cause of thrombocytopenia with absent radii syndrome.

Several methods have been developed to detect MAE in the rare disease context. Kremer et al. (2017) applied the DESeq2 package (Love et al., 2014) and used a negative binomial test with a fixed dispersion parameter for all genes. They filtered for heterozygous rare SNVs with an RNA-seq coverage > 10 reads and called MAE when more than 80% of the reads contained the variant in association with a Hochberg adjusted P-value of < 0.05 . The approach of Cummings et al. (2017) involved defining the 95% confidence interval of mean allele balance in GTEx individuals for each gene and identifying patients for whom the gene-level allele balance fell outside of the range based on a t-test for significance. Gonorazky et al. (2019) filtered variants in heterozygous sites with a coverage depth ≥ 20 , calculated minor-allele read-count ratios and a median of these ratios for genes having ≥ 5 variants. This value was used to characterize the MAE by computing a Z-score and comparing sample data with the GTEx controls. Fresard et al. (2019) ran the ASEReadCounter (Castel et al., 2015) and obtained Z-scores per site, with a focus on outlier MAE sites within the individuals with the disease when compared with all other rare-disease individuals and GTEx samples. Finally, recently developed, ANEVA-DOT (Mohammadi et al., 2019) implements a binomial-logit-normal test with gene-specific variance-VG.

1.6.2.4. Tissue-specific gene expression and splicing

As a gene of interest must be expressed in the chosen tissue of investigation, tissue-specific expression is the most frequently discussed limitation of the RNA-seq. A further complication is that splicing can also differ between different tissues. As a consequence, aberrant splicing in the affected tissue can remain elusive even if the gene is well expressed in a tissue of investigation. Therefore, the chosen tissue of investigation should be an

adequate proxy for both gene expression and splicing. In praxis, this is not so easy to achieve as clinicians and researchers can perform RNA-seq only on the tissue they can have access to from patients, which is usually limited to blood (whole blood, Epstein–Barr virus (EBV)-transformed lymphocytes), and skin biopsies (fibroblasts). These three are referred to as clinically accessible tissues (CATs) (Aicher et al., 2020). In parallel, interest often lies in the affected, but often inaccessible, tissues (non-CATs, e.g., muscle, brain, heart, liver, etc.). Fresard et al. (2019) demonstrated that RNA-seq on whole blood can be used for diverse disease categories. On the contrary, Cummings et al. (2017) favored performing RNA-seq on skeletal muscle, the disease affected tissue, in a cohort of neuromuscular disease patients. Gonorazky et al. (2019) recommended using muscle, fibroblasts and transdifferentiated t-myotubes, while considering blood as an inadequate representative of the neuromuscular disease. Furthermore, they identified aberrant splicing events in a muscle that were absent in fibroblasts from the corresponding patients. Working on published datasets from three CAT and 53 non-CATs, Aicher et al. (2020) found that splicing of 40% of genes per non-CAT is inadequately represented by at least one CAT, with fibroblasts being the best performing CAT. To assess splicing variation, the authors developed the MAJIQ-CAT online resource that shows how well CATs represent splicing based on the choice of gene and the non-CATs.

1.7. Treatment of mitochondrial diseases

Although advances in molecular and biochemical methodologies have largely improved our understanding of the etiology and mechanism behind mitochondrial diseases, these advances have not been matched in terms of treatment options. Unfortunately, there is still no cure for mitochondrial diseases as a whole. Available treatment for the vast majority remains mainly symptomatic and does not significantly alter the course of the disease. There is a limited base of evidence and little data from randomized clinical trials. Nevertheless, scientific advances have stimulated interest in developing new treatments and prevention methods (Russell et al., 2020). This gives hope for the future.

1.7.1. Symptomatic treatments

Application of symptomatic treatments, although not curative, can still improve quality of life and increase life expectancy. Examples include physical therapy for hypotonia and motor delay, hearing aids or cochlear implants for hearing loss, slow infusion of sodium bicarbonate during acute exacerbation of lactic acidosis, anticonvulsants for epilepsy, cardiac pacing in arrhythmias, surgical correction of ptosis, administration of pancreatic enzymes for exocrine pancreatic dysfunction, and treating diabetes with diet, sulfonylurea, and insulin (El-Hattab et al., 2017).

In some cases, especially in individuals with mtDNA mutations, exercise and endurance training can be helpful as they can improve the OXPHOS activity, stimulate mitochondrial biogenesis, mitophagy and the incorporation of satellite cells into existing muscle fibers, leading to a lowering of heteroplasmy levels (Jeppesen et al., 2006). No specific dietary manipulation has shown to have consistent benefits. Still, a high-lipid, low-carbohydrate diet such as a ketogenic diet is often suggested as a high-carbohydrate diet can

be metabolically challenging in individuals with an impaired OXPHOS (Munnich et al., 2009). As an example, patients with PDH deficiency had increased longevity and improved mental development when placed on a ketogenic diet (Wexler et al., 1997). Finally, liver transplantation has been performed for individuals with hepatocerebral mtDNA depletion syndromes which frequently progress to liver failure. However, this approach remains controversial, largely because of the multi-organ involvement (Parikh et al., 2016).

1.7.2. Strategies for the development of specific therapies

1.7.2.1. Small molecule therapies

In the last two decades, there have been several approaches targeting different mitochondrial processes in the development of new drugs. These small molecules have acquired an increasing interest as modulators of mitochondrial function can have a therapeutic role in a variety of more common diseases (Parkinson's disease, Alzheimer's disease, etc.), putting their usage in the wider context (Russell et al., 2020). Some studies focus on enhancing OXPHOS function by either augmenting its components (CoQ₁₀, idebenone, riboflavin) or increasing its substrate availability (dichloroacetate, thiamine) (reviewed in El-Hattab et al., 2017). As examples, CoQ₁₀ supplementation resulted in improvement in the clinical manifestations associated with CoQ₁₀ deficiency (Horvath, 2012). Idebenone, an analog of CoQ₁₀, has been licensed to treat LHON patients (Klopstock et al., 2013). Riboflavin (vitamin B₂) is a flavoprotein precursor that has proven useful in the treatments of multiple acyl-CoA dehydrogenase deficiencies and *ACAD9* patients (Olsen et al., 2007; Repp et al., 2018). Another approach is based on creatine, as it serves as an energy buffer for the movement of high energy phosphates from mitochondria to the cytoplasm, where they are used. Creatine monohydrate supplementation improved exercise capacity in mitochondrial myopathy cases (Tarnopolsky, 2011). Antioxidants, such as vitamin C, vitamin E, lipoic acid, cysteine donors, and EPI-743 can alleviate the toxic effect of excessive ROS produced during mitochondrial dysfunction (reviewed in El-Hattab et al., 2017). As nitric oxide (NO) deficiency plays a major role in mitochondrial disease pathogenesis, supplementation with its precursors arginine and citrulline could have therapeutic utility (El-Hattab et al., 2017). Elamipretide binds selectively to cardiolipin and protects it from oxidation, which preserves the mitochondrial cristae structure and promotes the OXPHOS. It has been promisingly tested in various conditions related to the mitochondrial damage (Szeto, 2014). Activators of mitochondrial biogenesis, such as bezafibrate and resveratrol, have been tested but with limited evidence (Baur et al., 2006; Dillon et al., 2012). Nicotinamide riboside (NR), nicotinamide mononucleotide (NMN), acipimox and ACMSD inhibitors have been successfully shown to increase the cellular load of NAD⁺ *in vitro* and in mice (Khan et al., 2014; van de Weijer et al., 2015; Katsyuba et al., 2018; Lee et al., 2019a). Rapamycin, an mTORC1 inhibitor, partially rescues mitochondrial function by promoting the removal of dysfunctional mitochondria and has been used in clinical practice for years (Khan et al., 2017). Urolithin A, a metabolite that stimulates mitophagy, has been shown to prevent the accumulation of dysfunctional mitochondria and improve muscle health in old mice and elderly individuals, inducing a molecular signature that resembles one during a regular

exercise regimen (Andreux et al., 2019). Nucleoside bypass therapy via deoxynucleoside monophosphates and deoxynucleoside treatments has been proposed for the mtDNA maintenance disorders and has shown encouraging results in MNGIE patients with *TK2* mutations (Dominguez-Gonzalez et al., 2019). Finally, Pirinen et al. (2020) recently reported successful treatment of adult patients with mitochondrial myopathy with niacin that restored disturbed NAD⁺ levels, increased mitochondrial biogenesis, and muscle strength.

1.7.2.2. Manipulation of mtDNA

Recent approaches for treatments of mtDNA-caused diseases are based on shifting the heteroplasmic ratio. These therapies rely on directed, sequence-specific nucleolysis of mtDNA using mitochondrially targeted zinc finger-nucleases (mtZFNs) and mitoTALENs that are delivered into cells by the adeno-associated virus (AAV). As mammalian mitochondria lack efficient DNA double-strand break (DSB) repair pathways (Dahal et al., 2018), the selective introduction of the DSBs into the mutant mtDNA via these sequence-specific nucleases leads to rapid degradation of such DNA and stimulates replication of the remaining mtDNA pool, eliciting shifts in the heteroplasmic ratio towards the wild-type mtDNA (Bacman et al., 2013; Gammage et al., 2016). Using a mouse model with a heteroplasmic mtDNA mutation (Kauppila et al., 2016), these approaches have been shown to be effective and safe, and have led to a robust and stable reduction of the mutant mtDNA, and a rescue of the molecular defect (Bacman et al., 2018; Gammage et al., 2018). In addition, remarkable efforts have been made in the field of gene therapy and CRISPR gene editing in a range of genetic disorders. Termed allotopic expression, a current “hot” approach is based on the delivery of the wild-type gene via a modified virus into the cell. It has shown promising results in treating LHON patients with m.11778G>A variant. The principle is that the AAV carries wild-type *MT-ND4* and that the viral capsid VP2 is fused with an MTS to target the AAV to mitochondria and achieve the *MT-ND4* expression. Initial results of unilaterally injected AAV vector into the eyes showed an improvement in 12 out of 14 patients (Guy et al., 2017).

1.7.2.3. Prevention of transmission

Ultimately, as mitochondrial diseases are often relentlessly progressive and incurable, affected families understandably seek options to prevent their transmission. As they vary vastly depending on the source of the defect (nDNA or mtDNA), it is clear why the determination of genetic diagnosis is so important. For families with nuclear defects, genetic counseling, prenatal testing, and preimplantation genetic diagnosis are helpful options (Gorman et al., 2018). However, in mtDNA-encoded disease, due to the involvement of genetic bottleneck and heteroplasmy, provision of such advice and diagnostic options are also challenges. Moreover, prenatal diagnosis based on samples taken by amniocentesis or chorion villus biopsy may not predict the degree of heteroplasmy in all tissues, or disease severity (Steffann et al., 2007), making their interpretation difficult. To reduce the chance of having an affected child, different strategies aim to prevent or reduce the transmission of pathogenic mtDNA. An egg donation avoids transmission, but also results in a child not

genetically related to the woman undergoing treatment. The development of mitochondrial replacement (or mitochondrial donation) therapy (MRT) offers affected women the possibility of having a genetically related child with a greatly reduced risk of transmitting disease. It involves transplantation of the nuclear genome from the egg of an affected woman to an enucleated egg from an unaffected donor. This can be performed by maternal spindle transfer, pronuclear transfer or polar body transfer (reviewed in Greenfield et al., 2017). However, this approach is still followed by a number of controversies and ethical issues. As it results in a child with mtDNA from a second woman, the child is often controversially termed the “three-parent baby”.

2. Aim

Unprecedented clinical heterogeneity together with a lack of curative treatments for mitochondrial diseases emphasizes the relevance of establishing the genetic diagnosis. Currently, only a limited number of treatments for specific genetic defects are available. Trials for the development of new therapies are increasingly requiring patient stratification based on their genetic diagnosis (Warr et al., 2015). Finally, living without knowing the genetic cause of disease represents a major burden to patients and their families, regarding both care and future prevention of transmission. In this respect, the genetic diagnosis presents a fundamental basis for both genetic counseling and prenatal testing. The most widely used open database curating genes associated with rare disorders, OMIM (Online Mendelian Inheritance in Man; <http://www.ncbi.nlm.nih.gov/omim>), has catalogued 5889 genes associated with a particular phenotype (disease) (accessed in September 2020). This number will not remain stagnant for long: ~300 new disease-associated genes are accumulating each year. In addition, gene mutations are currently reported at an annual rate of 10,000 (Cooper et al., 2011). Although seemingly formidable, these numbers are probably an underestimation, given lack of resources, infrastructure and expertise still limiting discoveries worldwide (Baynam et al., 2015). WES had a revolutionary impact on reaching the genetic diagnosis. However, a significant portion of patients remain undiagnosed following WES. Indeed, the diagnostic rate of WES in mitochondrial disease patients is around 50%, indicating technical limitations, but also challenges in gene and variant prioritization and annotation. RNA-seq implementation in a diagnostic setting enabled solving 10% of the WES-inconclusive cases (Kremer et al., 2017), indicating that a shift away from just DNA analysis can be beneficial. To expand the genetic spectrum of mitochondrial disorders and increase the diagnostic rate on one hand, and to assess the benefits and pitfalls of WES and RNA-seq in the diagnostic setting on the other, the starting point for my thesis was the analysis of cases whose diagnostic WES was inconclusive. These cases initially included 158 WES-negative cases within mitoNET- German network for mitochondrial diseases, and later 100 WES-negative cases with a cardiac phenotype from the DZHK- German Centre for Cardiovascular Research. The reanalysis of these WES data led to the development of two different approaches across three projects:

- 1) WES reanalysis identified a candidate gene: WES revealed biallelic variants in a gene encoding a mitochondrial protein *MDH2* in a patient with suspected mitochondrial disease. As this gene had not previously been associated with a Mendelian disorder, it was unclear whether these variants were pathogenic. To investigate the causality of discovered variants, I performed different functional assays to shed the light on the pathomechanism.
- 2) WES reanalysis identified a candidate gene: WES revealed biallelic variants in a gene encoding a mitochondrial protein *UQCRC1* in a patient with a cardiac phenotype. Again, functional assays were performed to associate the discovered variants with the disease.

- 3) Soon after the initial “WES reanalysis” and “candidate gene-centric” approach, I turned to RNA analysis as a new and unbiased approach to diagnostics. Following the work of Kremer et al. (2017), I continued to perform RNA-seq as a complementary tool to all WES-inconclusive cases referred to our institute with the hope to find a genetic cause. In collaboration with bioinformaticians from the Technical University of Munich, we have further expanded the diagnostic implementation of RNA-seq. Upon confirmation of skin derived-fibroblasts as a good investigation model, I continued to use these as an experimental model and took part in expanding our RNA-seq cohort. This resulted in a large RNA-seq compendium of rare disease. For analysis of RNA-seq data, OUTRIDER and FRASER, algorithms that search for aberrant expression and splicing events, respectively, were implemented. The RNA-seq analysis steps were integrated into a computational protocol that reduced the time from the acquisition of raw files to results to just several hours (Yepez et al., 2021). We also took advantage of sequence information, and implemented variant calling on RNA-seq data to expand the sequence spectrum and cover regions missed by WES. Upon such improvements, I analysed the RNA-seq results in a diagnostic setting.

These three projects had the overall aim to improve the genetic diagnosis of mitochondrial diseases via the discovery of novel disease-causing genes and the implementation of complementary diagnostic tools such as RNA-seq to WES.

3. Material and methods

3.1. Material

3.1.1. Nucleic acids

3.2.3.1. DNA

Genomic DNA of controls and patients with suspected mitochondrial diseases belong to DNA collection of the Institute of Neurogenomics/Institute of Human Genetics (Helmholtz Zentrum München, Neuherberg, Germany/ Klinikum Rechts der Isar, Munich, Germany). For all samples informed consent was obtained for diagnostic and research purposes.

3.2.3.2. cDNA

UQCRFS1 cDNA plasmid (catalog number: HsCD00043924) was purchased from DNASU Plasmid Repository (Tempe, AZ, USA).

3.2.3.3. Oligonucleotides

The oligonucleotides enlisted in Table 4 were synthesized by Metabion (Planegg, Germany).

Table 4. List of used primers.

Name	Sequence (5'→3')	Purpose
<i>UQCRFS1</i> cDNA forward	CACGATGTTGTCGGTAGCAGCC	Lentiviral transduction
<i>UQCRFS1</i> cDNA reverse	CTAACCAACAATCACCATATCGTC	Lentiviral transduction
<i>UQCRFS1</i> splice check forward	AGCCTGTGTTGGACCTGAAGC	RT-PCR
<i>UQCRFS1</i> splice check reverse	ACTGGGTGACGGCATTCTTG	RT-PCR
<i>UQCRFS1</i> variant sequence forward	CCTAGCCTCGCTGCTTTCTC	Segregation analysis
<i>UQCRFS1</i> variant sequence reverse	TCCATAAAGGGCTATTTCCTG	Segregation analysis
CMV	CGCAAATGGGCGGTAGGCGTG	Check of lentiviral construct
V5	ACCGAGGAGAGGGTTAGGGAT	Check of lentiviral construct

3.1.2. Cell lines

Following cell lines were used for the experiments:

- NHDFneo (NHDF): normal human dermal fibroblasts derived from neonatal foreskin used as a control cell line – referred to as NHDF or control. Commercially available and purchased from Lonza (Basel, Switzerland).
- HEK 293 FT: clonal isolate derived from human embryonal kidney cells stably expressing the SV40 large T antigen from the pCMVSPORT6TAg.neo plasmid to facilitate lentivirus production. Expression of the SV40 large T antigen is under control of human cytomegalovirus (CMV) promoter and is high-level and constitutive. Commercially available and purchased from Thermo Fisher Scientific (Waltham, MA, USA).

- Patient cell lines: primary fibroblasts derived from a skin biopsy. In many cases the culture was started by colleagues at the Institute of Neurogenomics/Human Genetics.
- One Shot TOP10 Chemically Competent *E. coli*: designed for high-efficiency cloning and plasmid propagation. These cells allow stable replication of high-copy number plasmids. Commercially available and purchased from Thermo Fisher Scientific (Waltham, MA, USA).
- One Shot Stbl3 Chemically Competent *E. coli*: designed for cloning direct repeats found in lentiviral expression vectors. These cells reduce the frequency of homologous recombination of long terminal repeats (LTRs) found in ViraPower™ Lentiviral Expression Vectors. Commercially available and purchased from Thermo Fisher Scientific (Waltham, MA, USA).

3.1.3. Antibodies and the protein ladder

The primary antibodies were used as enlisted in Table 5.

Table 5. List of used antibodies.

Antibody	Producer	Catalog number	Purpose	Dilution	Buffer
MDH2	Cell Signaling	8610S	Western blot	1:1000	5% non-fat dry milk in 1X TBS-T
UQCRFS1	Abcam	ab14746	Western blot	1:1000	5% non-fat dry milk in 1X TBS-T
SDHA	Abcam	ab14715	Western blot	1:3000	5% non-fat dry milk in 1X TBS-T
CS	Novus Biologicals	NBP2-43648	Western blot	1:3000	5% non-fat dry milk in 1X TBS-T
α -tubulin	Sigma Aldrich	T5168	Western blot	1:4000	5% non-fat dry milk in 1X TBS-T
GAPDH	Cell Signaling	2118	Western blot	1:1000	5% non-fat dry milk in 1X TBS-T
UQCRC2	Abcam	ab14745	BN-PAGE	1:1,500	Western Blocking Reagent (Merck)
ATP5F1A	Abcam	ab14748	BN-PAGE	1:2000	Western Blocking Reagent (Merck)
UQCRFS1	Abcam	ab14746	Immunofluorescence	1:100	Antibody Diluent, Background Reducing (Agilent)
VDAC1	Abcam	ab15895	Immunofluorescence	1:400	Antibody Diluent, Background Reducing (Agilent)

The secondary antibodies used for Western blot were anti-rabbit (111-036-045) and anti-mouse (115-036-062), conjugated with horseradish peroxidase (HRP), and purchased from Jackson Immuno Research Laboratories (West Grove, PA, USA).

Protein ladder used during the protein gel electrophoresis was Ladder PageRuler™ Plus Prestained Protein Ladder 10 to 250 kDa (catalog number: 26619), purchased from Thermo Scientific (Waltham, MA, USA).

3.1.4. Chemicals and solutions

Standard laboratory chemicals or solutions were purchased from Sigma-Aldrich (St. Louis, MO, USA) or Merck (Darmstadt, Germany). Most of the chemicals for cell culture were obtained from Gibco or Invitrogen (Thermo Fisher Scientific, Waltham, MA, USA).

3.2. Methods

All methods as described below were performed by myself if not stated otherwise.

3.2.1. Cell culture

3.2.3.1. Cell lines cultivation

Cell lines were cultured in High glucose Dulbecco's Modified Eagle Medium (DMEM, with 4500 mg/L glucose, L-glutamine, sodium pyruvate, and sodium bicarbonate) (41966-029, Life Technologies) supplemented with 10% Fetal Bovine Serum (10270106, Life Technologies), 1% Penicillin-Streptomycin (100 X; 5000 U/ml) (15070063, Life Technologies) and 200 μ M Uridine (referred as DMEM+++) at 37 °C and 5% CO₂.

3.2.3.2. Mycoplasma test and removal treatment

Before any experimental work, all primary cell lines were tested for mycoplasma contamination with MycoAlert Mycoplasma detection kit (LT07-218, Lonza, Basel, Switzerland). In short, 2 ml samples of medium in which cell lines had been growing for at least 48 hours were taken and tested according to the manufacturer's protocol. If cell line was positive for mycoplasma, it was treated with Mycoplasma removal agent (BUF035, Biorad, Hercules, USA) and tested again after the treatment.

3.2.3.3. Cell stock deposition

To keep aliquots of cells of a low passage number, freezing in 10% dimethyl sulfoxide (DMSO) in DMEM+++ was applied. 1.5 ml of cell suspension was transferred into cryo vials, stored in -80 °C for at least 24 hours and finally transferred into liquid nitrogen for a long-term storage. For further work, cell aliquots were slowly warmed up. After thawing the cell suspension, cells were transferred into 75 cm² flask with 15 ml of DMEM+++ . The culturing medium was changed the following day to wash out DMSO. Alternatively, thawed cells were transferred into 15 ml or 50 ml Falcon tube and centrifuged for 3.5 min at 500 rpm, after which supernatant was removed, and the cell pellet dissolved and transferred into 75 cm² flask with 15 ml of DMEM+++ .

3.2.3.4. Cell maintenance

Cells were kept in 75 cm² or 175 cm² flasks suitable for adherent cells. The medium was changed every second or third day. If necessary, cells were splitted by washing once with PBS and deataching with Trypsin-EDTA solution (2.5 ml for 75 cm² flask, 3.5 ml for 175 cm²). When all cells detached, Trypsin was deactivated by adding excess (at least 2x volume of Trypsin-EDTA solution) of DMEM+++ , cells were transferred into 15 ml or 50 ml Falcon

tube, and centrifuged for 3.5 min at 500 rpm. After centrifugation, the supernatant was discarded, and cell pellet was dissolved in DMEM+++ and one fourth (or whichever dilution planned for the experiments) of the cell suspension was transferred back into the new culturing flask. The rest was either divided into additional bottles or used for other experiments. In the case cell pellets were needed for further experiments (eg. DNA, RNA, protein isolation), after centrifugation cells were washed in 1 ml PBS, centrifuged once more for 4 min at 6000 rpm, after which supernatant was removed and pellets were kept in -20 °C.

3.2.3.5. Cell number determination

To determine the total number of cells, I measured cell concentration (number of cells per 1 ml) by dissolving the cell suspension 1:10 in PBS to a total volume of 200 µl in a 1.5 ml Eppendorf tube. 200 µl of cell suspension was analysed by Scepter™ Handheld Cell Counter with 60 µm Scepter™ Sensors by following instruction on the instrument. Measurement was typically performed in duplicates, with average value used.

3.2.3.6. Transfection

The delivery of DNA (usually plasmid) into HEK and HeLa 293FT cells was performed to induce expression of protein of interest. For the purpose, I used Lipofectamine 2000 Transfection reagent (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's protocol. The association of the lipid-based transfection reagent with nucleic acids results in a formation of lipoplexes, cationic complexes, that are then mainly internalized to cells by endocytosis. For transient transfection (DNA is introduced into the nucleus of the cell, but not integrated into the chromosome), expression of transfected gene was analysed 48 h after transfection.

3.2.2. Bacterial culture and techniques

3.2.3.1. Bacterial transformation

For bacterial transformation of lentiviral vectors, One Shot® Stbl3™ chemically competent *E. coli* cells were used. For all other plasmids, I used One Shot™ TOP10 chemically competent *E. coli*. 1 - 5 µl of plasmid (usually 10 pg - 100 ng) was gently mixed into bacterial cells (previously thawed on ice), followed by an incubation on ice for 30 min. Heat-shock of cells was performed for 30 s at 42 °C in a water bath, followed by incubation on ice for 2 min, after which 250 µl S.O.C. medium was added to cells. Then, cells were incubated with horizontal shaking at 37 °C, 45% shaking in Heiz Thermomixer HTML 133 for 1 hour. 50 µl and 100 µl of the cells suspension were spread on the pre-warmed LB agar plates containing antibiotic (Ampiciline or Kanamycin, dissolve stock of 100mg/mL 1:1000) for which plasmid contained resistancy gene and incubated overnight at 37 °C.

3.2.3.2. Colony PCR

During ligation, the amplified copy of the gene of interest can be inserted in both directions (some clonings, such as the TOPO cloning, are not specific in this step) or the amplified PCR product can be shortened or disrupted. Therefore, to ensure the proper size

and orientation of the insert, colony PCR with specific combination of primers was performed after transformation. Following overnight incubation, LB agar plate was inspected and distinct colonies were selected (typically 10-20). Selected colonies were picked, transferred into 20 µl of HPLC water and spread on a new agar plate for further usage and incubation. DNA was released from bacteria by heating the mixture on 95 °C for 20 min in thermocycler and used for a colony PCR. The spreaded colonies were left to grow on a new plate overnight at 37 °C. For colony PCR, 10x PCR Buffer, Q-Solution, dNTP mix and Taq DNA Polymerase were purchased from Qiagen, Hilden, Germany.

The PCR reactions were calculated as following:

Component	µl per 10 µl reaction	Final concentration
HPLC H ₂ O	5.05 µl	
10x PCR Buffer	1 µl	1x
2 mM dNTP mix	1 µl	0.2 mM each
10 µM forward primer	0.4 µl	0.1 µM
10 µM reverse primer	0.4 µl	0.1 µM
Taq DNA Polymerase	0.05 µl	0.25 U / rxn
DNA extracted from colony	2.5 µl	

PCR Program was run on thermocycler as following:

Step	Temperature	Time
Heat lid	110 °C	
Denature	95 °C	5 min
Start cycle (25x)		
Denature	95 °C	1 min
Anneal	primer-depending	30 s
Extend	72 °C	4 min
End cycle		
Extend	72 °C	5 min
Store	10 °C	1 h 10 min

The PCR products were analysed on a 2% agarose gel on the Fusion Pulse (Vilber) imaging system.

After the confirmation of plasmid sequence with Sanger sequencing, positive clones were selected for further amplification. For that, the next day the positive clones were scraped from LB agar plate and transferred into 4 ml of LB medium with adequate antibiotic and cultured overnight at 37 °C while being shaken. The next day, overnight culture was used for plasmid isolation by mini prep and glycerol stock creation, or further amplification.

3.2.3.3. Mini prep

For isolation of up to 20 µg of high-purity plasmid DNA, QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) was used, according to manufacturer's protocol. Concentration of isolated plasmid DNA was subsequently measured on NanoDrop™ OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

3.2.3.4. Midi prep

For isolation of up to 200 µg of high-purity plasmid DNA, QIAGEN Plasmid Midi Kit (Qiagen, Hilden, Germany) was used, according to manufacturer's protocol. Concentration of isolated plasmid DNA was subsequently measured on NanoDrop™ OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

3.2.3.5. Bacterial stock deposition

For long-term storage of bacteria harboring desired plasmid, bacterial glycerol stocks were created. 500 µL of the overnight culture was added to 1 mL of 80% glycerol in a cryovial and gently mixed. The glycerol stock was then frozen and stored at -80 °C. The stock is in this way stable for years. To recover bacteria from the glycerol stock, pipette tip was used to scrape some of the frozen bacteria off of the top and then cultivated in LB medium. As subsequent freeze and thaw cycles reduce shelf life, thawing of glycerol stock was avoided.

3.2.3. DNA and RNA analysis

3.2.3.1. DNA and RNA isolation

DNA and RNA were isolated from whole-cell pellets using AllPrep® DNA/RNA Mini Kit and RNase mini kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. Isolated DNA was then stored at 4 °C and RNA at -80 °C.

3.2.3.2. cDNA synthesis

The isolated RNA was converted into cDNA through reverse transcription with the First Strand cDNA synthesis Kit (Thermo Fisher Scientific, Waltham, MA, USA), according to manufacturer's protocol. The cDNA was stored at -20 °C afterwards.

3.2.3.3. Polymerase chain reaction (PCR)

10x PCR Buffer, Q-Solution, dNTP mix and Taq DNA Polymerase were purchased from Qiagen (Hilden, Germany).

The PCR reactions were calculated as following:

Component	µl per 20 µl reaction	Final concentration
HPLC H ₂ O	11.9 µl	
10x PCR Buffer	2 µl	1x
2 mM dNTP mix	2 µl	0.2 mM each
10 µM forward primer	1 µl	0.5 µM
10 µM reverse primer	1 µl	0.5 µM
Taq DNA Polymerase (5 U/ µl)	0.1 µl	0.5 U / rxn
DNA	2 µl	

Optionally, Q-Solution was added to PCR mix (4 μ l per sample, then used 7.9 μ l of HPLC H₂O instead).

The PCR Program was run on thermocycler as following:

Step	Temperature	Time
Heat lid	110 °C	
Denature	96 °C	2 min
Start cycle (25x)		
Denature	96 °C	30 s
Anneal	primer-depending	45 s
Extend	72 °C	1 min 30 s
End cycle		
Extend	72 °C	7 min
Store	20 °C	1 min

3.2.4. Protein analytics

3.2.3.1. Immunocytochemistry

This work has been in collaboration with Rene Feichtinger (Salzburger Landeskliniken and Paracelsus Medical University, Austria). Fibroblasts were plated at a density of 7,500 and 10,000 cells/well on permanox slides (Lab-Tek Chamber ®Slides, Sigma-Aldrich Chemie GmbH Munich, Germany), respectively, and incubated overnight at 37 °C in 5% CO₂, to reach 70% confluency before the staining. Subsequently medium was removed from the chambers and cells were washed twice with PBS. As PBS was removed, cells were fixed by 15 min room temperature (RT) incubation with 4% paraformaldehyde (PFA). After fixation, 4% PFA was removed and cells were washed three times with PBS-T. Then cells were incubated for 40 min at 95 °C in Antigen Retrieval Buffer (Abcam). Buffer was then removed and slides were washed three times in PBS-T, each wash 3 min. Slides were then incubated for 1 h at RT with primary antibody dilluted in Antibody Diluent, Background Reducing (Agilent). When the incubation was over, cells were washed three times in PBS-T, each wash 3 min. Then cells were incubated with corresponding secondary antibody dilluted in Antibody Diluent, Background Reducing (Agilent) for 1 h at RT. After incubation, slides were washed three times with PBS-T, each wash 3 min. Slides were mounted in ProLong Antifade Reagent containing DAPI (Invitrogen, Eugene, OR, USA) and covered with cover slips (Spezial 24 x 60 mm #1, Menzel-Gläser). Images were taken on a fluorescence microscope.

3.2.3.2. Western blot

Cell pellets were resuspended in 200-400 μ L of 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) Deoxycolate) supplemented with 1:100 Protease Inhibitor Cocktail Set III, Animal-free (Calbiochem, affiliated with Merck, Darmstadt, Germany), depending on the pellet size. Samples were then incubated on a rotating wheel for 1h at 4 °C to be disrupted chemically. Next, samples were centrifuged for 10 min at 11,000 g at 4 °C to remove the unbroken cells. Supernatants were carefully transferred into new tubes and were subsequently disrupted by ten strokes with a 0.30 x 8 mm syringe in order to break the DNA. The whole protein amount was quantified by Bradford assay using Protein Assay Dye Reagent Concentrate (Bio Rad) on Jasco V-550UV/VIS Spectrophotometer. Subsequently, samples to be loaded on a gel were adjusted to 1.5 μ g protein/ μ L in 1x Laemmli buffer (5% (w/v) SDS, 250 mM Tris-HCl pH 6,8, 50% (v/v) glycerol, 500 mM β -mercaptoethanol, 0,025% (w/v) bromphenol blue) and heated up for 10 min at 50 °C. Designated protein amounts were loaded on precast gels (Lonza, Basel, Switzerland). Electrophoresis in 1x ProSieve EX Running buffer (Lonza, Basel, Switzerland) was started at 80 V for 30 min and continued at 120 V for 60 min. Semi dry transfer was performed from gel onto the PVDF membranes (GE Healthcare Life Sciences, Chalfont St. Giles, UK) using 1x ProSieve EX Western Blot Transfer buffer (Lonza, Basel, Switzerland) at a constant current of 375 mA for 15 min. Efficiency of transfer was typically assessed with Ponceau staining of membrane, which unspecifically colors proteins. Membranes were then washed with water and 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 at RT. Immunoblotting with primary antibodies was performed overnight at 4 °C, or for 1 h at RT in case of GAPDH and α -tubulin antibodies (used as loading controls). Specific antibodies binding was visualized by incubation with corresponding secondary antibodies and by using ECL Plus Western Blotting Detection System (GE Healthcare Life Sciences, Chalfont St. Giles, UK) on the Fusion Pulse (Vilber) imaging system. Acquired images were analysed with the ImageJ software.

3.2.3.3. Proteomics

This work has been done in collaboration with Robert Kopajtich and Dmitrii Smirnov from my institute, and BayBioMS, Technical University of Munich, Freising, Germany. Fibroblast whole-cell pellets with 0.5 million cells were lysed under denaturing conditions in a urea containing buffer and quantified using BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). 15 μ g of protein extract were further reduced, alkylated and digested using Trypsin Gold (Promega). Digests were acidified, desalted and TMT-labeled as described (Zecha et al., 2019) with the TMT 10-plex labeling reagent (Thermo Fisher Scientific, Waltham, MA, USA). Each TMT-batch consisted of 8 patient samples and 2 reference samples, allowing for data normalization between batches. Peptide identification was performed using MaxQuant version 1.6.3.410 and protein group intensities were row- and column-wise normalized to account for the variability within and between the TMT-batches. Outliers and mitochondrial complexes were identified and quantified using PROTREIDER (in development), with significance threshold of adjusted p-value < 0.05.

3.2.5. Biochemical measurements

3.2.3.1. Seahorse measurements

Cellular oxygen consumption can be used as a read-out of the mitochondrial fitness as mitochondrial respiratory chain presents the biggest oxygen consumer in the cell. Electron donors can derive from glucose, fatty acids and glutamine. Mito Stress test was used to access the glucose-driven respiration. The afternoon/evening before the assay, 20,000 fibroblasts/well were seeded in 80 μ L of DMEM+++ in a XF 96-well cell culture microplate (Seahorse Bioscience, Agilent Technologies, Santa Clara, CA, USA) with corner wells left without cells, just filled with DMEM+++ as a background measurement. Plated cells were incubated at 37 °C and 5% CO₂ overnight. The sensor cartridge was rehydrated in the XF96 utility plate by adding 200 μ l of XF calibrant solution per well. Next day cells were carefully washed once with and subsequently incubated in a 180 μ L bicarbonate-free DMEM (Life Technologies, Carlsbad, CA, USA) at 37 °C for at least 30 min prior measurement. The oxygen consumption rate (OCR) was measured using the XF96 Extracellular Flux Analyzer (Agilent). 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 μ M) loaded in Port A; after addition of FCCP (0.4 μ M) loaded in Port B; after addition of rotenone (2 μ M) loaded in Port C, and after addition of antimycin A (2 μ M) loaded in Port D. Following the measurement, cells were washed with 200 μ l PBS and Seahorse plate kept at -80 °C. Data was analysed using the Wave Desktop Software (Seahorse Bioscience, Agilent Technologies, Santa Clara, CA, USA), as well as OCR-Stats (Yepez et al., 2018). For OCR stats, the calculated parameters of cellular respiration are: non-mitochondrial respiration (minimum rate measurement after antimycin A injection), basal respiration (last measurement before oligomycin injection minus non-mitochondrial respiration), ATP-production (last measurement before oligomycin injection minus minimum rate measurement after oligomycin injection), proton leak (minimum rate measurement after oligomycin injection minus non-mitochondrial respiration), maximum respiration (maximum measurement after FCCP injection minus non-mitochondrial respiration), and spare respiratory capacity (maximal respiration minus basal respiration).

3.2.3.2. Determination of cell number by CyQuant assay

Cells attach with a different strength and multiply with a different grow rate, therefore the cell number must be again determined after finishing experiments, especially the Seahorse assay. For the determination by CyQuant assay, cells were washed in PBS and frozen at -80 °C for at least 4 h to disrupt the cell membrane. The cell number quantification was done using the CyQuant Cell Proliferation Kit (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's protocol. Fluorescence was measured on Cytation 3 imaging reader (BioTek).

3.2.3.3. Malate Dehydrogenase 2 (MDH2) Activity Assay

To measure enzymatic activity of MDH2 in a sample, commercially available Malate Dehydrogenase 2 (MDH2) Activity Assay (ab119693, Abcam) was used, according to manufacturer's protocol. Absorbance at 450nm was measured on Cytation 3 imaging reader (BioTek).

3.2.6. Sequencing methods

3.2.3.1. Sanger sequencing

For Sanger sequencing, isolated plasmid or PCR product were used as templates. PCR product was purified using the MultiScreen® PCRµ96 Filter Plate (Millipore, Merck KGaA, Darmstadt, Germany) according to the manufacturer's protocol. If not used immediately, purified PCR products were stored at -20 °C. The plasmid or purified PCR product were used for subsequent cycle sequencing using the ABI BigDye Terminator v.3.1 Cycle Sequencing kit (Life Technologies, Carlsbad, CA, USA) as follows:

Component	µL per 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/plasmid	1 µL	
Ultrapure H ₂ O	1 µL	

Sequencing reaction was run as follows:

Step	Temperature	Time
Heat Lid	110 °C	
Denature	96 °C	1 min
Start Cycle (25x)		
Denature	96 °C	10 s
Anneal	50 °C	5 s
Extend	68 °C	2 min
End cycle		
Temperature	20 °C	1 min

If not used immediately, samples were stored at -20 °C. 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 4 °C. The pellet was washed with 125 µL 70% ethanol, centrifuged at 2000 g for 10 min, and left to dry at RT in dark. Subsequently, the pellet was resuspended in 25 µL ultrapure H₂O, transferred to a microtiter plate, and placed into the automated ABI 3730 sequencer. Resulting sequences were analysed using SnapGene Software (<https://www.snapgene.com/>).

3.2.3.2. Whole exome sequencing

Whole exome sequencing was performed in collaboration with NGS core facility in HelmholtzZentrum München, Neuherberg, Germany, and group of Dr. Tim Strom from Institute of Human Genetics, Technical University of Munich, Germany. DNA was isolated

from peripheral blood leucocytes or fibroblasts using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA concentration was measured using the Qubit™ dsDNA BR Assay Kit and 3 µg of DNA was used for library preparation. Exonic regions from human DNA samples were enriched using the SureSelect Human All Exon V5/V6 kits from Agilent (Agilent Technologies, Santa Clara, CA, USA) and subsequently sequenced as 100 bp-end runs on a Illumina HiSeq2500 and Illumina HiSeq4000 (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). Single nucleotide variants (SNVs) as well as small insertions and deletions (indels) were detected with SAMtools (version 0.1.19) and GATK 3.8. On average, 85,000 SNVs were detected per individual. For further analysis in diagnostic purposes, 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 was further facilitated by highlighting genes which are: 1) known mitochondrial disease-associated genes (Stenton and Prokisch, 2020); 2) encoding reported mitochondrial proteins; and 3) encoding predicted mitochondrial proteins (Elstner et al., 2009).

3.2.3.3. RNA-sequencing

RNA-sequencing was done in collaboration with NGS core facility in HelmholtzZentrum München, Neuherberg, Germany, group of Dr. Tim Strom from Institute of Human Genetics, Technical University of Munich, Germany, and group of Prof. Julien Gagneur from Department of Informatics, Technical University of Munich, Garching, Germany. Cell pellets necessary for RNA-sequencing were prepared by me, and my colleagues Robert Kopajtich, Caterina Terrile, Agnieszka Nadel and Dr. Laura Kremer. RNA was isolated from fibroblasts whole-cell pellets using the AllPrep RNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. RNA integrity number (RIN) was determined by the Agilent 2100 BioAnalyzer (RNA 6000 Nano Kit, Agilent Technologies, Santa Clara, CA, USA). Non-strand specific RNA-sequencing was performed as described in Kremer et al. (2017), according to TruSeq RNA Sample Prep Guide (Illumina). Library preparation for strand-specific RNA-sequencing was performed according to the TruSeq Stranded mRNA Sample Prep LS Protocol (Illumina, San Diego, CA, USA). In short, 1 µg of RNA was purified using poly-T oligo attached magnetic beads, and fragmented. RNA fragments were reverse transcribed with the First Strand Synthesis Act D mix. Second strand cDNA was synthesized with Second Strand Marking Mix that enables strand specificity by replacing dTTP with dUTP. The resulting double-stranded cDNA was subjected to end repair, A-tailing, adaptor ligation, and library enrichment. 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). RNA libraries were sequenced as 100 bp paired-end runs on an Illumina HiSeq2500 and HiSeq4000 platforms. Reads were demultiplexed and mapped with STAR (Dobin et al., 2013) to the hg19 genome assembly. FPKM (fragments per kilobase per millions of reads) values were obtained with DESeq2

(Love et al., 2014). Expressed genes were defined as being present in at least 5% of the samples with an FPKM value greater than 1 (Brechtmann et al., 2018). Identification of aberrant expression, splicing and mono-allelic expression (MAE) was based on the Detection of RNA Outliers Pipeline (Yepez et al., 2021). We used as a reference genome the GRCh37 primary assembly, release 29, of the GENCODE project (Frankish et al., 2019). To better target the counts, the exonic regions of every gene were extracted with the `exonsBy` function from the `GenomicFeatures` (Lawrence et al., 2013) R package. For counting reads that are paired with mates from the opposite strands, `SummarizeOverlaps` function from the `GenomicAlignments` (Lawrence et al., 2013) R package was used. `OUTRIDER` and `Fraser` algorithms were used for the detection of aberrant expression and aberrant splicing, respectively (Brechtmann et al., 2018; Mertes et al., 2021). For MAE detection, only heterozygous single nucleotide variants from WES were considered. Reads assigned to each allele were counted with the `ASEReadCounter` function (Castel et al., 2015) from `GATK`. Positions with less than 10 reads were discarded. MAE was called if a rare alternative variant was present in at least 80% of total reads at the position. RNA-seq variant calling was performed using `GATK` best practices for RNAseq short variant discovery (<https://gatk.broadinstitute.org/hc/en-us>). Variants with a ratio of quality to depth of coverage < 2, strand biased (phred scaled fisher exact score >30), or belonging to a SNP cluster (3 or more SNPs found within a 35 base window) were filtered out. Only variants absent from a repeat masked region (<http://www.repeatmasker.org>), and with at least three reads supporting the alternative allele were considered.

3.2.7. Functional complementation by lentiviral transduction

To experimentally validate the candidate variant as indeed disease-causing, functional complementation assays were performed as described (Kremer and Prokisch, 2017). The assay is based on the rescue of the observed phenotype by over-expression of a wild-type copy of a gene of interest in human fibroblasts. The “rescue” plasmid was generated by inserting a wild-type cDNA of a gene of interest into a pLenti6.3 vector and delivered into fibroblasts by lentiviral transduction, according to the pLenti6.3/V5-TOPO TA Cloning Kit (Invitrogen, Eugene, OR, USA) and ViraPowerTM Lentiviral Support Kit (Invitrogen, Eugene, OR, USA), respectively.

3.2.3.1. Primer design

To amplify a wild-type copy of a cDNA, specific primers were designed flanking the open reading frame (ORF) of the gene of interest. To secure protein expression, Kozak consensus sequence (CGCC, alternatively CACC) was added at the beginning of the forward primer directly before the start ATG codon. The reverse primer was designed by starting at the stop codon. Specificity of the primers was verified by UCSC In-Silico PCR (<https://genome.ucsc.edu/cgi-bin/hgPcr>) and melting temperature was determined by NEB Tm calculator (<http://tcalculator.neb.com/>). The melting temperatures of both primers were different by maximum of 5 °C.

3.2.3.2. High-fidelity PCR

High-fidelity PCR utilizes a DNA polymerase with a low error rate, ensuring a high degree of accuracy in the replication of cDNA of interest. Generated or purchased cDNA was used as template for amplification of a wild-type copy of a gene of interest. The annealing temperature was calculated based on the lower melting temperature of both primers ($T_a = T_m_lower + 3\text{ }^\circ\text{C}$). High Fidelity PCR buffer, MgSO₄ and Platinum® Taq DNA Polymerase High Fidelity were purchased from Invitrogen (Eugene, OR, USA).

The PCR reactions were calculated as following:

Component	μL per 10 μL reaction	Final concentration
Ultrapure H ₂ O	5.76 μL	
10x High Fidelity PCR Buffer	1 μL	1x
2 mM dNTP mix	1 μL	0.2 mM each
50 mM MgSO ₄	0.4 μL	2 mM
10 μM forward primer	0.2 μL	0.2 mM
10 μM reverse primer	0.2 μL	0.2 mM
Platinum® Taq DNA Polymerase High Fidelity (5 U/ μL)	0.04 μL	0.2 U
Template cDNA	1 μL	<5 ng

The PCR program was run as following:

Step	Temperature	Time
Heat lid	110 $^\circ\text{C}$	
Denature	95 $^\circ\text{C}$	2 min
Start cycle (30x)		
Denature	95 $^\circ\text{C}$	30 s
Anneal	primer-depending	30 s
Extend	68 $^\circ\text{C}$	1 min/kb
End cycle		
Extend	68 $^\circ\text{C}$	10 min
Store	20 $^\circ\text{C}$	1 min

The PCR products were checked for the correct size by being run on a 2% agarose gel (100 V, 45 min) and visualized on the Fusion Pulse (Vilber) imaging system.

3.2.3.3. TOPO TA cloning

Cloning was performed according to the User Manual for pLenti6.3/V5-TOPO® TA Cloning Kit (Invitrogen #K531520), as following:

Component	μL per 5 μL reaction
Sterile water	To 5 μL
Fresh PCR product	0.5 – 1 μL
Salt solution	1 μL
pLenti 6.3/V5 TOPO® vector	1 μL

The ligation mixture was kept for 5 min at RT. 2 μL of the mixture was used for transformation of One Shot® Stbl3™ chemically competent *E. coli* cells. After identification of clones with the correct insert (harboring no mutations) by the colony PCR and Sanger sequencing, *E. coli* were amplified in a 50 ml overnight culture containing ampicillin. Next day, midi prep was performed and plasmid concentration measured, as described previously.

Transduction and transfection:

The following media (1-4) were filtered through a 0.22 μm filter using a SteriCup® and SteriTop® from Millipore (Merck KGaA, Darmstadt, Germany).

- Medium 1: DMEM, 10% (vol/vol) FBS
- Medium 2: DMEM, 10% (vol/vol) FBS, 1% (vol/vol) penicillin-streptomycin and 200 μM Uridine
- Medium 3: DMEM
- Medium 4: DMEM, 2% (vol/vol) FBS
- Medium 5: normal fibroblast cell culture medium (nonfiltered) – DMEM+++

Day 1. HEK293 FT cells were thawed and kept in medium 1. Patient and control cell lines were also thawed and cultured in DMEM+++.

Day 3. HEK293FT cells were split and seeded on a 10 cm dish with approximately 8 ml of medium 1. One dish was sufficient for the transduction of two cell lines (typically patient and control). The cells were seeded to reach 70-80% confluence in the evening on the following day. This is done as too confluent HEK293 cells form a monolayer that can easily detach from the surface.

Day 4. Medium 1 and 3 are warmed up and Packaging Mix is thawed on ice. In 15 ml Falcon tube 1.5 ml mixture of medium 3 and 36 μl of Lipofectamine 2000 was prepared and incubated at RT for 5 min. Meanwhile, in 1.5 ml of medium 3 9 μl of Packaging Mix and 3 μg of pLenti vector (as calculated previously) were included. This mixture was dropwise added to the Lipofectamine 2000 solution and gently mixed by inverting the tube. Next, this mixture was incubated for 20 min at RT. This procedure leads to an enclosing of the vector DNA into small packages that can then be transferred in the target cells by lipofection, thus making small holes in the lipid layer of the cell. In the meantime, the medium of HEK 293 FT cells prepared the previous day is aspirated and replaced with 5 ml of fresh medium 1. When incubation ends, the solution containing lipoplexes is dropwise added to the cells. All steps are carried out carefully to avoid the detachment of the cells. Transfected cells are then incubated overnight. From this moment, HEK 293 FT cells belonged to biosafety level 2.

Day 5. On the next day, the patient fibroblasts and NHDF cells were split and plated in a 10 cm dish with 8 ml of medium 2 in an amount that would make them reach 70-80% confluence on the day of transduction. The same day the medium of transfected HEK 293 FT cells was carefully replaced with 7 ml of medium 4, and incubated for 72 hours.

Day 8. The lentiviral transduction is performed with level 2 of biosafety. First, the centrifuge was cooled down to 4 $^{\circ}\text{C}$ and medium 2 warmed up. The supernatant from HEK 293 FT cells was carefully transferred into 15 ml Falcon tubes and centrifuged at 2000g for 15 minutes at 4 $^{\circ}\text{C}$. 3-4 ml of medium 2 was prepared in a 50 ml Falcon tube. After centrifugation, supernatant was filtered through a 45 μm filter into the 50 ml Falcon tube containing medium 2 and mixed gently. The medium from fibroblasts was aspirated and interchanged with the one containing viral supernatant (7 ml).

Day 9. The medium of transduced fibroblasts was replaced by medium 5.

Day 10. The selection was started by addition of blasticidine (Thermo Fisher Scientific, Waltham, MA, USA) to the medium 5 to a final concentration of 5 µg/ml. This resulted in survival of only cells stably expressing the gene of interest.

The cells are typically virus free in two weeks. Cells were kept in culture in selection medium until they reached sufficient number to create stocks and perform experiments.

3.2.8. Statistical analysis

Where the statistics is indicated, the two-tailed two-sample Student's t-test was followed to determine the statistical significance, assuming a normal distribution. Error bars indicate standard deviation or standard error of the mean. All experiments were performed at least three times unless stated otherwise. A p value of < 0.05 was considered statistically significant (*p < 0.05; **p < 0.01; ***p < 0.001, NA- not significant). Z-score was computed as: value minus mean across samples, divided by standard deviation across samples.

3.2.9. Online resources

- ClinVar, <https://www.ncbi.nlm.nih.gov/clinvar/>
- DROP, <https://github.com/gagneurlab/drop>
- GeneMatcher, <https://genematcher.org/>
- GATK best practices for RNAseq short variant discovery, <https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels->
- GATK, <https://gatk.broadinstitute.org/hc/en-us>
- GTEEx, <https://www.gtexportal.org/home/>
- RepeatMasker, <http://www.repeatmasker.org/>
- HGMD, <http://www.hgmd.cf.ac.uk/ac/index.php>
- Human Splicing Finder, <https://www.genomnis.com/access-hsf>
- MutationTaster, <http://www.mutationtaster.org/>
- OMIM, <https://omim.org/>
- SnapGene, <https://www.snapgene.com/>
- Tm Calculator, <https://tmcalculator.neb.com/#!/main>
- UCSC In-Silico PCR, <https://genome.ucsc.edu/cgi-bin/hgPcr>

4. Results

4.1. *MDH2* variants cause severe neurological presentations

Published as “Mutations in *MDH2*, Encoding a Krebs Cycle Enzyme, Cause Early-Onset Severe Encephalopathy” by Ait-El-Mkadem*, Dayem-Quere*, Gusic* et al. in the *American Journal of Human Genetics* in 2017 (PMID: 27989324).

My work focused on one of three described patients (referred to as “P3”), where I reanalysed the WES, discovered the candidate gene, performed segregation analysis, connected with other researchers via GENOMIT and GeneMatcher, did cell culture work with P3’s fibroblasts, performed Western blot, enzymatic activity assay and microscale respirometry.

4.1.1. Case reports

Three patients (P1, P2 and P3) described in this study were independently referred for WES with a suspicion of a mitochondrial disease. My main focus was on P3, whose WES and fibroblasts cell lines I received as a part of GENOMIT and WES-negative cases. All three presented with an early-onset generalized hypotonia, psychomotor delay, and refractory epilepsy (clinical presentations are summarized in Table 6, and in Ait-El-Mkadem et al. (2017)). Upon identification of a candidate disease gene- *MDH2* in common, P1 and P2 were matched thanks to the GeneMatcher (Sobreira et al., 2015), and P3 was matched via GENOMIT, a global network of mitochondrial disorders researchers (<http://genomit.eu/>). Informed consent for diagnostic and research studies was obtained in accordance with the Declaration of Helsinki protocols and approved by local ethics committees. P1 and P3 were born in France, and P2 in the Netherlands.

P3 is a male child who was born at term to healthy, unrelated French parents. He initially presented with severe hypotonia, macrocephaly, macrosomia, and two supernumerary nipples. Within 1st year of life, he exhibited psychomotor delay with refractory epilepsy, dystonia, and failure to thrive despite tube feeding through percutaneous gastrostomy. P3 also carried von Willebrand disease and congenital cystic adenomatoid malformation (CCAM). Marked cerebral and cerebellar atrophy were observed on brain MRI at the age of 4 years. Lactate was elevated in blood (2.8 mmol/L [normal range < 2.20mmol/L]). Marked cerebral and cerebellar atrophy were observed on brain MRI at the age of 4 years. A slight decreases in complex V and complex I activities were measured in muscle and fibroblasts, respectively.

Table 6. Genotype and clinical presentations of patients with biallelic MDH2 variants.

	Patient 1 (P1)	Patient 2 (P2)	Patient 3 (P3)
Gender	Male	male	male
Last examination	4.5 years	died at 1.5 years	7.5 years (lost view)
MDH2 variants	c.398C>T (p.Pro133Leu) c.620C>T (p.Pro207Leu)	c.398C>T (p.Pro133Leu) c.596delG (p.Gly199Alafs*10)	c.109G>A (p.Gly37Arg) c.398C>T (p.Pro133Leu)
Age at onset	5 months	neonatal	neonatal
Refractory epilepsy	+ (partial, afterward myoclonic; onset at 7 months)	+ (generalized tonic and spasms; onset at 2 months)	+ (myoclonic epilepsy and generalized tonic; onset ?)
Hypotonia	+	+	+
Developmental delay	+	+	+
Head control	10 months	not acquired at 6 months	12 months
Sitting position	18 months	not acquired at 12 months	–
Crawling	18 months	not acquired at 12 months	no
Good eye contact	yes	yes	no
Language	not acquired	babbling at 12 months	not acquired
Muscle weakness	+	+	+
Failure to thrive	+	–	+
Gastrostomy (age)	+ (3 years)	–	+ (?)
Last examination	4 years	18 months	7.5 years
Length	< –2 SDs	+ 1 SD	–2 SDs
Head circumference	< –2 SDs	+ 1 SD	+2 SDs
Weight	< –3 SDs	+ 1 SD	–2.5 SDs
Movement disorders	dystonia and dyskinesia	–	dystonia
Obstinate constipation	+	+	–
Ophthalmologic examination (age)	retinitis pigmentosa (4 years), strabismus (5 months)	strabismus (1 year)	?
Pyramidal signs	+	+	–
Deep tendon reflexes	decreased	N	–
Plantar responses	bilateral extensor	bilateral extensor after 1 year	normal
Other findings	–	–	two supernumerary nipples, von Willebrand disease, CCAM
Ketogenic diet (onset and response)	+ (3 years), reduction epileptic seizure frequency	+ (18 months), reduction epileptic seizure frequency	+ (3 years), ?
Evolution	alive at 5 years	died at 1.5 years	alive at 12 years
Brain MRI abnormalities	+	+	+
Lactate concentration			
Plasma (N < 2.20 mmol/l)	elevated (3.0)	elevated (5.7)	elevated (2.8)
L/P ratio (N < 18)	elevated (63)	elevated (23)	elevated (20)
CSF (N < 2.10 mmol/l)	elevated (2.48)	elevated (3.3)	ND
Krebs cycle intermediates			
Malate (N < 7 μmol/mmol creatine)	elevated (56)	elevated (15–38)	ND
Fumarate (N < 14 μmol/mmol creatine)	elevated (20)	N or elevated (9–55)	N
Succinate	N	N	N
OXPHOS activity			
Muscle	N	ND	slightly reduced CV activity
Liver	reduced CV activity	ND	ND
Fibroblasts	N	ND	slightly reduced CI activity

*Abbreviations: +, present; –, absent; ?, unknown; N, normal; ND, not done; SD, standard deviation; CCAM, congenital cystic adenomatoid malformation; MRS, magnetic resonance spectroscopy; CC, corpus callosum; L/P, lactate/pyruvate.

4.1.2. WES reveals biallelic variants in *MDH2* in three unrelated families

WES was performed with a DNA sample from P3 at our institute. Initial diagnostic analysis of WES failed to detect pathogenic variants in known disease genes that could explain patient's phenotype. Considering the autosomal recessive mode of inheritance, the variants were next filtered to be rare (MAF <0.001 within gnomAD and a frequency <0.05 among our internal database of over 16,000 exomes), revealing 45 genes with potentially biallelic, non-synonymous variants. Based on a suspicion of a mitochondrial disease, *MDH2*, harboring two missense variants, was pinpointed as the most promising candidate, being the only mitochondrial protein-coding gene in the list. Identified variants were both predicted to be probably damaging: c.109G>A (p.Gly37Arg) and c.398C>T (p.Pro133Leu) (RefSeq ID: NM_005918.3) by PolyPhen2 (scores: 0.995 and 0.996, respectively) and SIFT (scores: 0.01 and 0, respectively). c.109G>A variant was found three times, and the c.398C>T variant thirteen times in gnomAD, but both only in a heterozygous state. To confirm the variants and their compound heterozygosity, Sanger sequencing was performed. The segregation analysis revealed a heterozygous c.398C>T inherited from mother and *de novo* c.109G>A (Figure 14). Paternity was confirmed by genotype analysis of seven informative microsatellites.

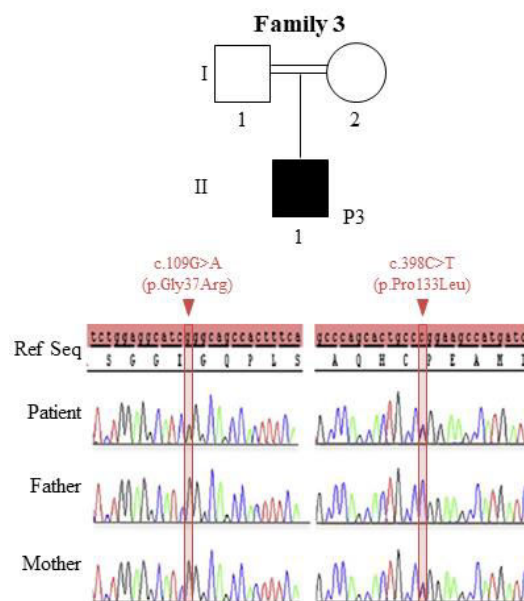


Figure 14. P3 inherited variants from each parent. Pedigree of family 3, with Sanger sequencing analysis below, showing identified variants on the DNA level in P3 and segregation analysis.

Independently, additional carriers of potentially deleterious biallelic *MDH2* variants were identified (Ait-El-Mkadem et al., 2017). WES analysis in P1 (as in Bannwarth et al., 2014) identified compound heterozygous missense variants in *MDH2*: c.398C>T (p.Pro133Leu) (as in P3) and c.620C>T (p.Pro207Leu). His healthy sister was a heterozygous carrier of the paternal variant. In P2, WES identified compound heterozygous *MDH2* variants: c.398C>T (p.Pro133Leu) missense variant (as in P1 and P3), inherited from father, and c.596delG (p.Gly199Alafs*10) nonsense variant, inherited from mother.

4.1.3. MDH2 in detail

MDH2 (mitochondrial malate dehydrogenase, malate dehydrogenase 2, NP_005909.2) is a Krebs cycle enzyme which catalyzes the reversible oxidation of malate to oxaloacetate, utilizing the NAD⁺/NADH cofactor system. It also plays a role in the malate-aspartate NADH shuttle (Goward and Nicholls, 1994). It is encoded by the *MDH2* gene, located on the Chr7 (NC_000007.14; Chr7:76048106-76067508). *MDH2* is ubiquitously expressed with high expression primarily in muscle and heart (<https://www.gtportal.org/home/>). It is a 36-kDa, 338 amino acid (AA) enzyme with an NAD binding site at positions 140 to 142 (Figure 15). Its first 24 AAs represent an MTS. Crystal structure of human MDH2 by the X-ray diffraction at 1.9 Å resolution (PDB:2DFD; Figure 15) reveals the dimer interface, which consists mainly of interacting alpha-helices that form a compact interaction and well separated catalytic domains. Dimer structure of MDH2 indicates the important connection between the protein stability and catalytic activity.

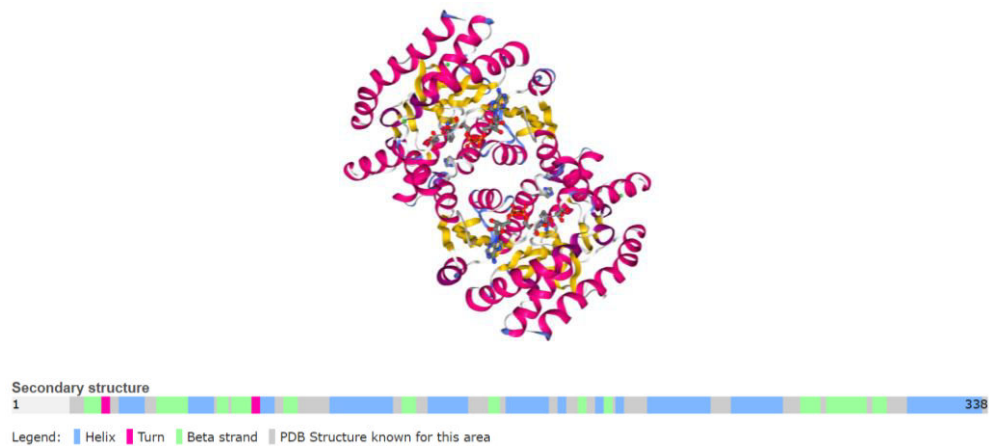


Figure 15. MDH2. Upper figure: 3D representation of the crystal structure of MDH2 (PDB: 2DFD), shown as a homodimer with color code representing the secondary structure. Lower figure: Secondary structure of the MDH2 with the corresponding amino acid residues (from Uniprot, P40926).

4.1.4. Observed MDH2 variants are predicted as deleterious

MDH2 is located on chromosome 7 and consists of nine exons, where the first one is not translated. As depicted in Figure 16, the four observed variants are located in exon 2, exon 4, and exon 6, respectively. These missense variants affect highly conserved amino acid residues, whereas the c.596delG leads to a frameshift and a stop codon ten AA downstream.

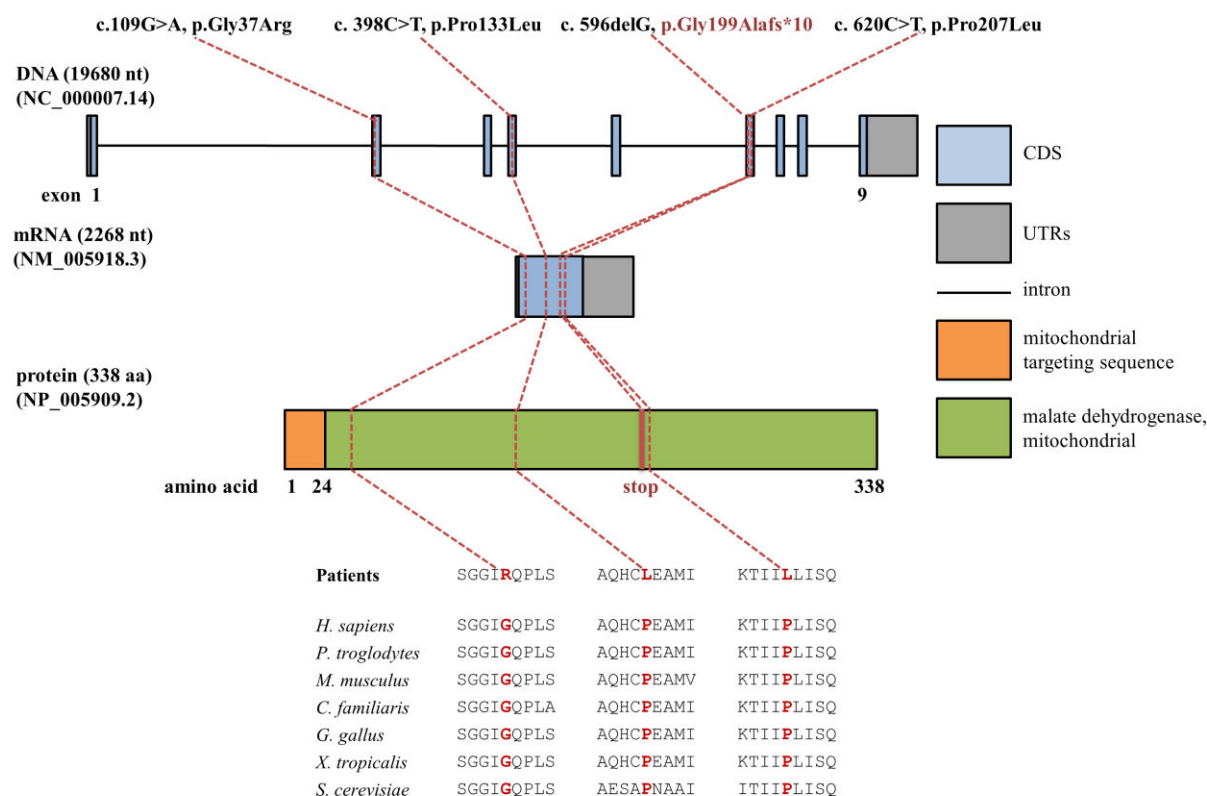


Figure 16. Effect of *MDH2* variants on gene and protein. Genomic organization of *MDH2* into nine exons, the mRNA, and the protein structure showing the positions of the identified variants. The localization of the affected amino acids is highlighted in red on the protein structure. Phylogenetic conservation of the affected amino acid residues is shown in red in the alignment of homologs across different species.

4.1.5. *MDH2* variants causes the loss of MDH2 and its enzymatic activity

To confirm the predicted deleterious effect of the variant, I performed Western blot analysis on P3-derived fibroblasts protein lysates. Immunoblotting was performed using antibodies against MDH2, and alpha-tubulin, which served as a loading control. The experiment showed that MDH2 levels were not detectable in patient's fibroblasts (Figure 17), confirming that the identified *MDH2* variants adversely affect the stability of the protein. Same results were observed in fibroblasts protein lysates of P1 by our collaborators (Ait-El-Mkadem et al., 2017).

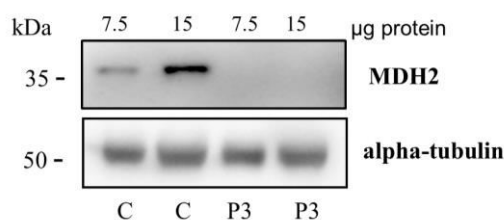


Figure 17. *MDH2* variants affect protein levels. Western blot analysis of fibroblasts' protein lysates of patient 3 (P3) and control cell line (C) reveals levels of MDH2 protein in P3 below the sensitivity of the antibody.

In addition, I measured the MDH enzymatic activity with the mitochondrial malate dehydrogenase (MDH2) Activity Assay Kit in P3 and control (C1 and C2) fibroblasts, which expectedly showed significantly decreased levels of MDH2 in patients compared to controls, close to the blank values (Figure 18). Similarly, markedly decreased MDH2 enzymatic activity was also found in P1 compared to the control cell lines by our collaborators, as well as its notably lower levels in fibroblasts of patient's parents, suggesting that each missense variant has a deleterious effect (Ait-El-Mkadem et al., 2017).

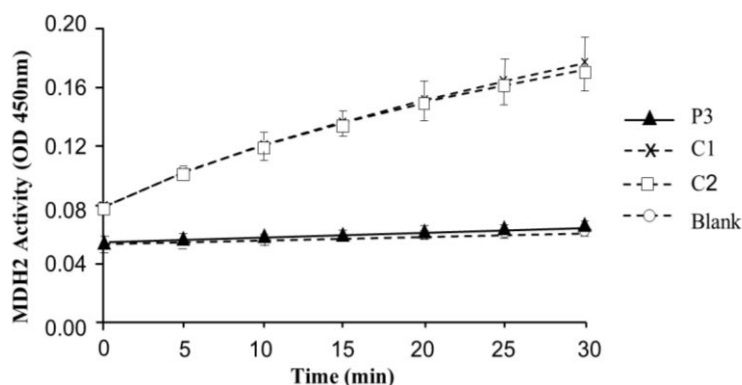


Figure 18. *MDH2* variants affect MDH2 enzymatic activity. MDH2 activity in the fibroblasts from patient 3 (P3), and two control cell lines (C1 and C2, respectively) shows ablation of MDH2 enzymatic activity in P3. Data represent three independent experiments performed in a duplicate. Error bars present standard deviation.

4.1.6. Loss of MDH2 does not affect fibroblasts respiration

Having in mind close connection between the Krebs cycle and OXPHOS, I analysed cellular respiration in glucose-rich medium by performing microscale respirometry (Seahorse experiment). The fibroblast cell lines underwent mitochondrial stress test. The data did not reveal significant impairment of mitochondrial respiration in comparison to control, while defects in other cell lines were well recorded (Figure 19).

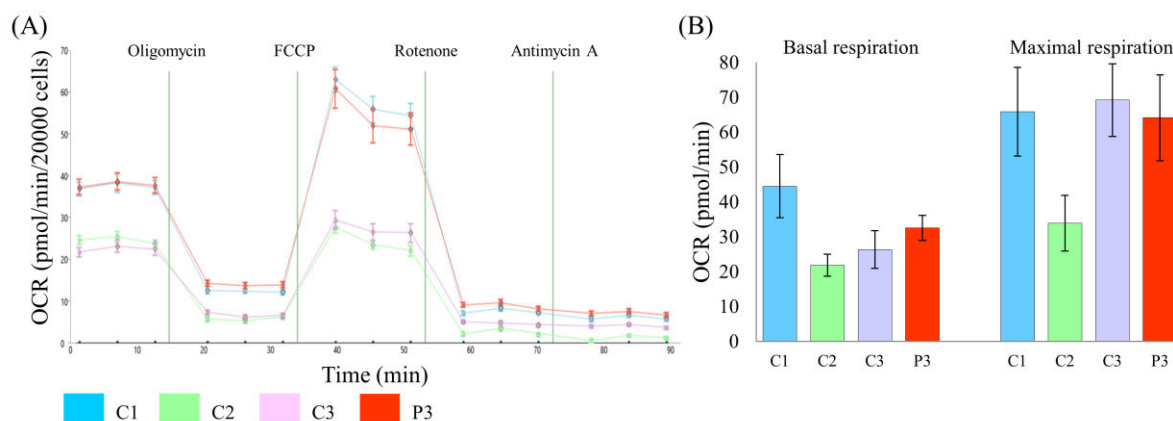


Figure 19. *MDH2* variants do not impact mitochondrial respiration. Oxygen consumption rates (OCR, in pmol/min) behaviour of patient 3 (P3), control cell line (C1), and other fibroblasts that served as a positive control with an OXPHOS defect (C2, C3) over time normalized to the cell count show no significant difference in P3 compared to healthy control (C1), but yes compared to C2 and C3. At least 24 wells were seeded with each sample per plate, and at least three plate replicates were performed for P3. The first three time intervals are obtained after basal conditions and the next twelve after injecting oligomycin, FCCP, rotenone and antimycin A, respectively. Figure shows one representative experiment, with each data point representing mean \pm standard error of the mean (SEMs) of measurements from technical replicates. (B) Basal and maximal respiration (OCR, in pmol/min), calculated with OCR-stats (Yopez et al., 2018), normalized to cell count in a combination of experiments, shown as mean \pm SEM. No significant impairment of P3 compared to C1 was observed.

4.1.7. *MDH2* variants in P1 cause disruption in Krebs cycle metabolites

To investigate the effect of *MDH2* deficiency on metabolite ratios, liquid chromatographic tandem-mass spectrometry was performed by our collaborators as described (Cascon et al., 2015; Ait-El-Mkadem et al., 2017). It expectedly revealed that ratios of malate and fumarate to citrate, respectively, were significantly higher in P1 fibroblasts compared to the control cell lines (C), without accumulation of succinate (Figure 20). In addition, the remaining MDH activity in fibroblasts of parents (father F1 and mother M1) of P1 is enough to maintain normal ratios of malate/citrate and fumarate/citrate.

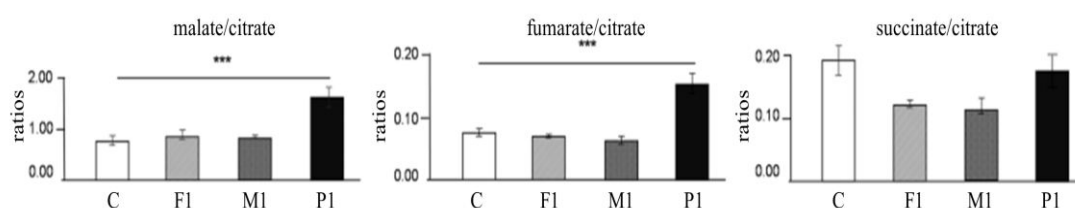


Figure 20. *MDH2* variants disrupt Krebs cycle metabolites. Metabolite ratios in patient 1 (P1) fibroblasts compared with those from his parents (F1 and M1) and control cell line (C) show significant increase in malate/citrate and fumarate/citrate ratios in P1. Abbreviation: ***, $p < 0.001$ (Student's t-test). Figure taken from Ait-El-Mkadem et al. (2017).

4.2. *UQCRFS1* variants cause complex III deficiency, cardiomyopathy and *alopecia totalis*

Published as “Bi-Allelic *UQCRFS1* Variants Are Associated with Mitochondrial Complex III Deficiency, Cardiomyopathy, and Alopecia Totalis” by Gusic*, Schottmann*, Feichtinger* et al. in the *American Journal of Human Genetics* in 2020 (PMID: 31883641).

In the context of the DZHK project I was focused on genetic analysis of WES-undiagnosed patients with cardiac phenotypes.

My work focused on one of two described patients (referred to as “P1”), where I reanalysed the WES, discovered the candidate gene, performed segregation analysis, connected with other researchers via mitoNET, did cell culture work with P1’s fibroblasts, performed RT-PCR, RNA-seq, microscale respirometry, and lentiviral complementation.

4.2.1. Case reports

Two patients (P1 and P2) described in this study were independently referred for WES with a suspicion a mitochondrial disease with cardiac manifestations. My main focus was on P1. Both patients exhibited early-onset CIII deficiency, cardiomyopathy and *alopecia totalis* (clinical presentations are summarized in Table 7 and in Gusic et al. (2020)). Upon identification of *UQCRFS1* as a candidate disease gene- *UQCRFS1* in common, both cases were matched thanks to the German network for mitochondrial disorders, mitoNET. Informed consent for diagnostic and research purposes was obtained in accordance with the Declaration of Helsinki protocols and approved by local ethics committees. P1 and P2 were both born in Germany, with their genotype and clinical presentations summarized in Table 7.

P1 was a male, first child of consanguineous Afghan parents. He was born at term by an emergency Caesarean section due to fetal bradycardia. At the first day of life physicians noted hypothermia, borderline thrombocytopenia [152/nl; N >150], lactic acidosis [24 mmol/l; N<2.0], and elevated creatine kinase levels [>5.000 IU/l; N <190]. Skin and skeletal muscle biopsies were performed during the first day of life. Measurements of OXPHOS activities showed normal values in the muscle and low-normal values in skin fibroblasts for combined CII+III activity. Metabolic urine measurements revealed increased excretion of ketone bodies and lactate. Plasma alanine was markedly elevated. Thiamine and CoQ₁₀ supplementation was initiated. EEG showed slightly pathologic baseline activity initially, but got normal at 7 weeks of age. Neonatal screening revealed hearing impairment. Echocardiography at the first day of life showed septum and right ventricle hypertrophy, increased right-ventricular pressure, and a patent ductus arteriosus with bi-directional shunting. At day 13, hypertrophic cardiomyopathy had progressed and was treated with metoprolol. Echocardiography at 2 months showed reduced biventricular function with severe ventricular hypertrophy. Although P1 was born with scalp hair, 8 weeks after birth his scalp hair was completely lost. P1 died at the age of 3.5 months from severe hypertrophic cardiomyopathy.

P2 is a male, the youngest child of healthy non-consanguineous German parents. During pregnancy, fetal growth retardation and a persistent left upper *vena cava* were

diagnosed. He was born at 37 weeks of gestation by Caesarean section due to fetal bradycardia. Postnatal complications included hypertrophic cardiomyopathy, ventricular septum defect (VSD), persistent fetal circulation, and lactic acidosis, as well as thrombocytopenia and severe normochromic anemia. During early infancy, feeding difficulties, muscular hypotonia, and a moderately delayed psychomotor development were reported. Febrile infections triggered a series of more than 10 severe metabolic crises with high lactate levels of up to 15 mmol/l [N 0.5–2.2]. However, under normal conditions, serum lactate levels were within the normal reference range. Cranial MRI performed at 6 months and 5 years did not reveal any abnormality. Subsequently, the boy's condition stabilized, and he was able to walk independently at 23 months, with language and cognitive development adequate for his age. Both VSD and persistent fetal circulation resolved spontaneously in the first year of life while the hypertrophic cardiomyopathy remained stable. Normocytic anemia and thrombocytopenia are persistent, but do not require therapy. Measurements of cultured skin fibroblasts revealed isolated CIII deficiency. At birth, P2 had very fine and curly hair of the scalp, which he lost entirely during early infancy. Since then, he has total alopecia of the scalp with very fine and sparse hair growing only to be lost again. Yet, his eyelashes and eyebrows are present. At 5 years of age, bilateral papilledema was diagnosed despite normal cerebrospinal fluid pressure. After initiation of CoQ₁₀ supplementation when he was 6 years old, the boy has remained in a good health for the last 3 years, with better exercise tolerance. He displays slightly impaired growth and fine motor skills. His general muscle strength is impaired, but his walking ability overall is normal.

Table 7. Genotype and clinical presentations of patients with biallelic *UQCRFSI* variants.

	Patient 1 (P1)	Patient 2 (P2)
Gender	male	male
Family history	no	no
Consanguinity	yes	no
Age at onset	congenital	congenital
Age at last assessment	3.5 mo	8 y
Age at death	3.5 mo	
<i>UQCRFSI</i> variants	c.215-1G>C (p.Val72-Thr81del10) c.215-1G>C (p.Val72-Thr81del10)	c.41T>A (p.Val14Asp) c.610C>T (p.Arg204*)
Fetal development		
Intrauterine growth retardation		+
low birth weight	59 th percentile	+ (3 rd percentile)
fetal bradycardia		+
Perinatal development		
premature birth		
persistent fetal circulation		+
hypothermia	+	
feeding difficulties	+	+
hyperventilation	+	
Metabolism		
lactic acidosis [highest level]	24 mmol/l	15 mmol/l
metabolic crises during febrile infections		+
Cardiovascular system		
hypertrophic cardiomyopathy	+	+
ventricular septal defect		+
persistent left superior <i>vena cava</i>		+
pericardial effusion	+	
Motor system		
muscular hypotonia		+
muscular weakness		+
delayed motor development		+
elevated creatine kinase levels [highest]	>5000 U/l	
Hematologic system		
thrombocytopenia		+
normochromic anemia		+
abnormality of blood coagulation	+	
Visual system		
bilateral papilledema		+
Skin and appendages		
<i>alopecia totalis</i>	+	+
Gastrointestinal system		
Cholelithiasis	+	

*Abbreviations: +, present; -, absent.

4.2.2. WES reveals biallelic variants in *UQCRRFS1* in two unrelated families

WES of P1 was performed at our institute. Initial diagnostic analysis of WES failed to identify likely pathogenic variants in genes associated with mitochondrial disease. Assuming an autosomal recessive mode of inheritance, 65 genes carrying rare potentially biallelic non-synonymous variants were observed (MAF <0.001 within gnomAD and a frequency <0.05 among our internal database of over 16,000 exomes). Among these genes was *UQCRRFS1* with a homozygous variant at the splice acceptor site c.215-1G>C (RefSeq: NM_006003.2). As *UQCRRFS1* was the only mitochondrial protein candidate gene, I considered this gene as a promising candidate. To confirm the variant and check segregation, Sanger sequencing analysis was performed in Family 1. The analysis confirmed identified variant in patient's sample in homozygous state and in each parent in heterozygous state (Figure 21).

WES of P2 has been performed by our collaborators in Charité-Universitätsmedizin Berlin. In short, the pathogenic potential of variants was assessed with MutationTaster (<http://www.mutationtaster.org/>), discovering two heterozygous *UQCRRFS1* variants: missense c.41T>A (p.Val14Asp) located in the MTS and nonsense c.610C>T (p.Arg204*). Segregation analysis via Sanger sequencing confirmed compound heterozygosity (Figure 21). Overall, all detected variants are absent from gnomAD and our in-house database, which contains >16,000 exome datasets of individuals with unrelated phenotypes.

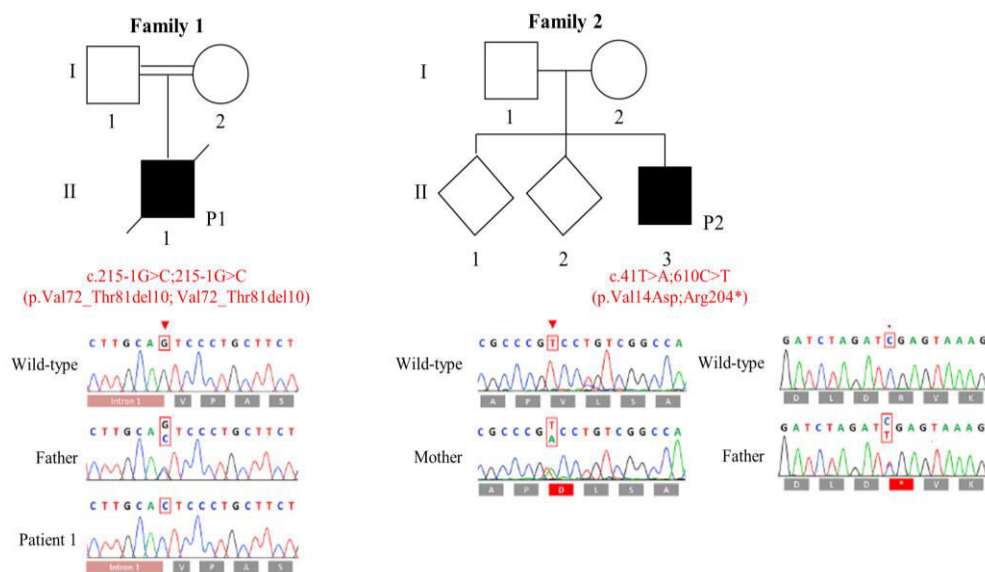


Figure 21. *UQCRRFS1* variants segregate in two unrelated families. Pedigree of two affected families, with Sanger sequencing analysis below, showing identified variants on the DNA level in patients and the segregation analysis.

4.2.3. *UQCRRFS1* in detail

UQCRRFS1 (Ubiquinol-Cytochrome C Reductase, Rieske Iron-Sulfur Polypeptide 1; NP_005994) is an iron-sulfur protein, and the last (catalytic) subunit to be incorporated to the mitochondrial complex III (Figure 22). It contains a 2Fe-2S prosthetic group that accepts a single electron from ubiquinol and transfers it to cyt c via CYC1 subunit (Iwata et al., 1998).

Its size is 274 AA and 30 kDa. Its first 78 amino acids represent a MTS. Residues 187-272 belong to the Rieske 2Fe-2S iron-sulfur domain that binds 2Fe-2S cluster per subunit. Crystal structure has been resolved as a part of respirasome (Guo et al., 2017; Figure 22). *UQCRFS1* is ubiquitously expressed with the highest expression in muscle and heart (GTEx portal).

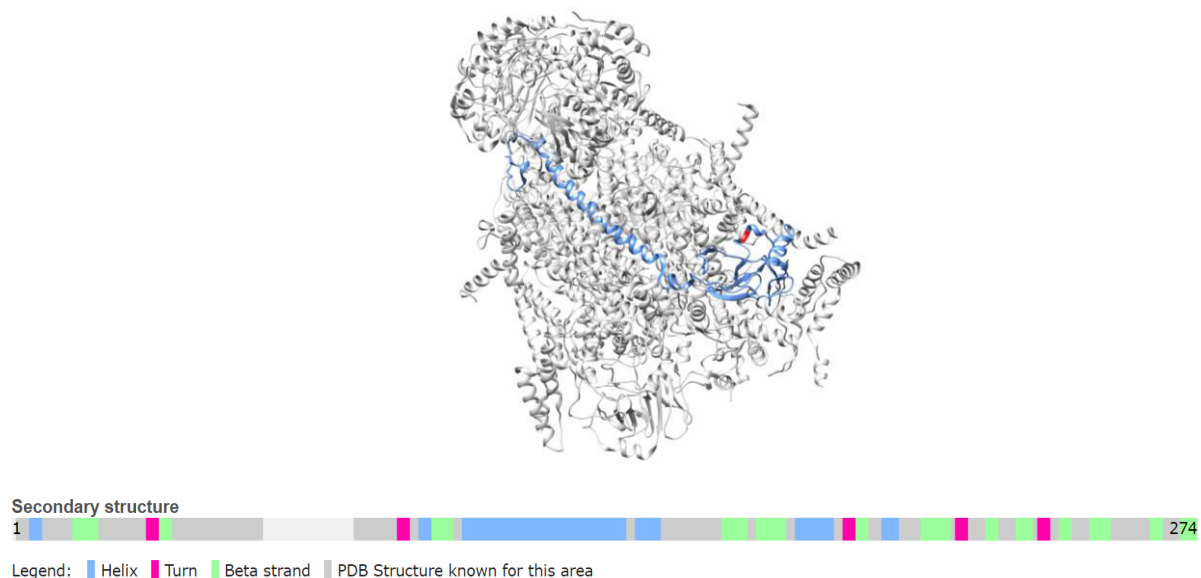


Figure 22. UQCRFS1. Upper figure: 3D representation of the crystal structure of human complex III (PDB: 5XTE), with UQCRFS1 in blue and p.Arg204 in red. Downer figure: Secondary structure of the MDH2 with the corresponding amino acid residues (from Uniprot, P47985)

A mutation that changes a conserved proline into serine in *isp-1*, the *C. elegans* homologue, has been reported to extend lifespan. However, functional evidence indicates that the mutant protein retains partial function (Feng et al., 2001; Yang and Hekimi, 2010). This very same mutation had opposing effect in mammals, as the homozygous mice were embryonically lethal. Heterozygous males had a decreased average lifespan, with a decreased activity of CIII in the heart (Hughes et al., 2011).

4.2.4. Observed *UQCRFS1* variants are predicted as deleterious

UQCRFS1 is encoded by a 2-exon, 6,282 nt long nuclear *UQCRFS1* gene located on Chr19q12 (Chr19:29205320-29213151). Overall, all three affected positions are highly conserved across vertebrates (Figure 23). The c.215-1G>C variant disrupts the direct splice acceptor site. c.41T>A changes the amino acid within MTS. Analysis of p.Val14Asp change on the TargetP-2.0 Server (Emanuelsson et al., 2007) showed that change of the non-polar valine to a negatively charged aspartic acid reduces the probability that the MTS will be correctly folded into an amphiphilic α -helix (von Heijne, 1989). c.610C>T creates a stop variant 70 AA before the reference one.

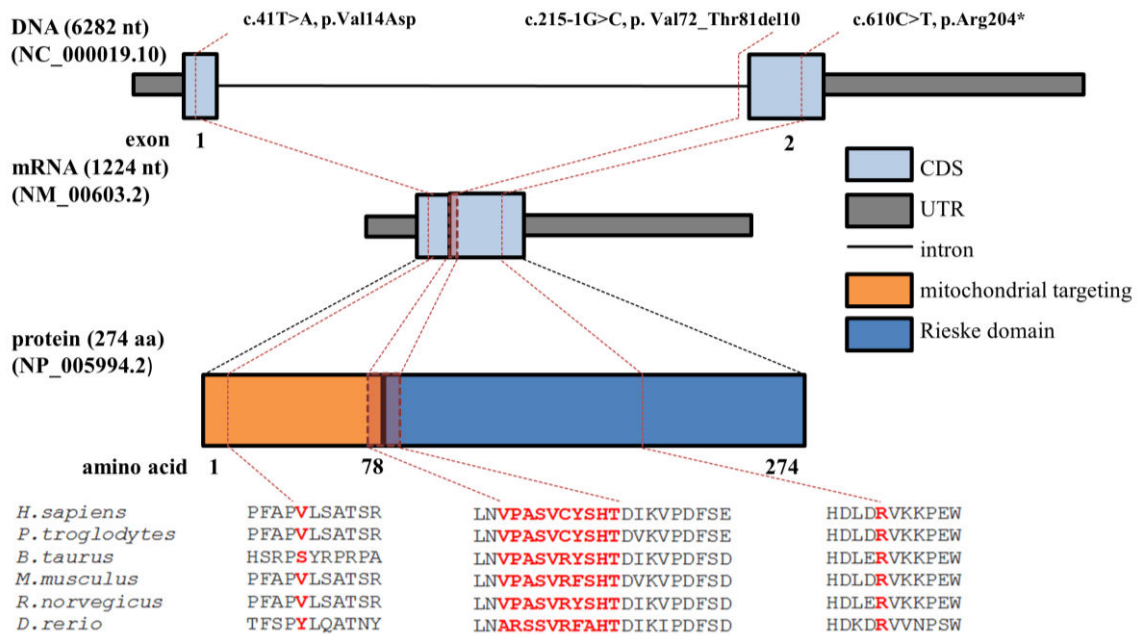


Figure 23. Effect of *UQCRFS1* variants on gene and protein. Genomic organization of *UQCRFS1* into two exons, the mRNA, and the protein structure pinpointing the positions of the identified variants. The positions of the affected amino acids are highlighted in red on the protein structure. Phylogenetic conservation of these amino acids is shown in red in the alignment of homologs across different species.

4.2.5. RNA analysis of P1 fibroblasts reveals a splice defect

To investigate the effect of the splice site variant on the transcript level, I performed splicing analysis via RT-PCR followed by Sanger sequencing, as well as total mRNA sequencing. RT-PCR revealed presence of a shorter transcript (Figure 24A). Sequence analysis confirmed a splicing defect, most likely resulting in the activation of a cryptic downstream splice site and an in-frame deletion of the first 30 nucleotides of exon 2 (Figure 24B). This deletion consequently leads to a deletion of 10 amino acids at a highly conserved region of the protein (Figure 23). Moreover, unbiased RNA-seq confirmed the RT-PCR findings, characterized by a splice defect called by Fraser, an algorithm that detects aberrant splicing, as all reads contained 30 nucleotides deletion (Figure 24C). In addition, an algorithm that detects aberrant expression, OUTRIDER, detected significantly reduced levels of *UQCRFS1*, the lowest expressed gene in P1 and across over 400 samples, with fold change (FC) of 0.46 compared to the median (Figure 24D and E).

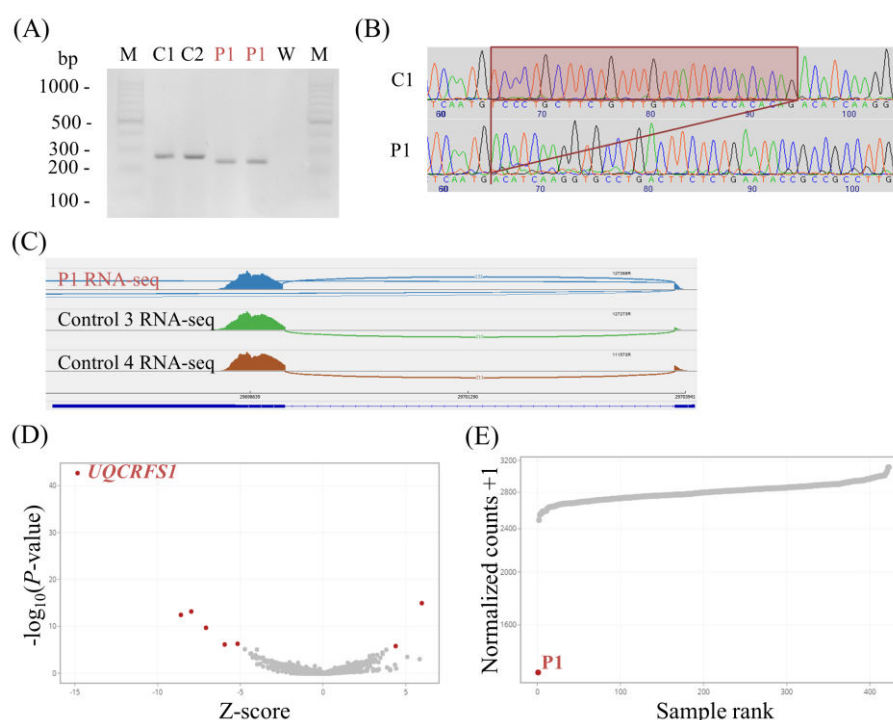


Figure 24. RNA analysis shows aberrant splicing of *UQCRFS1* in P1. (A) Splicing analysis by RT-PCR of *UQCRFS1* region between exon 1 and exon 2 by RT-PCR shows existence of a single, shorter band in patient 1 (P1) compared to random control fibroblasts (C1 and C2). (B) Sanger sequencing revealed a deletion of first 30 nucleotides of exon 2 in P1. (C) Sashimi plot of *UQCRFS1* in P1 revealed exon 2 truncation compared to the control RNA-seq samples. (D) Volcano plot of P1’s transcriptome, each dot representing a gene. Expression of each gene is shown as gene-specific level significance ($-\log_{10}(P\text{-value})$, y-axis) versus Z-score (x-axis) with *UQCRFS1* among expression outliers (red dots). (E) Sorted normalized counts (y-axis) of *UQCRFS1* across all samples, showing P1 as the sample with the lowest expression (red dot). Normalized counts taken from OUTRIDER after the autoencoder correction.

4.2.6. Patients’ fibroblasts exhibit impaired mitochondrial respiration

As both patients were suspected to suffer from a mitochondrial disease and CIII takes the central part in OXPHOS system, I analysed cellular respiration by performing microscale respirometry (Seahorse experiment). The fibroblasts cell lines were grown in glucose-rich medium and underwent mitochondrial stress test. Data (Figure 25; summarized in Figure 26C) revealed significantly decreased oxygen consumption rate capacity. Several parallel measurements revealed that the basal respiration, ATP production, and maximal respiration of the patients’ cells were significantly lower compared to the control cell line (Figure 26C).

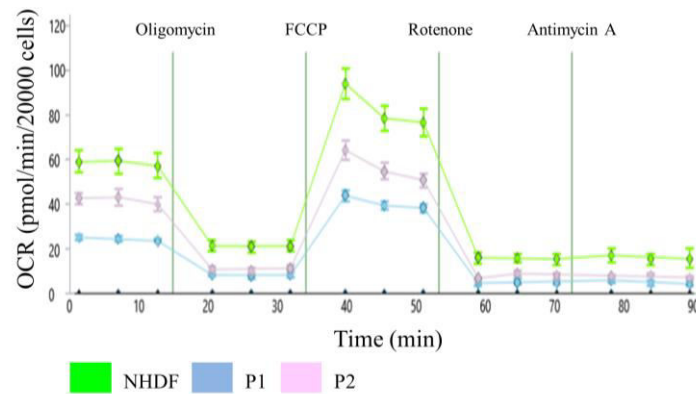


Figure 25. Microscale respirometry analysis reveals defect in mitochondrial respiration in patients' fibroblasts. Oxygen consumption rates (OCR, in pmol/min) behaviour of control cell line (NHDF), and patients' fibroblasts (P1 and P2) over time normalized to the cell count. At least 16 wells were seeded with each sample, and at least two plate replicates were performed for each patient cell line. The first three time intervals are obtained after basal conditions and the next twelve after injecting oligomycin, FCCP, rotenone, and antimycin A, respectively. Figure shows one representative experiment, with each data point representing mean \pm SEMs of measurements of technical replicates.

4.2.7. The overexpression of wild-type *UQCRFS1* in patients' fibroblasts restores mitochondrial respiration

Upon observing the defect in mitochondrial respiration, I checked whether expression of wild-type *UQCRFS1* can restore decreased parameters. To do so, I overexpressed a wild-type cDNA copy of *UQCRFS1*. The PCR product of *UQCRFS1*^{wt} was cloned into a lentiviral vector. The vector was amplified, purified and transduced into the patients and the control fibroblasts cell lines (now named P1-T-*UQCRFS1*, P2-T-*UQCRFS1*, NHDF-T-*UQCRFS1*, respectively). Upon lentiviral transduction, only NHDF-T-*UQCRFS1* and P2-T-*UQCRFS1* were growing in sufficient amounts for all experiments. P1-T-*UQCRFS1* reached a high passage number after transduction and did not replicate sufficiently enough to be included in all the experiments. As later the growth potential of these fibroblasts got exhausted, it allowed only two repetitions of the experiment with comparative low cellular yields. Lentiviral complementation rescued the respiration defect in both patients' fibroblast cell lines. Indeed, Seahorse experiments show significant improvement of basal respiration, ATP production, maximum respiration, and spare respiratory capacity in the transduced cell lines. The maximum respiration got augmented into the control range (Figure 26). The recovery of the patients' fibroblasts phenotype upon complementation suggests that the defects in *UQCRFS1* are indeed responsible for the observed phenotype.

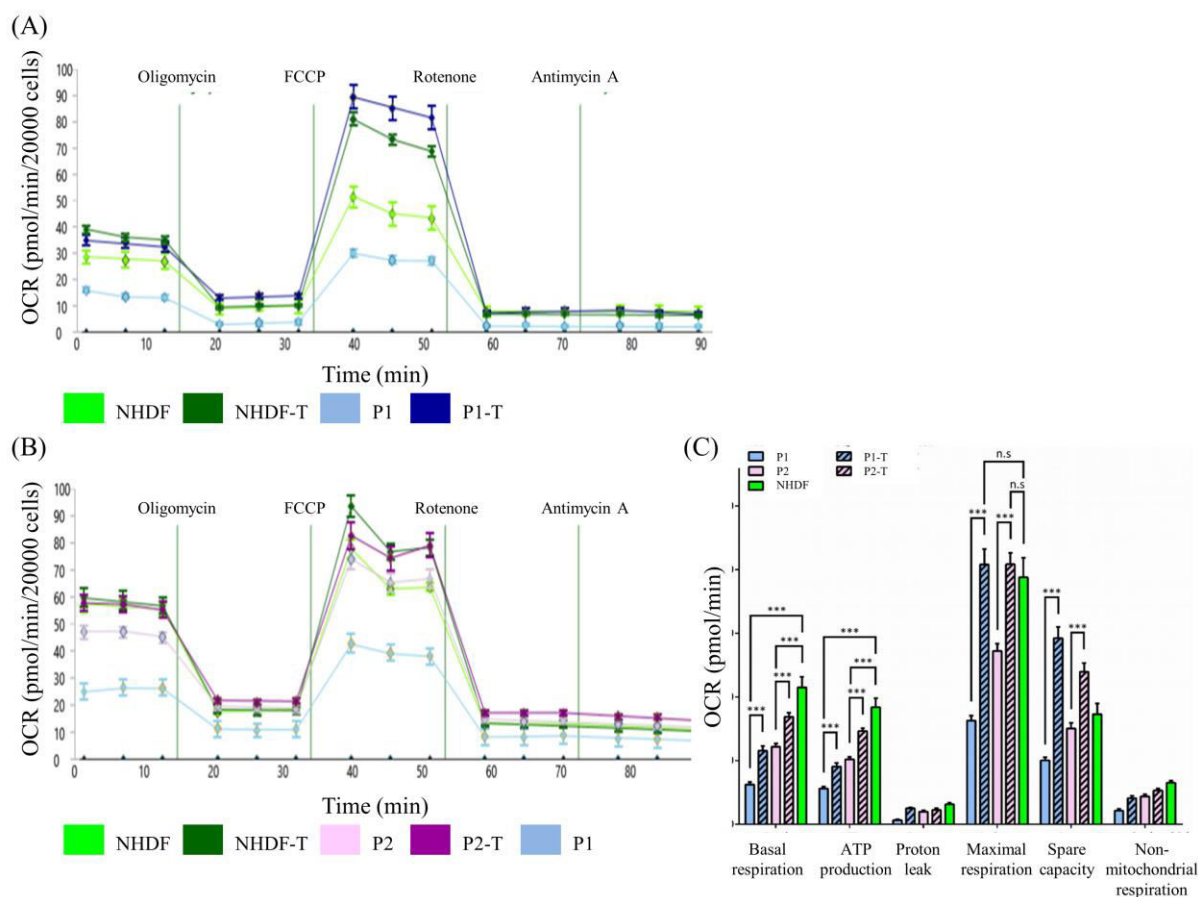


Figure 26. (A) and (B) Microscale respirometry analysis reveals rescue of mitochondrial respiration in patients' fibroblasts upon lentiviral transduction. Oxygen consumption rates (OCR, in pmol/min) behaviour of control cell line (NHDF), and patients' fibroblasts (P1 in (A), P2 in (B)) with (-T) or without lentiviral transduction over time normalized to the cell count. At least 16 wells were seeded with each sample, and at least two plate replicates were performed for each patient cell line. The first three time intervals are obtained after basal conditions and the next twelve after injecting oligomycin, FCCP, rotenone, and antimycin A, respectively. Experimental setup and number of replicates were limited due to poor growth of transduced patients' cell lines. Each figure shows one representative experiment, with each data point representing mean \pm SEMs of measurements of technical replicates. (C) Parameters of cellular respiration calculated with OCR-stats (Yepez et al., 2018). The bars depict the mean \pm SEMs from the two plate replicate experiments. Significance levels were determined with the two-tailed non-parametric Man-Whitney Utest. Abbreviations: ***, $p < 0.001$; n.s., not significant. Figure 26C adjusted from Gusic et al. (2020).

4.2.8. *UQCRFS1* variants lead to a loss of UQCRFS1

This experiment was performed by Rene Feichtinger (Salzburger Landeskliniken and Paracelsus Medical University, Austria), as described (Gusic et al., 2020). To validate the predicted damaging effect of the aberrant splicing on a protein, Western blot analysis was performed. It revealed strongly reduced UQCRFS1 levels in both patients, with no signal from P1 (Figure 27). This indicates that UQCRFS1 in P1 is below the detection limit of the antibody. In addition, Western blot confirmed that the transduction of cell lines with *UQCRFS1*^{wt} restored normal UQCRFS1 levels in P2 (Figure 27).

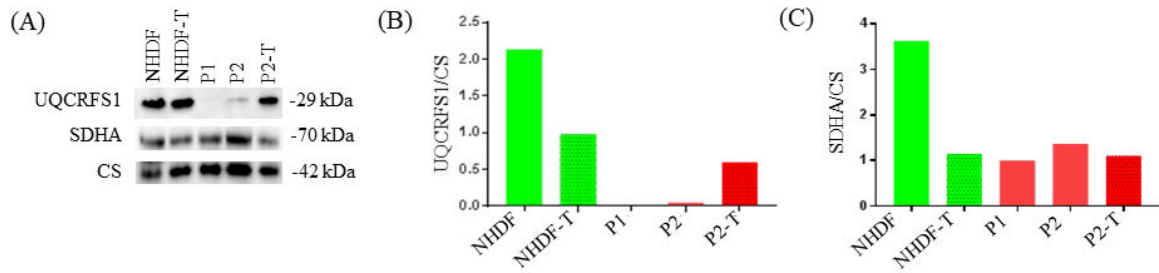


Figure 27. Depletion of UQCRC1 in patients' fibroblasts and its restoration upon lentiviral transduction in P2. (A) SDS-PAGE and subsequent Western blot analysis of fibroblast lysates showing UQCRC1, with SDHA and CS as mitochondrial loading controls. Densitometric analysis of the UQCRC1/CS (B) and SDHA/CS (C) ratios revealed strong reduction of UQCRC1 protein levels in patients' fibroblasts (P1 and P2) compared to control fibroblasts (NHDF). Lentiviral transduction rescued UQCRC1 levels in P2 cells (P2-T), and had no effect on NHDF (NHDF-T). Figure adjusted from Gusic et al. (2020).

4.2.9. UQCRC1 depletion affects complex III

This experiment was performed by Rene Feichtinger (Salzburger Landeskliniken and Paracelsus Medical University, Austria), as described (Feichtinger et al., 2017; Gusic et al., 2020). Having in mind the crucial role of UQCRC1 in CIII structural and functional maturation, we wanted to gain insight into the consequences of UQCRC1 depletion on CIII composition as a whole. The multi-protein complexes from lauryl-maltoside-solubilized mitochondria were separated in their native conformation using BN-PAGE and immunoblotting was performed using anti-UQCRC2 antibody for CIII, and anti-ATP5F1A for Complex V as a loading control. In concordance with importance of UQCRC1, assembled CIII₂ was barely detectable in both patients' samples (Figure 28). Lentiviral complementation restored the CIII assembly defect in P2 (Figure 28).

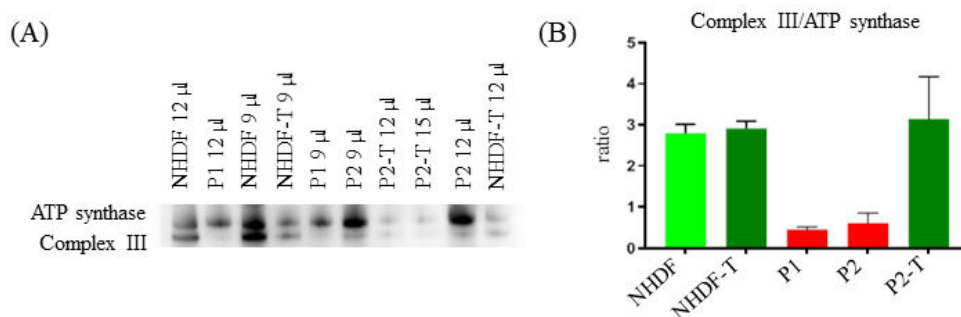


Figure 28. Disruption of CIII in patients' fibroblasts and its restoration upon lentiviral transduction in P2. (A) BN-PAGE analysis performed on digitonin solubilized mitochondria stained for ATP synthase and complex III (with ATP5F1A and UQCRC2, respectively). (B) Densitometric analysis of complex III/ATP synthase ratio in BN-PAGE. Complex III reduction is rescued by lentiviral transduction in P2 fibroblasts (P2-T). Error bars indicate the SEM. Figure adjusted from Gusic et al. (2020).

4.2.10. UQCRC1 depletion is confirmed by immunofluorescence

This experiment was performed by Rene Feichtinger (Salzburger Landeskliniken and Paracelsus Medical University, Austria), as described (Gusic et al., 2020). To investigate the

subcellular distribution of the wild-type and mutant UQCRFS1 the control and patients' fibroblasts were stained with anti-UQCRFS1 and anti-VDAC1 antibodies (Figure 29). This confirmed a strong reduction of UQCRFS1 protein in the mitochondria of both patients (Figure 29C, D, I, J). Next, UQCRFS1 restoration was observed by immunofluorescence in both patient cell lines that underwent lentiviral transduction (Figure 29G, H, M and N).

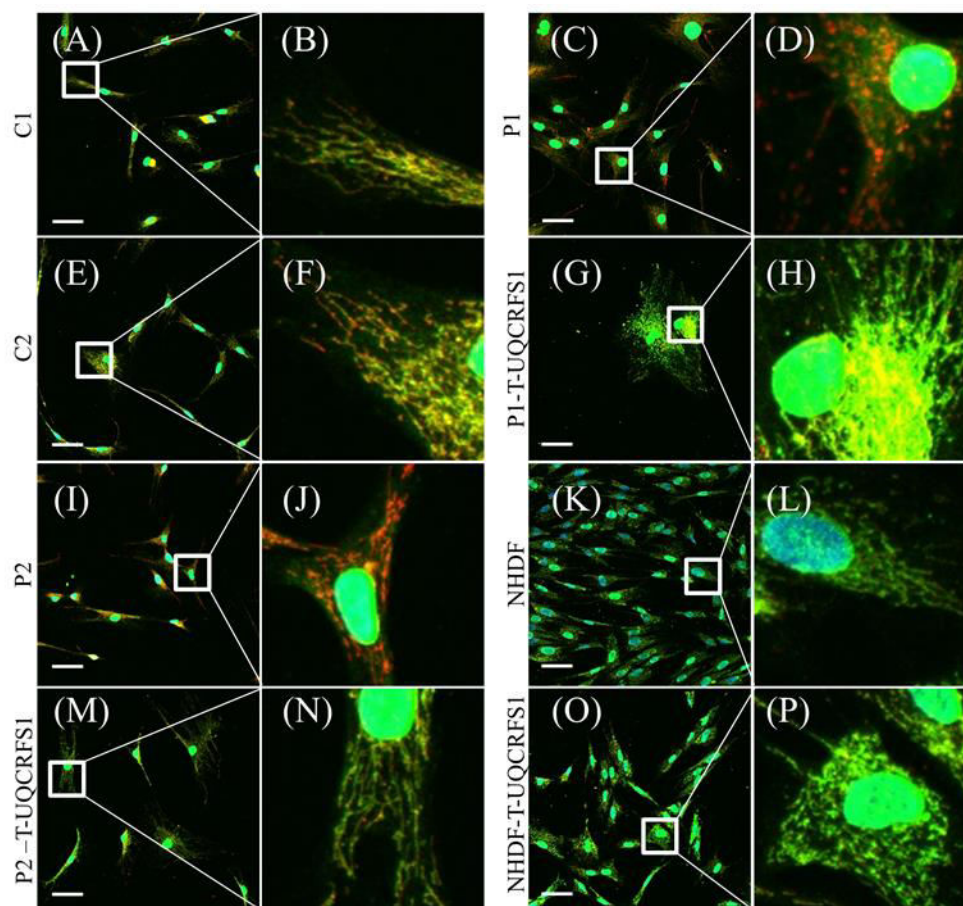


Figure 29. Immunofluorescent stainings confirm depletion of UQCRFS1 in patients' fibroblasts and its restoration upon lentiviral transduction. Merged immunofluorescent stainings of control (C1, C2, NHDF) and patients' fibroblasts (P1 and P2) with antibodies against UQCRFS1 (green) and mitochondrial marker VDAC1 (red) with (-T) or without lentiviral transduction. (A, B) C1, random control fibroblasts (C, D) P1 fibroblasts, (E, F) C2, random control fibroblasts, (G, H) P1 fibroblasts expressing an additional wild-type copy of *UQCRFS1* after lentiviral transduction, (I, J) P2 fibroblasts, (K, L) NHDF control fibroblasts, (M, N) P2 fibroblasts expressing an additional wild-type copy of *UQCRFS1* after lentiviral transduction, (O, P) NHDF cells expressing *UQCRFS1* after lentiviral transduction. The red staining in patients' fibroblasts (C, D, I, J) indicated strongly reduced UQCRFS1 levels, that are restored upon transduction of a wild-type copy of UQCRFS1 (G, H, M, N), evident by a green mitochondrial network. The area indicated with the white box was further magnified for a better visualization. Nucleus stained in green due to unspecific antibody binding. Scale bar = 50 μ m.

4.3. RNA-seq as a complementary tool in diagnostics

Submitted as “Clinical implementation of RNA sequencing for Mendelian disease diagnostics” by Yépez*, Gusic* et al. (2020).

During the implementation of RNA-seq in diagnostics in addition to WES in a large cohort, I analysed 310 RNA-seq samples from fibroblasts of affected individuals with WES data available.


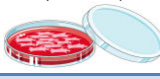
Sequencing and bioinformatics work were performed by the group of Dr. Tim Strom (Institute of Human Genetics, Technical University of Munich, Germany), including Dr. Riccardo Berutti and Dr. Thomas Schwarzmayr, together with the group of Prof. Julien Gagneur (Department of Informatics, Technical University of Munich, Garching, Germany), including Dr. Daniel Bader, Dr. Vicente Yépez, Christian Mertes, Felix Brechtmann and Nicholas Smith.

4.3.1. RNA-Seq compendium for rare disease diagnostics

In our research group we accumulated RNA-seq data from 253 blood and 462 fibroblasts samples. 103 fibroblasts-derived RNA samples were sequenced by strand-nonspecific protocol; 359 were strand-specific. Only 10 blood samples were sequenced by strand-specific protocol.

We counted genes as expressed when they are detected in at least 5% of the samples with an FPKM >1. 55% and 66% of OMIM genes and 85% and 90% of mitochondrial-disease genes are expressed in blood and fibroblasts, respectively.

Table 8. Number of expressed genes stratified by tissue and gene class from our in-house RNA-seq cohort.

Tissue of investigation	Blood (n=253) 	Fibroblasts (n=462) 
RefSeq genes (n=60,830)	13,585	14,087
Protein-coding genes (n=20,337)	10,766 (53%)	12,122 (60%)
OMIM disease genes (n=3,792)	2,092 (55%)	2,511 (66%)
Mitochondrial disease genes (n=338)	286 (85%)	305 (90%)

Fibroblasts are a good model for functional studies, but they have also proven as a better model for RNA-based studies compared to blood (Table 8). To assess the value of RNA-seq in addition to WES, in a larger cohort, I analysed 310 RNA-seq samples from fibroblasts of affected individuals for which WES data was available in-house. This cohort presents an expansion of the 105 samples from Kremer et al. (2017). All individuals were suspected to suffer from a mitochondrial disease and have undergone WES analyses beforehand, which was inconclusive for 214. All individuals and/or their guardians provided written informed consent before undergoing diagnostic and research studies, in agreement

with the Declaration of Helsinki and approved by ethical committees of the participating centres from where biological samples originated.

4.3.2. Systematic analysis of RNA-seq data

To analyse RNA-seq dataset in a rare disease diagnostic setting, the strategy was to focus on three transcript events that lead to disease-onset: aberrant expression, aberrant splicing and mono-allelic expression (MAE). Considering the genetic diversity of rare diseases, we considered that each sample is quite unique in its disease-causing variant, and thus compared it against the median of all other samples. We developed and applied the bioinformatics tools OUTRIDER (Brechtmann et al., 2018), Fraser (Mertes et al., 2021) and MAE pipeline (Kremer et al., 2017) to systematically search for rare and significant transcript events. Noteworthy, pre-processing steps, aberrant event detection tools and analysis scripts are available online and publically as part of the DROP computational workflow (Yepez et al., 2021). Such methods improvements have enabled us to reduce the RNA-seq processing time from months to hours. Application of mentioned methods revealed a median four aberrantly expressed genes, 23 aberrantly spliced gene and one aberrantly mono-allelically expressed gene (Figure 30). These 28 aberrant transcript events were my initial focus during analysis. I also manually inspected transcripts of candidate genes from WES, as well as genes well associated with patient's phenotype. When lacking promising RNA-seq candidates, I released the initial filters, such as the padj value (from 0.05 to 0.1) and Z-score. We initially assumed autosomal recessive mode of inheritance and looked for a known disease and candidate genes. Nevertheless, I also took into account that the disease-cause can be due to haploinsufficiency.

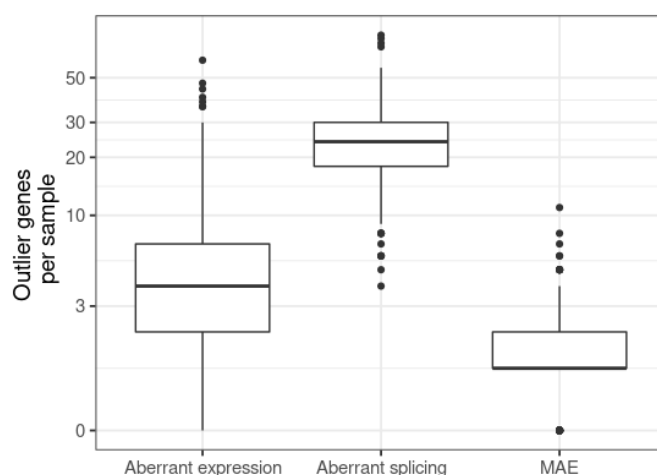


Figure 30. Median of outliers per sample per transcript event. Bioinformatics tools reveal a median of four aberrantly expressed genes, 23 aberrantly spliced genes and one mono-allelically expressed gene per sample.

For detection of aberrant expression, OUTRIDER (Brechtmann et al., 2018) was applied. It is an algorithm based on the assumption that RNA-seq read counts follow a negative binomial distribution with a gene-specific dispersion. It uses a denoising autoencoder to model read-count expectations while correcting for covariation patterns that arise due to technical, environmental, and common genetic factors. It then applies a statistical

test to detect outliers as read counts that significantly deviate from the modeled distribution ($\text{padj} < 0.05$). With such approach, I first looked at the transcriptome profile of a single case and searched for downregulated outlier genes, assuming transcript depletion as a disease-causing mechanism. Upon identification of a promising gene in inspected sample, I looked at gene expression distribution across all samples or alternatively, pre-defined subsets, in order to investigate how inspected sample is behaving compared to the cohort distribution, but also to identify the presence of additional samples with the same aberrant event. In desired scenario, gene is aberrantly expressed in just one sample. Next, I looked at the fold change of gene's expression, which informed me how much of median normalized counts is present in a sample. Fold change of less than 0.3 indicates a strong expression downregulation that could be due to defects in both alleles. If its fold change was over 0.4, I inspected the transcript on IGV in search for additional effects on transcript that could in combination with aberrant expression be causative, or checked if the gene can be disease-causing in a dominant manner.

Concerning the aberrant splicing, FRASER was applied (Mertes et al., 2021). It identifies splice sites *de novo*, by calling and defining introns by a donor (5') and acceptor sites (3'). It quantifies alternative donor and acceptor usage independently by counting spliced reads and transforming them into the percent-spliced-in (PSI, Ψ) metric. This metric detects exon skipping, exon creation and exon truncation. Fraser also counts reads overlapping the exon-intron boundaries for both donor and acceptor and transforms them into the splicing efficiency (theta, θ) metric. This metrics enables detection of exon elongation and intron retention. Like OTRIDER, Fraser uses denoising autoencoder to control for latent cofounders and uses the beta-binominal distribution to call aberrant splicing events with a statistical test ($\text{padj} < 0.05$). The beta-binominal fit then provides p values, Z-scores, and $\Delta\Psi_3$, $\Delta\Psi_5$, $\Delta\theta_3$ and $\Delta\theta_5$ that are defined as effect sizes and represent absolute difference between the expected and observed value, and thus indicate the percentage of misspliced introns. I initially focused on genes that exhibited aberrant splicing, with effect size greater than 0.3. In case of promising gene defect, I first looked for potential additional carriers of such event in a gene, but also of other aberrant splicing events in same gene. In ideal case, the inspected sample is the only carrier of such event. Then, aberrant splicing was inspected in IGV, compared to several other samples, alternative isoforms were quantified and their effect on ORF investigated.

Mono-allelic expression detection is based on counting the reads aligned to each allele at genomic positions of heterozygous variants. Using methods described in Kremer et al. (2017), MAE was called if a rare variant (maximal AF ≤ 0.001) was expressed in more than 80% of the reads. Upon detection of such MAE variants, I prioritized them as I did variants during WES analysis, with the main focus on non-synonymous ones.

In addition, our bioinformatics collaborators developed calling variants in RNA-seq, yielding a median of 44,183 variants per sample, in comparison to a median of 63,632 variants in WES (Figure 31A). Power of RNA-seq variant call compared to WES lies in intergenic- upstream and downstream, and the 3'UTR regions (Figure 31B). The enrichment in sole RNA-seq variant detection of 3'UTR compared to the 5'UTR is simply explained by their size difference: average length of 5'UTR of 200 nt compared to 800 nt of 3'UTR

(Mignone et al., 2002) makes the former more likely to still be captured by the exome capture kit. Overall, RNA-seq detects additional variants in practically every gene region, even coding one (Figure 31B), indicating that it can help overcome technical limitations in detection of certain variants by WES. For the purpose, we searched for genes that did not pass the WES filter search for biallelic variants, but in which RNA-seq led to obtaining at least two or more rare variants in total, which could be of a diagnostics potential for an autosomal recessive disease. In total 123 genes passed this filter in mitochondrial disease genes, but none proved to be disease-causing (Figure 31C).

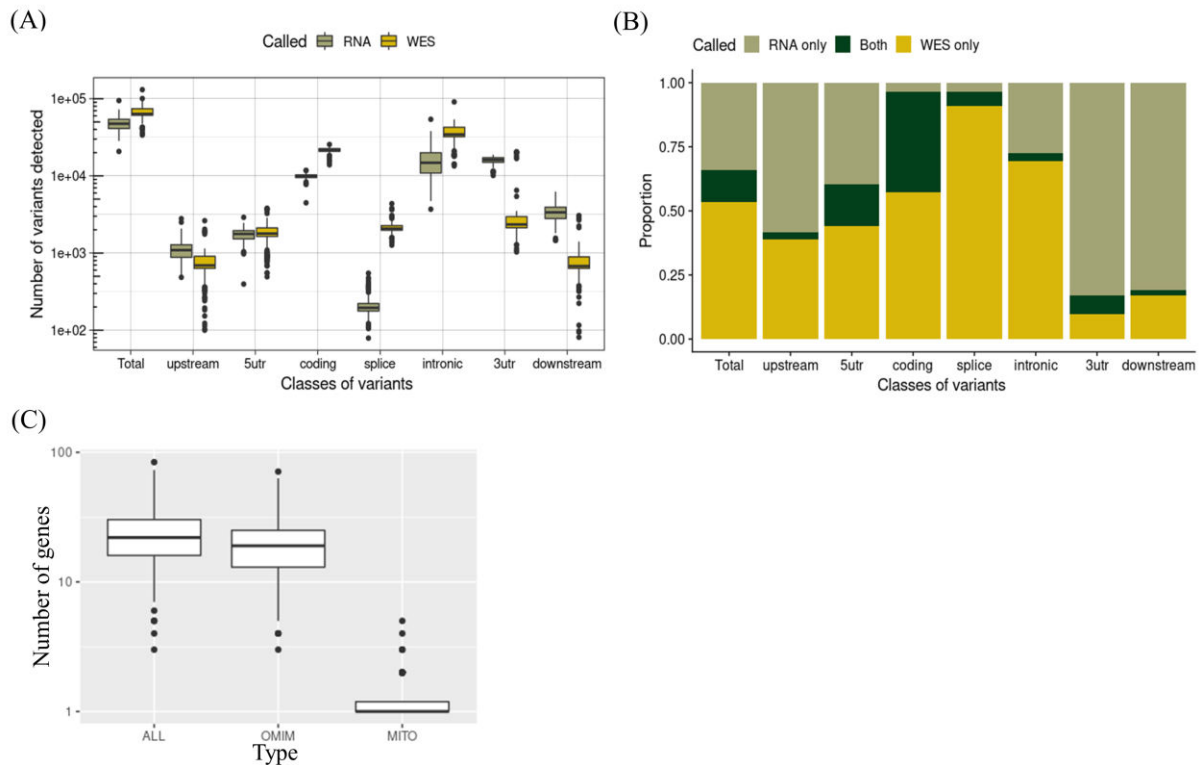


Figure 31. RNA-seq variant calling. (A) Number of variants called by WES-only and RNA-seq only, stratified by variant classes. (B) Proportion of different variant classes detected by WES-only, RNA-seq only, and in their combination. (C) Distribution of genes that pass the biallelic variant filter after integration of RNA-seq and WES variant calls, but did not pass the filter with WES-only initially.

As a rule, every candidate aberrant transcript was manually inspected on IGV. For biallelic variants, if more than 70% of the transcript was affected, or a combination with a missense variant was observed, I used WES/WGS and RNA-seq variant calling. I established genetic diagnosis upon observing a transcript defect, and associating it with rare variant(s) in the same gene.

4.3.3. Selected case studies

Main aim of implementation of RNA-seq was to increase diagnostic yield. Following cases describe how RNA-seq enabled pinpointing disease gene and establishing genetic diagnosis.

4.2.3.1. Case 1. Deletion in promoter causes aberrant expression in *UFMI*

Case 1 is a boy with neonatal-onset leukodystrophy, nystagmus, and hearing impairment. As WES analysis (WES ID: EXT_CAT_AD9741) was negative, RNA-seq was initiated on the on RNA extracted from patient's fibroblast (RNA-seq ID: EXT_CAT_AF6383). RNA-seq analysis identified 25 aberrant transcript events, among which was *UFMI* (MIM: 610553), encoding for ubiquitin-like protein whose depletion has been associated with hypomyelinating leukodystrophy. This gene was called as an expression outlier with FC expression level 0.61 compared to the median and the lowest expression among our strand-specific subcohort (Figure 32A and B). As the patient's phenotype matches the *UFMI*-associated one, I searched for a causative variant. Reanalysis of WES revealed a 3-bp homozygous deletion in the promoter region (c.-273_-271delTCA; Figure 32C). Although this variant has been reported as ClinVar pathogenic (VCV000495149) and described to significantly reduce promoter and transcriptional activity (Hamilton et al., 2017), it was overlooked during initial analysis as it is outside the protein-coding region. This case exemplifies how the detection of aberrant expression enables the reprioritization of variants located in the poorly defined genomic regions.

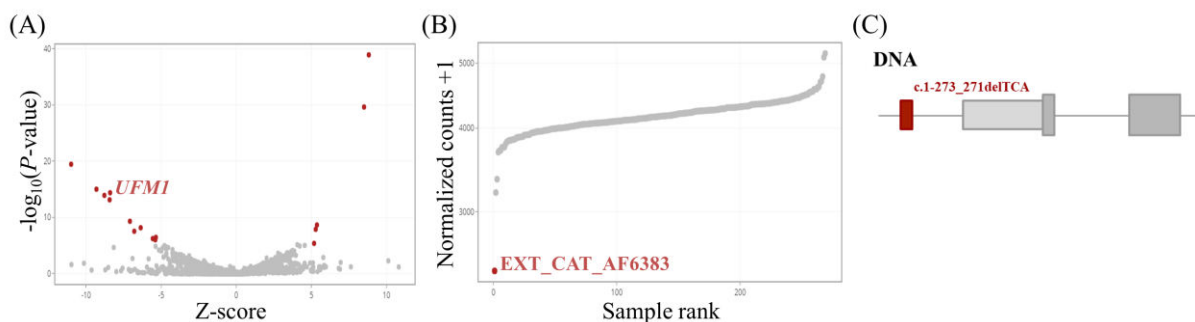


Figure 32. RNA-seq prioritizes homozygous *UFMI* variant. (A) Volcano plot of EXT_CAT_AF6383, each dot representing a gene. Gene expression is shown as gene-level significance ($-\log_{10}(P\text{-value})$, y-axis) versus Z-score, with *UFMI* among expression outliers (red dots). (B) Sorted normalized counts (y-axis) of *UFMI* across all samples in a strand-specific subcohort, with the lowest expression in EXT_CAT_AF6383 (red dot). Normalized counts taken from OTRIDER after the autoencoder correction. (C) Schematic representation of the c.-273_-271delTCA variant in *UFMI* promoter. Figure not shown at genomic scale.

4.2.3.2. Case 2: Synonymous variant causes aberrant splicing in *TWNK*

Case 2 is a male with an early-onset acute liver failure. Initial WES analysis (WES ID: 110277) remained inconclusive, upon which RNA-seq was performed (RNA-seq ID: 113015R). Among 56 aberrant transcript events RNA-seq revealed, I pinpointed *TWNK* (*C10ORF2*), encoding for the mtDNA helicase, whose mutations have been described as a cause of hepatocerebral type of mtDNA depletion syndrome (MIM: 606075). As an OTRIDER and Fraser outlier, *TWNK* FC expression level was 0.43 compared to the median across all samples (Figure 33A and B). Moreover, *TWNK* was called as a splicing outlier by Fraser (Figure 33C and D), and on IGV I observed a skipping of the 1st 62 nucleotides of exon 2 in 80% of the remaining transcript. This exon truncation causes a frameshift and stop codon (p.416Gfs*7) (Figure 33E).

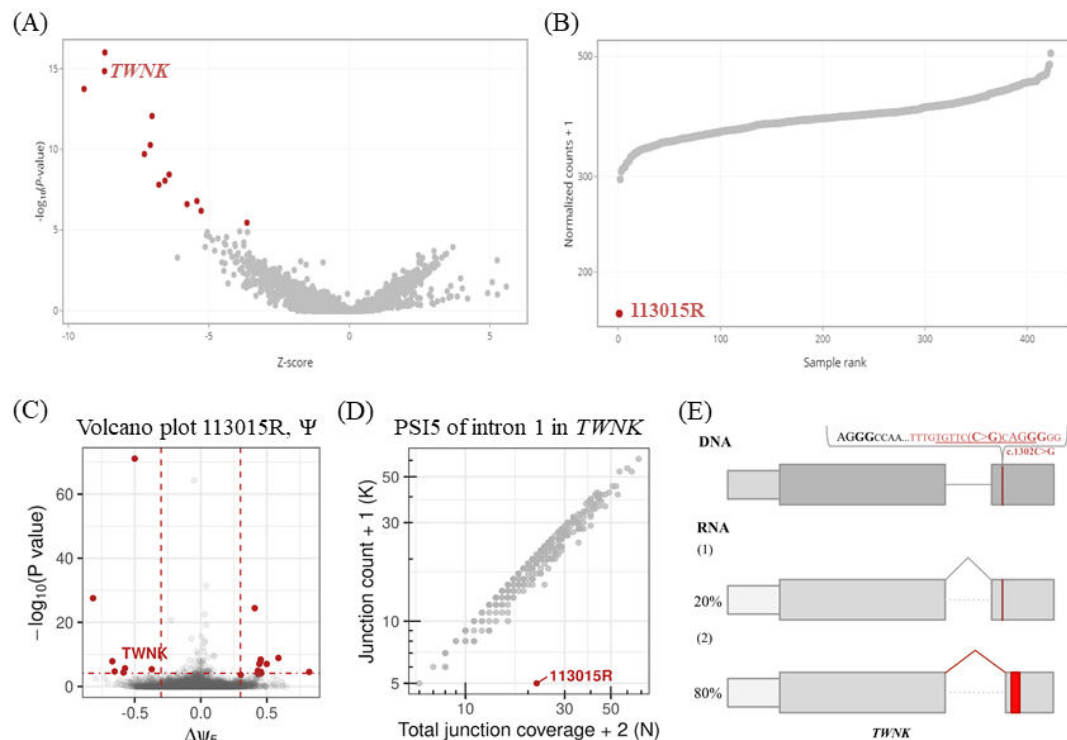


Figure 33. Synonymous *TWNK* variant affects expression and splicing (A) Volcano plot of 113015R, each dot representing a gene. Gene expression is shown as gene-level significance ($-\log_{10}(\text{P-value})$, y-axis) versus Z-score, with *TWNK* among expression outliers (red dots). (B) Sorted normalized counts (y-axis) of *TWNK* across all samples, with the lowest expression in 113015R (red dot). Normalized counts taken from OUTRIDER after the autoencoder correction. (C) Volcano plot for Ψ in 113015R, with gene-level significance ($-\log_{10}(\text{P})$, y-axis) versus effect (ψ_5 , x-axis) for the alternative splice acceptor usage in 113015R with *TWNK* among outliers (in red) (D) PSI5 value of intron 1 in *TWNK*, shown as intron split-read counts (y-axis) against the total acceptor split-read coverage for the intron 1 of *TWNK*. (E) Schematic representation of the c.1302C>G *TWNK* variant (in brackets, red) as well as its nearby region and canonical (in bold black) and cryptic splice site (in bold red), and its effect on RNA with the PTC in red. Figure not shown at genomic scale.

As the gene's function and associated clinical phenotype matched well with the patient's one, I went back to the WES to look for a causative variant(s). The reanalysis detected a homozygous rare synonymous variant in exon 2 (c.1302C>G, p.Ser434Ser) that is not prioritized during standard WES analysis as it does not affect the AA composition. Nevertheless, on the transcript level it creates a novel acceptor four bp downstream AG/GG, leading to a splice defect and consequent LoF (33E). Human Splicing Finder (<https://www.genomnis.com/access-hsf>) predicts that the variant alternates ESE *tgttcCca* and *ccCagg*. Based on the remaining transcript, I hypothesize that the variant disrupts the ESE which leads to an activation of a new acceptor. To conclude, this case highlights the role of RNA-seq in variant interpretation and impact of often overlooked variants on the transcript.

4.2.3.3. Case 3: Mono-allelic expression of *NFUI*

Case 3 is a male with a severe Leigh disease. He is currently bedridden at the age of 15 years. Complex I deficiency was measured in fibroblasts. Upon negative WES (WES ID: EXT_JAP_PT162), RNA-seq was performed on the on RNA extracted from patient's fibroblast (RNA-seq ID: 103170R), detecting 78 aberrant transcript events. Among those, *NFUI* was detected as an expression outlier, with 63% transcript left compared to the median, the lowest among all the samples (Figure 34A and B). In addition, MAE was detected, with 97% of the reads harboring a missense, predicted probably damaging variant, c.290A>G, p.Val91Ala in *NFUI* (Figure 34C). As the pathogenicity of the variant has still not been described, proteomics was performed on the fibroblasts from the patient, in collaboration with Robert Kopajtich. The results (proteome ID: P102392) show that the *NFUI* is lowest expressed among all the samples, with FC=0.13 (Figure 34D), confirming the *in silico* prediction of the variant deleteriousness.

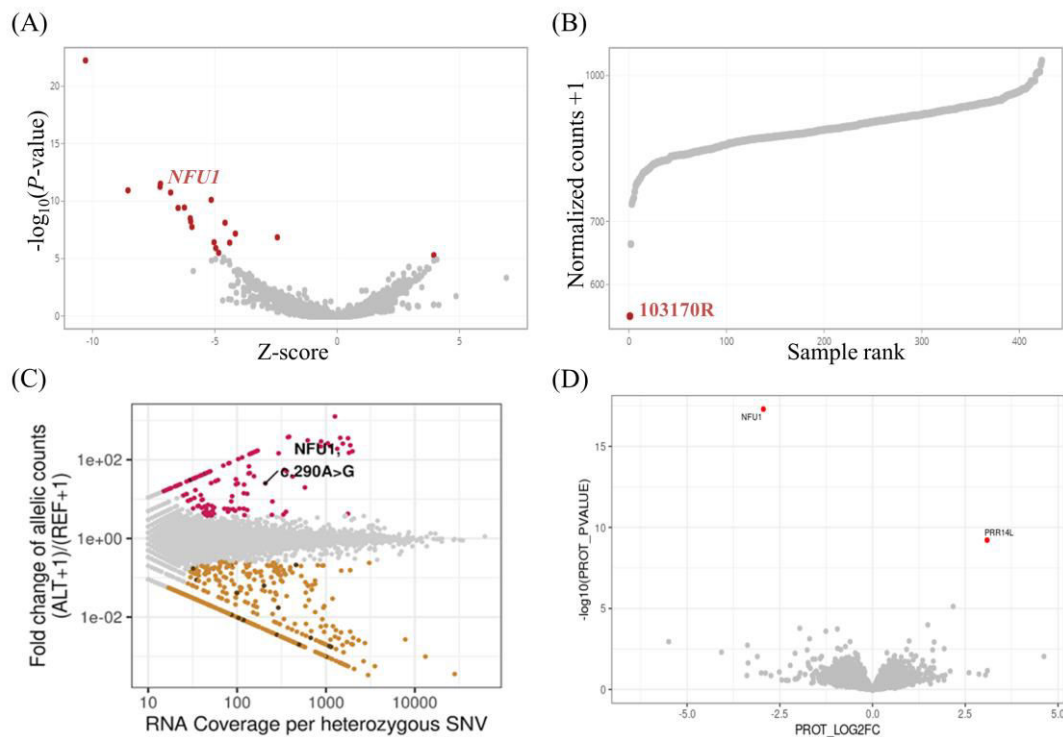


Figure 34. Mono-allelic expression of a deleterious *NFUI* variant (A) Volcano plot of 103170R, each dot representing a gene. Gene expression is shown as gene-level significance ($-\log_{10}(\text{P-value})$, y-axis) versus Z-score, with *NFUI* among expression outliers (red dots). (B) Sorted normalized counts (y-axis) of *NFUI* across all samples, with the lowest expression in 103170R (red dot). Normalized counts taken from OUTRIDER after the autoencoder correction. (E) Fold change between alternative (ALT+1) and reference (REF+1) allele read counts for 103170R compared to the total read counts per SNV within the sample. Pink represents MAE towards the alternative allele and mustard towards the reference one. In darker tones are the rare variants, including *NFUI* c.290A>G. (D) Volcano plot of P102392 showing protein levels as peptide-level significance ($-\log_{10}(\text{PROT_P-VALUE})$, y-axis) versus fold change ($\text{prot_log}_2\text{FC}$, x-axis) with *NFUI* as a protein outlier (in red).

The *NFUI* gene (MIM: 608100) encodes a protein that plays an essential role in the production of Fe-S clusters for the normal maturation of lipoate-containing 2-oxoacid

dehydrogenases and for the assembly of the mitochondrial respiratory chain complexes (Cameron et al., 2011). As Fe-S clusters defects are known to cause Leigh syndrome-like phenotype, the discovered defects in both RNA and proteins fit well the patient's phenotype enable establishing *NFUI* diagnosis.

Finally, the question of the 2nd variant remained that would explain the depletion of transcript coming from one allele. For the purpose, WGS was performed by our collaborators from the Chiba Children's Hospital, Chiba, Japan (head of the group Dr. Kei Murayama). It revealed a heterozygous 5 kb deletion, affecting exon 6 of *NFUI* (Figure 35). Segregation analysis confirmed the autosomal recessive mode of inheritance, identifying father and mother as carriers of heterozygous missense variant and deletion, respectively. Case 3 is a striking example of how implementation of different OMICS technologies altogether leads to a genetic diagnosis. Starting from negative WES, I detected *NFUI* as a candidate on RNA-seq, and the finding was functionally and genetically validated via proteomics and WGS, respectively.

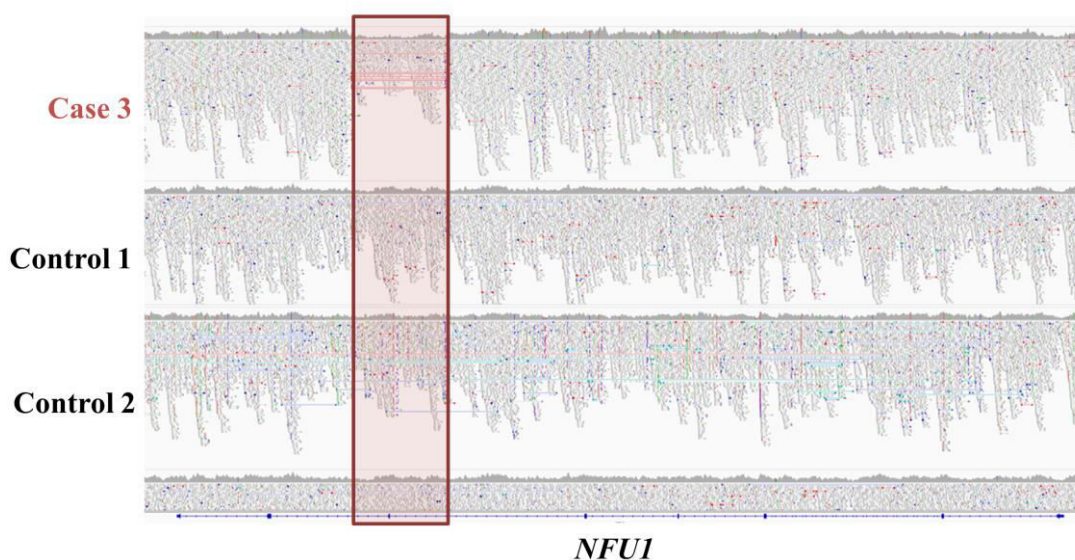


Figure 35. WGS of Case 3 identifies cause of depletion of the second allele. IGV of WGS data of the case 3 and two controls revealing a 5 kb deletion affecting *NFUI* exon 6 in Case 3, highlighted by red rectangle.

4.2.3.4. Case 4: RNA-seq variant calling pinpoints the disease-causing variant

Case 4 is a female patient from UK. She presented general deterioration, with the disease onset in infancy, failure to thrive, elevated lactate, and a CI defect. Initial WES analysis (WES ID: 69801) did not revealed any clear genes as causative ones. Yet, it did reveal a heterozygous unstart c.2T>C variant in known mitochondrial complex I deficiency disease gene, *NDUFAF5* (MIM: 612360). With the main start codon disrupted, and the next available ATG starting out-of-frame at the position c.30, this variant is a loss-of-function by consequence. RNA analysis (RNA-seq ID: 68607R) revealed 24 aberrant transcript events, with *NDUFAF5* as a splicing outlier, with 35% of reads going to a cryptic exon arising in intron 1 (Figure 36A and B). The newly included exon is 258 nt in length and contains a new acceptor site at c.223–964(A)_963(G). As this region is far from the limits of WES capture to

call a variant by WES, the variant calling on RNA-seq revealed a rare deep intronic variant c.223-907A>C. This variant is the 56th base pair into the cryptic exon and I hypothesise that it creates a new ESE site. As seen on Figure 36C, this region is elusive to WES, due to poor coverage of just 3 reads. On the protein level, with the new exon in frame, the ORF is extended with 31 amino acids before it ends in a stop codon (Figure 36C). Interestingly, this variant with the exact same functional effect was recently reported in Simon et al. (2019). It was discovered in a CI defect patient from UK via WGS and functionally validated by RT-PCR, and as well as *in silico* proposed to induce an ESE.

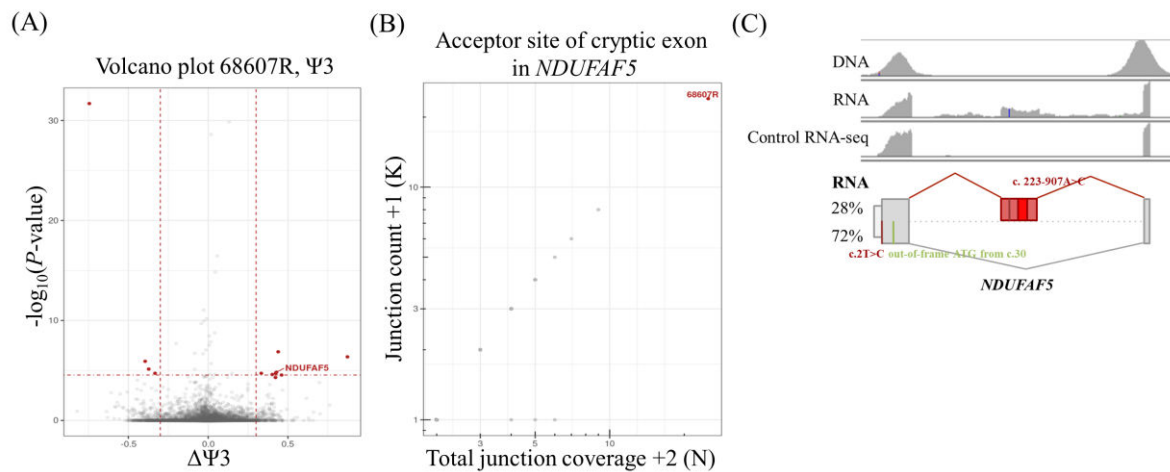


Figure 36. Deep intronic variant creates a cryptic exon in *NDUFAF5*. (A) Volcano plot of Ψ_3 in 68607R, showing gene-level significance ($-\log_{10}(P)$, y-axis) versus splicing metrics (Ψ_3 , x-axis) for the quantification of alternative splicing donor in 68607R with *NDUFAF5* among outliers (in red) (B) Values for acceptor site of cryptic exon in *NDUFAF5*, with reads spanning the exon 1- cryptic exon boundary (y-axis) against the total coverage of the cryptic exon acceptor site of *NDUFAF5*. (C) IGV pinpointing cryptic exon as well as start-loss and deep intronic variant in Case 4, with schematic depiction of variants consequence on the RNA and protein with new ATG in green, and cryptic exon and a PTC in red.

Independently, WGS performed on another CI deficiency case from UK revealed presence of a heterozygous c.223-907A>C (WGS ID: 69340G), in *trans* with previously WES-discovered (WES ID: 69340) heterozygous frameshift c.604_605insA. The RNA-seq (RNA-seq ID: MUC1375) reproduced the results from the Case 4. Altogether, the intronic variant is absent from the gnomAD and reported in 1 published case, and now in 2 more cases from our cohort.

4.2.3.5. Case 5: Complex pattern of aberrant splicing in *MRPL44*

Case 5 was suspected of multiple acyl-CoA dehydrogenation deficiency (MADD) with respiratory chain deficiency. It was a female presenting at 10 months with developmental delay, hypotonia and mild facial dysmorphic features. At that time she had normal blood spot acylcarnitines, but suberyl glycine was increased in her urine. Patient died when 13 months old with respiratory distress. Postmortem blood spot acylcarnitines showed increases in C3, C4, C5, C6, C8 and C10. Initial Sanger sequencing of the known MADD-causing genes *ETFA*, *ETFB*, *ETFDH*, *RFVT1*, *RFVT2*, *RFVT3* and *FLAD1* did not support

the suspected MADD diagnosis. WES analysis (WES ID: 96993) detected *MRPL44* encoding the mitochondrial ribosomal protein L44 (MIM: 611849), a part of the large subunit of the mitochondrial ribosome, as a candidate. The homozygous variant c.179+3A>G is located in the splice region of exon 1. The variant has not previously been associated with disease and is not reported in the gnomAD. To examine the predicted effect of variant on splicing, RNA-seq was performed on RNA extracted from patient's fibroblast (RNA-seq ID: MUC2229). Implementation of the RNA-seq analysis tools detected 25 aberrant transcript events, including *MRPL44*, as both expression and splicing outlier. *MRPL44* was the 2nd most significant downregulated gene in MUC2229 (Figure 37A), and with lowest expression among all other samples (Figure 37B), with 64% of the transcript left compared to the median. In addition, Fraser called *MRPL44* as the most significant splicing outlier. By careful inspection of the remaining transcript from the IGV, I observed a complex splicing pattern consisting of four distinct isoforms (Figure 37C and D). Isoform 1, a wild-type transcript, was present in only 18% of the reads. Isoform 2 was present in 36% of the reads and contained exon 1 without the last 164 nucleotides, leading to a frameshift and stop after 114 nucleotides (38 AA). Isoform 3 was present in 20% of the reads and had the entire intron 1 retained, leading to a stop codon after 180 nucleotides (60 AA). Finally, the isoform 4 was present in 26% of the reads and had only first 128 nucleotides of intron 1 retaining, again leading to a stop codon after 180 nucleotides (60 AA) (Figure 37C and D).

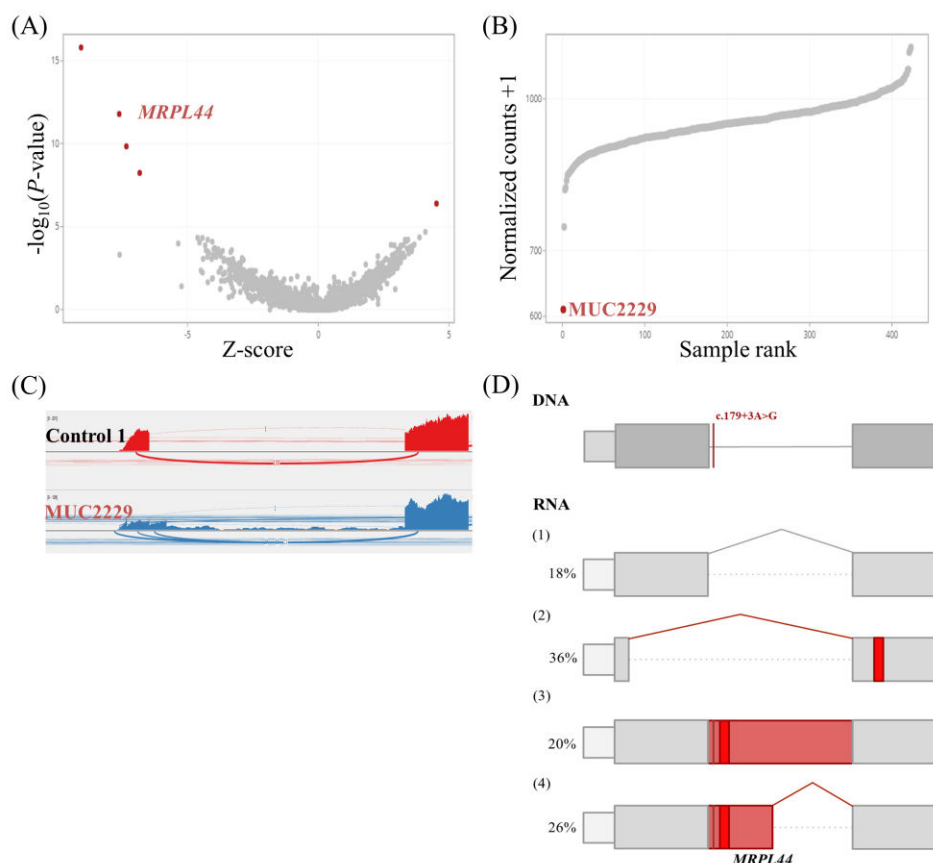


Figure 37. RNA-seq allows quantification of different transcript isoforms. (A) Volcano plot of MUC2229, each dot representing a gene. Gene expression is shown as gene-level significance ($-\log_{10}(P\text{-value})$, y-axis) versus Z-score, with *MRPL44* among expression outliers (red dots). (B) Sorted normalized counts (y-axis) of *MRPL44* across all samples, with the lowest expression in 113015R (red dot). Normalized counts taken from OUTRIDER after the autoencoder correction. (C) Sashimi plot of *MRPL44* in MUC2229 showing aberrant splicing with complex pattern compared to the control RNA-seq. Bar graph represents the coverage for each alignment track. Arcs represent splice junctions that connect exons and also display junction depth (the number of reads split across each junction). (D) Schematic depiction of the c.179+3A>G variant and its consequence on the RNA level and splicing. The quantification of splicing events reveals existence of 3 additional isoforms, all resulting in PTC (red rectangle). Figure not shown at genomic scale.

In addition, proteomic analysis of the patient's fibroblasts (proteome ID: P96993), in collaboration with Robert Kopajtich and Dmitrii Smirnov, confirmed the depletion of *MRPL44* (FC: 0.35), pinpointing it as an expression outlier. It also showed an overall decrease in mitochondrial ribosome as well as CI and CIV (Figure 38). This goes in line with the work of Yeo et al. (2015) that reported impaired mitochondrial translation and the reduction of OXPHOS assembly, as well as reports of *MRPL44*-caused combined OXPHOS deficiency (CI and IV) in three unrelated families with hypertrophic cardiomyopathy (Carroll et al., 2013; Distelmaier et al., 2015). In our case, the observed MADD-like biochemistry may have been induced by the secondary respiratory chain deficiency combined with catabolic stress. To conclude, I have found a diagnosis and expanded the phenotypic spectrum for *MRPL44*.

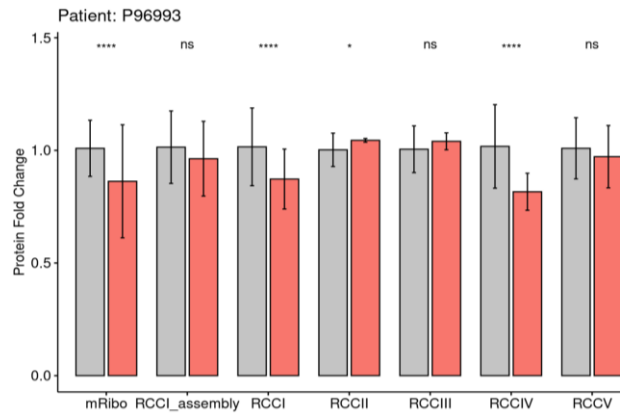


Figure 38. Defect in MRPL44 affects mitochondrial ribosome and OXPHOS. Protein fold change across mitochondrial ribosomal subunits (mRibo), CI assembly factors (RCCI_assembly) and OXPHOS subunits (RCCI-V) in P96993 in red.

4.2.3.6. Case 6: Multinucleotide variants cause aberrant splicing in *DARS2*

Case 6 is a boy with psychomotor delay, hypotonia, leukodystrophy and hyperlactacidemia, with CI deficiency measured in muscle. Upon negative WES (WES ID: 55563), RNA-seq was initiated. Among 32 aberrant transcript events (RNA-seq ID: 102950R) I observed a splicing defect in *DARS2* (MIM: 610956) (Figure 39A, B and C). Inspection on IGV revealed skipping of exon 5 in 75% of all reads (Figure 39D). As this gene, encoding for mitochondrial aspartyl-tRNA synthetase 2, is associated with similar phenotype, I searched for the causative variants. Initially I discovered only a rare heterozygous c.492+2T>C splice site variant, reported as pathogenic in ClinVar and explaining the splice defect. However, the other variant was missing, and, although not significant, FC of 0.75 of transcript compared to the median indicated that the 2nd allele causes slight depletion of transcript that is pathogenic in combination with the splice defect. Releasing the MAF filter, I discovered two additional variants in intron 2 *in cis*, c.228-12C>G and c.228-20T>C. Intron 2 has been previously reported as frequent carrier of pathogenic splice mutations (van Berge et al., 2014). These two variants, although separately not rare, are not present together in our internal database or gnomAD. Indeed, such variants sharing the same haplotype, termed multi-nucleotide variants (MNVs) are slowly but surely becoming recognised and valued during interpretation of sequencing data (Wang et al., 2020). Carefully inspecting this region, I observed exon 3 skipping in 10% of the reads (Figure 39D). Thus, I hypothesize that the combined effect of these two variants leads to a splice defect and depletion of transcript, which together with defective exon 5 leads to a disease (Figure 39D). Proteomics analysis by Robert Kopajtich and Dmitrii Smirnov (proteome ID: 102950R) confirmed RNA-seq findings, revealing *DARS2* as expression outlier, with FC of 0.26. Segregation analysis identified father as heterozygous carrier of c.492+2T>C, and mother of c.228-12C>G.

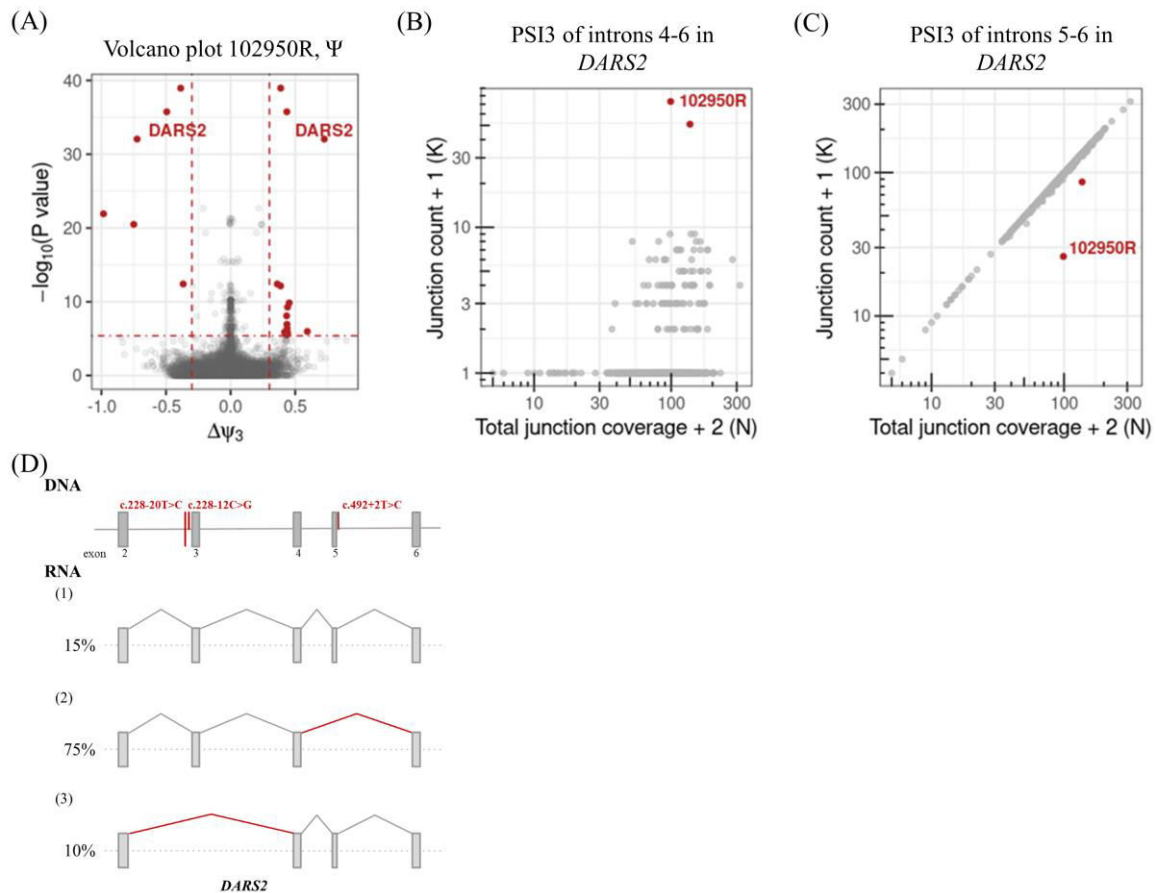


Figure 39. MNVs cause aberrant splicing in *DARS2*. (A) Volcano plot of Ψ in 102950R, with gene-level significance ($-\log_{10}(\text{P-value})$, y-axis) versus effect (ψ_3 , x-axis) for the usage of alternative donor in sample 102950R. Genes in red dots, among them pinpointed *DARS2*, passed both the genome-wide significance cut-off (horizontal red line) and the effect size cut-off (vertical red lines). (B) PSI3 of introns 4-6 in *DARS2*, shown as intron split-read counts (y-axis) against the total donor split-read coverage for the introns 4-6 of *DARS2*. (C) Same as (B), but for introns 5-6. (D) Schematic depiction of the c.228-20T>C and c.228-12C>G MNVs and c.492+2T>C variant and their consequence on the RNA level and splicing. Figure not shown at genomic scale.

4.2.3.7. Case 7: Direct splice site does not affect splicing in *BUB1*

This case presents the other side of diagnostics, where RNA-seq can lead to a dismissal of a WES-candidate gene. WES (WES ID: 97876) identified a promising candidate homozygous splice site variant c.1876+2A>G in *BUB1* (MIM: 602452). Unexpectedly, RNA-seq (RNA-seq ID: 126109R) revealed that this variant does not alter splicing (Figure 40), thus preventing false assignment of variant pathogenicity. Inspected exon-exon junction is expressed in all three known transcript isoforms of the gene, so tissue-specificity as a reason for normal splicing can be excluded. Apart from this case, within our cohort only 10% of genes with homozygous variants in the direct splice site were called by FRASER as aberrantly spliced, proving that functional validation of a predicted transcript defect is still a highly recommended step before establishing diagnosis.

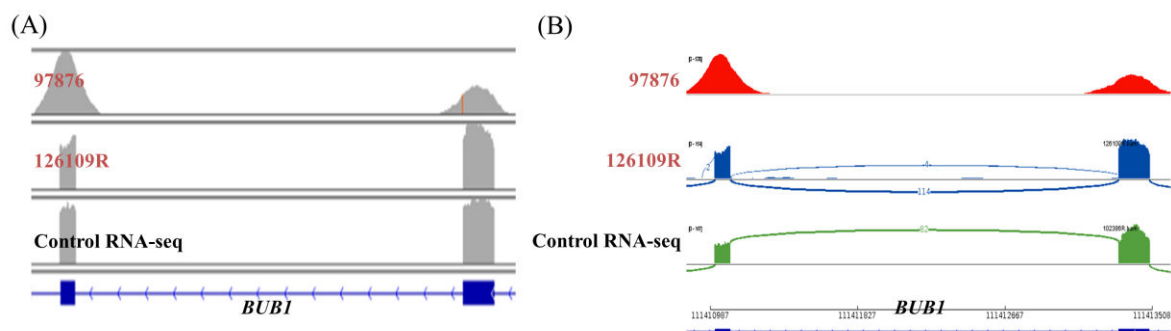


Figure 40. RNA-seq prevents false diagnosis. IGV (A) and Sashimi plot (B) showing normal splicing of *BUBI* in 126109R despite carrying a homozygous variant in the direct splice site.

4.2.3.8. Detecting dominant cases with OUTRIDER

Although we usually focus on events affecting both alleles, more than 660 disease genes are reported as haploinsufficient (Matharu et al., 2019). This is especially important in the diagnosis of neurodevelopmental disorders, where *de novo* variants are often found in the haploinsufficient genes, or regulatory elements (Heyne et al., 2018; Short et al., 2018). Three cases in our cohort- 126106R, 111888R and 99300R were diagnosed with a *de novo*, heterozygous LoF in *MEPCE* (MIM: 611478) (published in Schneeberger et al., 2019), *SON* (MIM: 182465), and *CHDI1* (MIM: 602118), respectively (Table 9 and 11). When subsetting for strand-specific RNA-seq samples, these genes were called as expression outliers with FC of 0.61, 0.64 and 0.75, respectively, although well above the median of 0.54 of underexpression outliers (Figure 41). These examples prove that our RNA-seq approach is sensitive enough to identify such “mild” expression downregulation as aberrant, and can be also used for diagnosis of autosomal dominant diseases. In addition, they show the value of subsetting the dataset to discover causative genes in diagnostics.

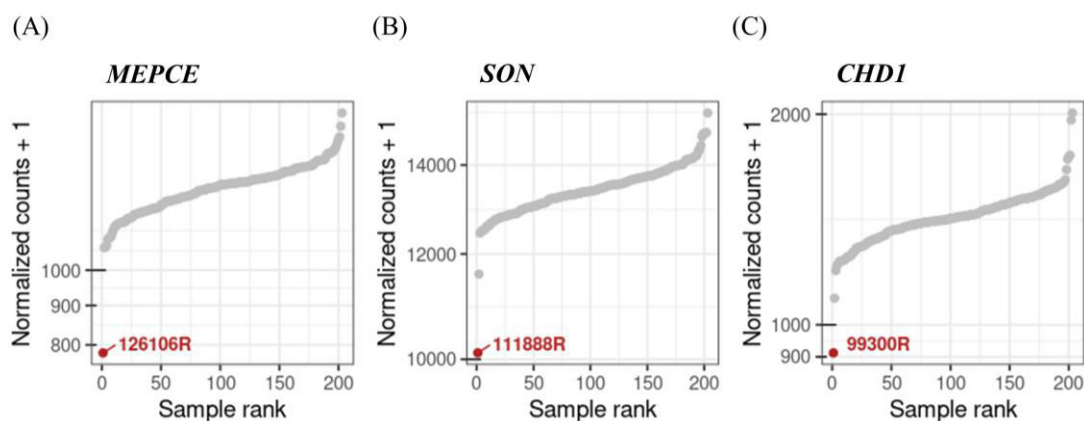


Figure 41. OUTRIDER calls dominant disease genes as outliers. Expression of *MEPCE* (A), *SON* (B) and *CHDI1* (C) shown as normalized counts ranked across all strand-specific samples, with the significant and lowest expression in 126106R, 111888R, and 99300R, respectively (red dots).

4.3.4. Summary of RNA-seq findings in diagnostic setting

As shown in Figure 42, RNA-seq detection of aberrant events and analysis provided functional evidence for 71 out of 310 cases.

Apart from 96 WES-diagnosed cases that served as controls, after RNA-seq 34 cases were diagnosed, encompassing 16% of WES-inconclusive cases. In addition, I identified promising candidate disease genes in 14 individuals (6% of WES-inconclusive) (Figure 42). I considered the sample as “new candidate” if I detected a transcript defect in a known disease gene that matches the symptoms of the individual, but without identifying causative variant(s), or in a gene not associated with a mitochondrial disease, but with identified deleterious variants. These “candidate gene” cases, although currently not solved, promise elucidation of novel mechanisms behind transcript perturbations and expansion of the disease gene spectrum, respectively. Follow-up studies on these cases are ongoing. Finally, methods recall a defect on transcript in 23 of our 96 WES-solved cases (24%). Summaries of the cases diagnosed via RNA-seq, candidates discovered via RNA-seq, and WES-diagnosed cases with a called RNA-defect are given in Tables 9, 10 and 11, respectively.

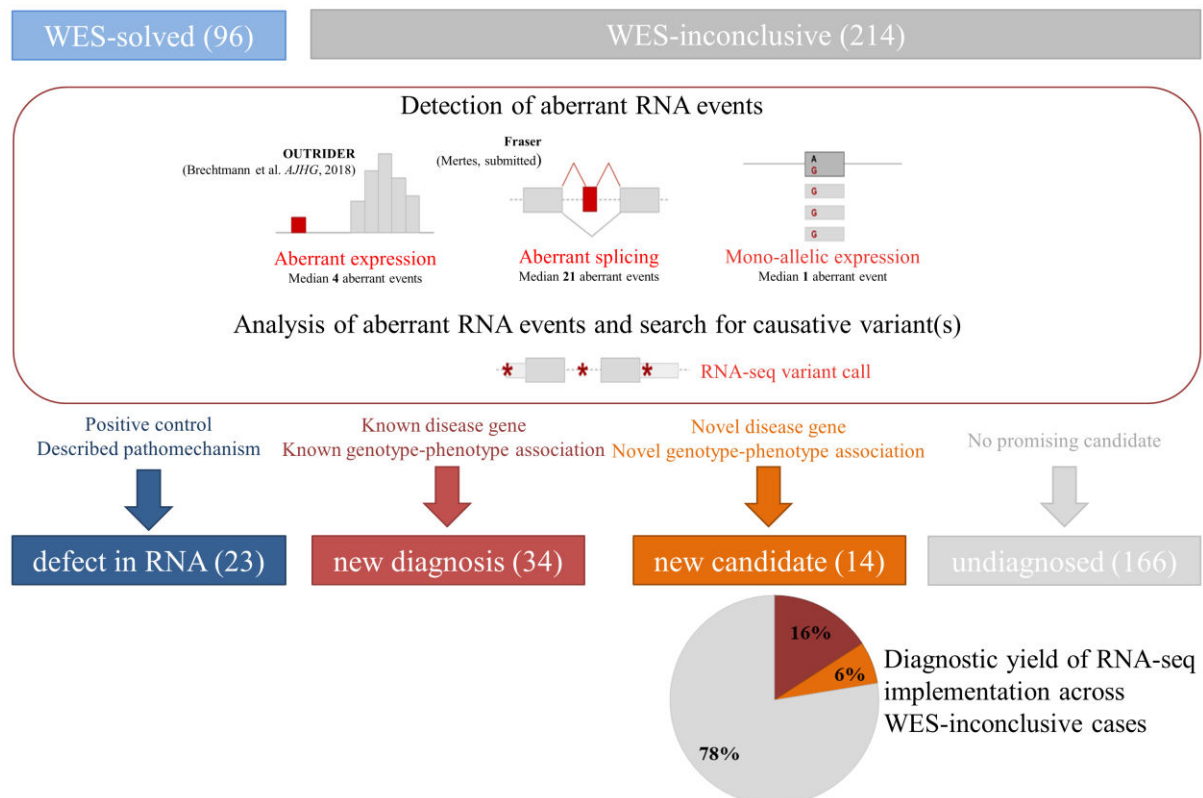


Figure 42. Summary of the RNA-seq diagnostic pipeline and diagnostic yield of the cohort.

Table 9. Summary of cases diagnosed via RNA-seq.

Index	RNA-seq ID	Gender	Primary symptoms	Genetic diagnosis	Variant	Genomic variant type	RNA defect			RNA-seq variant call
							AE	AS	MAE	
1	102631R	F	neurodevelopmental delay	<i>C19orf70</i> NM_205767.1	c.143del, p.Val48Ala ^{fs} *42 c.29+272G>C, p.Arg10Ser ^{fs} *13	frame-shift intron	x	x		x
2	102875R	F	Leigh syndrome, neurodevelopmental delay, intellectual disability, seizures, encephalopathy	<i>MRPL38</i> NM_032478.3	c.770C>G, p.Pro257Arg c.-174_-148del, p.?	missense 5utr deletion			x	
3	102950R	M	hypotonia, cardiomyopathy, white matter abnormality MRI	<i>DARS2</i> NM_018122.4	c.492+2T>C, p.Lys133_Lys164del c.228-12C>G, p.? c.228-20T>C, p.?	splice_donor intron intron			x	
4	103170R	M	Leigh syndrome, neurodevelopmental delay, speech delay, intellectual disability, encephalopathy, hypotonia, nystagmus	<i>NFUI</i> NM_001002755.2	c.362T>C, p.Val121Ala c.485-2588_545+1655del, p.?	missense deletion		x		x
5	104438R	M	myopathic facies, muscle weakness, motor developmental, neurodevelopmental delay and speech delay, intellectual disability, microcephaly, hypotonia, cardiomyopathy, dysmorphic features	<i>SLC25A4</i> NM_001151.3	c.598G>A, p.Gly200Arg c.598G>A, p.Gly200Arg	splice_region splice_region			x	
6	104982R	F	muscle weakness, myopathy, muscular dystrophy, hypotonia	<i>TIMDC1</i> NM_016589.3	c.596+2146A>G, p.Gly199ins5* c.596+2146A>G, p.Gly199ins5*	intron intron	x	x		x
7	113015R	M	acute liver failure	<i>THNK</i> NM_001163812.1	c.1302C>G, p.416G ^{fs} *7 c.1302C>G, p.416G ^{fs} *7	synonymous synonymous	x	x		
8	126112R	F	encephalopathy, respiratory distress	<i>NAXE</i> NM_144477.2	c.292-12C>G, p.? c.292-12C>G, p.?	intron intron	x	x		
9	126147R	F	acute liver failure	<i>DLI</i> NM_000108.3	c.685G>T, p.Gly229Cys unknown	missense unknown	x		x	
10	127282R	M	motor developmental and neurodevelopmental delay, respiratory distress, brainstem and white matter abnormality MRI, leukoencephalopathy	<i>MRPS25</i> NM_022497	c.329+75C>T, p.Gln110_Gln111ins* c.329+75C>T, p.Gln110_Gln111ins*	intron intron			x	x
11	127289R	M	haematological abnormality, cardiomyopathy, cardiac malformation, arrhythmias, alopecia totalis, hearing impairment, died as infant from cardiac failure	<i>UQCRFS1</i> NM_006003.2	c.215-1G>C, p.Val72_Thr81del ¹⁰ c.215-1G>C, p.Val72_Thr81del ¹⁰	splice_acceptor splice_acceptor			x	x
12	62336R	M	motor developmental delay, neurodevelopmental delay, seizures, feeding difficulties	<i>NDUFA10</i> NM_004544.3	c.-99_-75delACCCCGCCGTGACGTCACGGCAGC, p.? c.-99_-75delACCCCGCCGTGACGTCACGGCAGC, p.?	5'UTR 5'UTR			x	
13	98521R	F	MDDS, seizures, encephalopathy, hypotonia, died as neonate	<i>LIG3</i> NM_002311.4	c.86G>A, p.Trp29* c.1611+208G>A, p.?	nonsense intron		x		x
14	EXT_CAT_AF6383	M	nystagmus, hearing impairment, white matter abnormality MRI	<i>UFMI</i> NM_016617.2	c.-273_-271del, p.? c.-273_-271del, p.?	promoter promoter			x	
15	EXT_CAT_AF6392	F	Usher syndrome, immune abnormality, neutropenia, abnormality retina, cataract, visual impairment, hearing impairment	<i>PEX1</i> NM_000466.2	c.1842del, p.Gln615Lys ^{fs} *30 c.1240-1551A>G, p.Trp413_Ile414ins11*	frame-shift intronic			x	
16	MUC1344	M	myopathy, neurodevelopmental delay, hypotonia, movement disorder, failure to thrive, feeding difficulties, died as a young child due to recurrent respiratory infections	<i>TIMDC1</i> NM_016589.3	c.596+2146A>G, p.Gly199ins5* c.596+2146A>G, p.Gly199ins5*	intron intron			x	x
17	MUC1350	F	muscle weakness, neurodevelopmental delay, hypotonia, microcephaly, cardiomyopathy, hearing impairment	<i>CLPP</i> NM_006012.2	c.661G>A, p.Gln221Lys c.661G>A, p.Gln221Lys	missense missense			x	x
18	MUC1358	M	neurodevelopmental delay, feeding difficulties	<i>NDUFA10</i> NM_004544.3	c.-99_-75delACCCCGCCGTGACGTCACGGCAGC, p.? c.-99_-75delACCCCGCCGTGACGTCACGGCAGC, p.?	5'UTR 5'UTR			x	

19	MUC1361	M	ophthalmoplegia, speech delay, developmental regression, ataxia, abnormality retina, visual impairment	<i>MCOLN1</i> NM_020533.2	c.681-19A>C, p.Lys227_Leu228ins16* c.832C>T, p.Gln278*	intron nonsense	x	x	
20	MUC1365	M	died as infant, basal ganglia and brainstem abnormality MRI, neurodevelopmental delay, encephalopathy, hypotonia, myoclonus, nystagmus, abnormality eye movement, neuropathy	<i>TIMDC1</i> NM_016589.3	c.596+2146A>G, p.Gly199ins5* c.596+2146A>G, p.Gly199ins5*	intron intron		x	x
21	MUC1375	F	basal ganglia abnormality MRI, encephalopathy, brainstem abnormality MRI	<i>NDUF45</i> NM_024120.4	c.223-907A>C, p.Gln74_Val75ins19* c.604_605insA, p.Leu203Valfs*22	intron frame-shift	x		x
22	MUC1398	M	ophthalmoplegia, myopathic facies, myalgia, diabetes, arrhythmias	<i>TAZ</i> NM_181313	c.348C>T, p.Gly116_Arg123del c.348C>T, p.Gly116_Arg123del	synonymous synonymous		x	
23	MUC1404	M	muscle weakness, myopathy, neurodevelopmental delay, intellectual disability, seizures, hypotonia, dystonia, spasticity, microcephaly, failure to thrive, respiratory distress, cataract, delayed myelination, died as a child from pneumonia	<i>ALDH18A1</i> NM_001017423.1	c.1982C>A, p.Ser661 c.1858TC>T.Arg620Trp	stop missense	x		x
24	MUC2202	F	hypotonia, microcephaly, growth delay, visual impairment	<i>SFXN4</i> NM_213649.1	c.739dup, p.Arg247Lysfs*19 c.471+1G>A, p.Thr158Metfs+38	frame-shift splice_donor	x	x	x
25	MUC2203	F	basal ganglia abnormality MRI, ophthalmoplegia, ataxia, growth delay, arrhythmias, optic atrophy, visual impairment, neuropathy, white matter abnormality MRI	<i>NDUF54</i> NM_002495.2	c.466_469dup, p.Ser157* c.466_469dup, p.Ser157*	frame-shift frame-shift		x	
26	MUC2222	M	basal ganglia abnormality MRI, muscle weakness, myopathy, rhabdomyolysis, neurodevelopmental delay, seizures, infection related deterioration	<i>SLC25A42</i> NM_178526.4	c.380+2T>A, p.Gln127_Ala128ins* c.380+2T>A, p.Gln127_Ala128ins*	splice_donor splice_donor			x
27	MUC2229	F	MADD, respiratory distress, dysmorphic features	<i>MRPL44</i> NM_022915.3	c.179+3A>G, p.? c.179+3A>G, p.?	splice_region splice_region	x	x	
28	68607R	F	failure to thrive	<i>NDUF45</i> NM_024120.4	c.2T>C, p.0 c.223-907A>C, p.Gln74_Val75ins19*	start_loss intron		x	x
29	127272R	M	died as infant of heart failure	<i>MRPS30</i> NM_016640.3	c.602-468T>G, p.Asp201Valfs*14 c.602-468T>G, p.Asp201Valfs*14	intron intron	x	x	x
30	MUC1391	M	muscle weakness, myopathy, rhabdomyolysis, infection related deterioration, died as child	<i>LPIN1</i> NM_001261427.1	c.2550-865_2667-34del, p.? c.2550-865_2667-34del, p.?	deletion deletion		x	
31	126106R	M	global developmental delay, intellectual disability, seizures, hypotonia, muscle weakness, autistic features, speech delay, brain abnormalities on MRI	<i>MEPCE</i> NM_019606.6	c.1552 C> T, p.Arg518*	nonsense	x		
32	126118R	M	hypotonia, developmental delay, hearing impairment, white matter abnormality on MRI	<i>RRMB</i> NM_015713	c.328C>T, p.Arg110Cys chr8:103160286 G>A, p.?	missense intergenic		x	x
33	112223R	F	clotting defect	<i>DLD</i> NM_000108.3	c.685G>T, p.Gly229Cys c.1046+5G>T, p.Ile293_Asn349del	missense splice_region		x	
34	126763R	F	Leigh syndrome	<i>NDUF8</i> NM_005004.2	c.212+5C>T, p.? c.469-22A>C, p.?	splice_region splice_region		x	

*Abbreviations: AE, aberrant expression; AS, aberrant splicing; MAE, mono-allelic expression; M, male, F, female.

Table 10. Summary of candidate genes discovered via RNA-seq.

Index	RNA-seq ID	Gender	Phenotype match with reported ones	Candidate gene	Variant	Genomic variant type	RNA defect			RNA-seq variant call
							AE	AS	MAE	
1	MUC1396*	M		<i>MGST1</i> NM_001260511.1	not detected		x			
2	102860R	M	cardiac malformation, hypoglycaemia, respiratory distress, CIV defect	<i>MTOI1</i> NM_001123226.1	c.1712G>C, p.Arg571Thr not detected	missense	x			
3	110444R	F	Leigh syndrome, CI defect	<i>NDUFAF2</i> NM_174889.4	not detected		x			
4	110446R	M	mitochondrial cytopathy / MELAS suspected	<i>SGCE</i> NM_001099400.1	not detected		x			
5	126133R	F		<i>LETMI</i> NM_012318.2	c.2071-9G>C, p.Ile691Cysfs*3 c.2071-9G>C, p.Ile691Cysfs*3	splice_region splice_region	x	x		
6	127273R	F	Leigh syndrome	<i>NDUFAF2</i> NM_174889.4	not detected		x			
7	EXT_CAT_AE6710	M		<i>NRG1</i> NM_001160002.1	c.332T>C, p. Met111Thr c.400+5G>C, p.?	missense splice_region			x	
8	EXT_CAT_AE6712	M	congenital disorder of glycosylation	<i>ATP6API1</i> NM_00183.4	c.291-135C>T, p.? c.291-135C>T, p.?	intronic intronic	x			x x
9	EXT_CAT_AF6387	M	recurrent respiratory infections, cerebral atrophy, epileptic encephalopathy, hypotonia	<i>NAXD</i> NM_001242881.1	c.495-2A>G, p.? not detected	splice_acceptor		x		
10	MUC2208	F		<i>COQ10A</i> NM_144576.3	c.576G>A, p.? c.576G>A, p.?	splice_region splice_region	x	x		
11	MUC2213	M		<i>SFXN3</i> NM_030971.3	c.733+2T>C, p.Val232_Ala245delfs*5 c.733+2T>C, p.Val232_Ala245delfs*5	splice_donor splice_donor	x	x		
12	MUC2224	F	MADD	<i>ETFDH</i> NM_004453.2	c.687_688delAA, p.Thr230Ilefs2*	frame-shift	x			
13	103199R	M	Leigh syndrome	<i>NDUFAF2</i> NM_174889.4	not detected		x			
14	EXT_CAT_AF6384	M	Leigh syndrome	<i>PTCD3</i> NM_017952	c.1519-1G>C, p.Ser484_Ala485ins12* c.1918C>G, p.Pro640Ala	splice_acceptor missense		x		x

*Abbreviations: AE, aberrant expression; AS, aberrant splicing; MAE, mono-allelic expression; M, male, F, female.

Table 11. Summary of WES-diagnosed cases with an RNA-defect.

Index	RNA-seq ID	Gender	Genetic diagnosis	Variant	Genomic variant type	RNA defect		
						AE	AS	MAE
1	102389R	M	<i>DGUOK</i> NM_080916.2	c.353G>T, p.Arg118Leu c.143-309_168del	missense deletion		x	
2	116684	F	<i>RNF170</i> NM_030954.3	delEx4_7 delEx4_7	deletion deletion	x		x
3	126100R	F	<i>MED23</i> NM_001270521.1	c.3734dup, p.Leu1245Phefs*32 c.635T>C, p.Phe212Ser	frame-shift missense		x	
4	126101R	F	<i>ASCC1</i> NM_001198798.2	c.490-1471_747-5742del c.490-1471_747-5742del	deletion deletion		x	
5	57415	M	<i>ELAC2</i> NM_018127.6	c.1559C>T, p.Thr520 c.631C>T, p.Arg211*	missense nonsense	x		x
6	76622	M	<i>MGME1</i> NM_052865.2	c.456G>A, p.Trp152* c.456G>A, p.Trp152*	nonsense nonsense		x	
7	76636	M	<i>DNAJC3</i> NM_006260.4	c.580C>T, p.Arg194* c.580C>T, p.Arg194*	nonsense nonsense		x	
8	61695	M	<i>DNAJC3</i> NM_006260.4	c.580C>T, p.Arg194* c.580C>T, p.Arg194*	nonsense nonsense		x	
9	MUC1370	M	<i>GAMT</i> NM_000156.5	c.522G>A, p.Trp174* c.522G>A, p.Trp174*	nonsense nonsense		x	
10	76632	F	<i>NBAS</i> NM_015909.3	c.558_560del, p.Ile187del c.686dup, p.Ser230Glnfs*4	indel frame-shift		x	
11	MUC1415	M	<i>IARS</i> NM_002161.5	c.3521T>A, p.Ile1174Asn c.1252C>T, p.Arg418*	missense nonsense			x
12	MUC1399	F	<i>AP4B1</i> NM_001253852.1	c.1216C>T, p.Arg406* c.1216C>T, p.Arg406*	nonsense nonsense	x		x
13	76638	M	<i>NBAS</i> NM_015909.3	c.3164T>C, p.Leu1055Pro c.3010C>T, p.Arg1004*	missense nonsense			x
14	MUC1410	M	<i>TALDO1</i> NM_006755.1	c.345dup, p.Asp116Argfs*80 c.345dup, p.Asp116Argfs*80	frame-shift frame-shift	x		x
15	MUC1436	F	<i>TANGO2</i> NM_152906.4	c.4del, p.Cys2Aafs*35 c.56+1_57-1del, p.?	indel cnv	x		x
16	99301R	M	<i>VPS33B</i> NM_018668.3	c.1236del, p.Lys413Argfs*6 c.1236del, p.Lys413Argfs*6	frame-shift frame-shift		x	
17	99304R	F	<i>SELENON</i> NM_020451.2	c.713dup, p.Asn238Lysfs*? c.1397G>A, p.Arg466Gln	frame-shift missense		x	x
18	111888R	F	<i>SON</i> NM_032195.2	c.5753_5756del, p.Val1918Glufs*87	frame-shift		x	
19	MUC1379	M	<i>NSUN3</i> NM_022072.3	c.123-615_466+2155del, p.? c.295C>T, p.Arg99*	deletion nonsense		x	
20	MUC1348	M	<i>SLC25A10</i> NM_001270953.1	c.304A>T, p.Lys102* c.574C>T, p.Pro192Ser	nonsense missense			x
21	MUC0490	M	<i>TXN2</i> NM_012473.3	c.71G>A, p.Trp24* c.71G>A, p.Trp24*	nonsense nonsense		x	
22	99300R	M	<i>CHD1</i> NM_001270.2	c.184_187del, p.Glu63Leufs*185	frame-shift		x	
23	MUC1364	M	<i>SBDS</i> NM_012473.3	c.258+2T>C, p.Cys84Tyrfs*4 c.258+2T>C, p.Cys84Tyrfs*4	splice_donor splice_donor		x	x

*Abbreviations: AE, aberrant expression; AS, aberrant splicing; MAE, mono-allelic expression; M, male, F, female.

The majority of RNA-seq diagnosed cases (27/34) were pinpointed as expression outliers, suggesting that the downregulation of the transcript is the most frequent consequence of a pathogenic variant on transcript (Figure 43, Table 9). This is also observed in cases diagnosed via WES where we observed a defect on RNA (12/23), as well as in our RNA-seq candidates (8/14) (Figure 43, Table 10 and 11). Aberrant splicing was proven causative in 18 cases, out of which 12 were accompanied with significant expression downregulation. Upon this, we can hypothesize that the aberrant splicing generates a significant proportion and number of isoforms that disrupt the ORF, leading to the introduction of PTC and triggering transcript degradation by NMD (Popp and Maquat, 2013). Indeed, Lewis et al. (2003) computed that 35% of alternative splicing events create NMD-provoking PTCs.

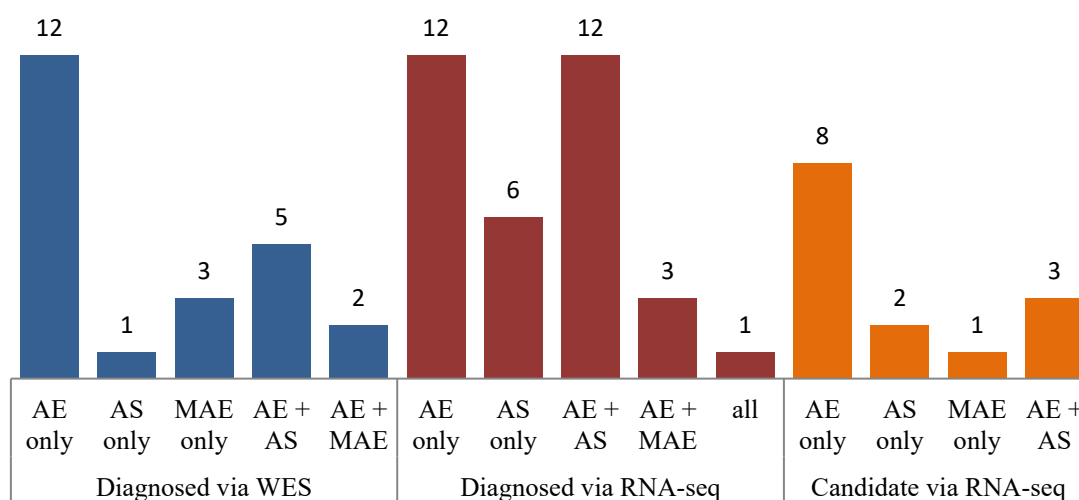


Figure 43. Disease-causing and candidate RNA-defects across the RNA-seq cohort. Overview of RNA-defects detected via three strategies that lead to confirmation of established pathogenesis (diagnosed via WES, in blue), establishing genetic diagnosis (diagnosed via RNA-seq, in red), and discovery of a new candidate (candidate via RNA-seq, in orange). Abbreviations: AE, aberrant expression; AS, aberrant splicing; MAE, mono-allelic expression.

RNA-seq diagnosed cases harbored variants that are not captured by WES (25%), prioritized during WES-analysis (25%), or would have required functional validation upon WES detection (29%) (Figure 44). With RNA-seq, they were prioritized and re-classified as pathogenic. With RNA-seq variant call, I have pinpointed in total 13 deep intronic causative variants in 9 cases. In 7 out of 9 cases, these variants led to a splice defect, more specifically the creation of a cryptic exon. In that line, variants positioned more than 100 bp away from splice sites are becoming more and more recognized as pathogenic, most frequently leading to a creation of a cryptic exon due to the activation of non-canonical splice sites (Vaz-Drago et al., 2017; Kremer et al., 2017). In two remaining cases (*LIG3* in 98521R and *PEX1* in EXT_CAT_AF6392) the detected variants lead to transcript downregulation. Although these variants get detected in WGS, only one variant (*NDUFAF2*, c.223-907A>C) was reported and could have been annotated without the functional validation (Simon et al., 2019).

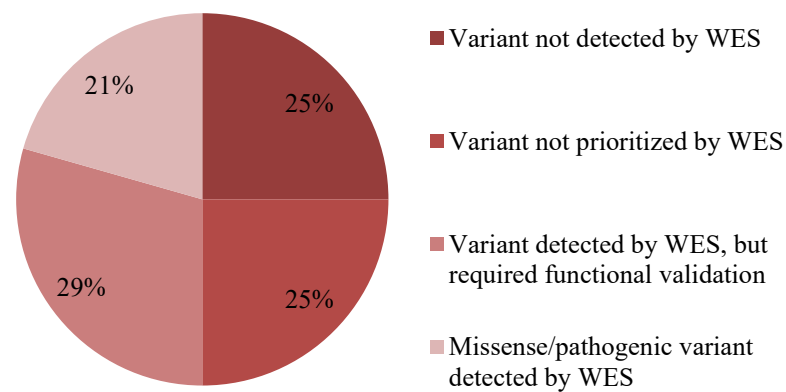


Figure 44. RNA-seq enables validation, (re)prioritization and discovery of pathogenic variants. Stratification of causative variants in cases diagnosed via RNA-seq.

Finally, we wondered if our methods are able to detect the known aberrant events known from our control cases. Out of the 96 WES-diagnosed cases, 30 contained at least 1 rare protein-truncating variant (PTV, MacArthur et al., 2012) and could thus be used as a positive control. A transcript defect was observed in 23 out of those 30 cases (Table 11). In 2 of the 7 cases where we could not find a defect in the transcript, the causal genes were not expressed in fibroblasts. Two cases solved with *TXNIP* had a very low FC (0.1 and 0.22), but the gene had a very high dispersion, which led to the cases not to be called as outliers. In rest of the cases, the PTVs are located in regions that could evade NMD. In the *GFER* case, the PTV is in the first exon, decreasing the NMD efficiency (Lindeboom et al., 2016). For *SUCO* case, one of the frame-shift variants is in the last exon. The *TASPI* case has a rare homozygous stop variant 54 bp upstream of the last exon-exon junction, close to the 50 bp rule of escaping NMD (Nagy et al., 1998). For such cases, one should also consider a possibility of false diagnosis, and carefully re-inspect them.

5. Discussion

Individual susceptibility to a disease is, at least to some degree, influenced by genetic variation. The identification and characterization of DNA variants influencing phenotype, especially in the context of onset and progression of disease, has become the main objective of human genetics. Deciphering the relationship between genetic variation and disease predisposition paves the way towards a better understanding of the underlying pathomechanism and development of treatment and prevention strategies. Over the past 25 years, technological and computational advances have enabled discoveries of pathogenic variants and the creation of vast amounts of genotype and phenotype data. Global efforts to increase the power of genetic information, such as large GWAS studies and sequencing resources, are greatly improving our understanding of genetics (Karczewski et al., 2020). This has special implications for Mendelian disorders, where revolutionary progress in novel disease gene discovery and increased diagnostic rates are witnessed.

5.1. Disease gene discovery and functional validation

Ever since the work of Douglas Wallace in the 1980s, understanding of the genetics behind the mitochondrial diseases has expanded on an admirable scale. During the writing of this thesis, four novel mitochondrial disease genes were discovered, resulting in total 338 genes associated with mitochondrial disease, and affecting a variety of mitochondrial functions (Stenton and Prokisch, 2020). Moreover, over 450 proteins with mitochondrial localization have been associated with a disease in general. In the past, novel disease gene discovery was achieved by gathering of patients with similar clinical and biochemical phenotypes, followed by candidate gene sequencing with prior linkage analysis or homozygosity mapping and mtDNA sequencing. With the implementation of NGS, the disease gene discovery rate has significantly accelerated (Figure 45). Since 2009, NGS enabled the identification of disease genes without any additional linkage or pathomechanistic information. Such a striking effect risks becoming a victim of its own success, reflected by the declining rate of novel disease gene discoveries in Mendelian disorders since 2013 (Boycott et al., 2017), indicating that most of the “low-hanging fruit” has been picked. This does not yet seem to hold true for the mitochondrial diseases, where an average of 22 novel gene discoveries are still reported per year since the onset of NGS (Frazier et al., 2019; Stenton and Prokisch, 2020).

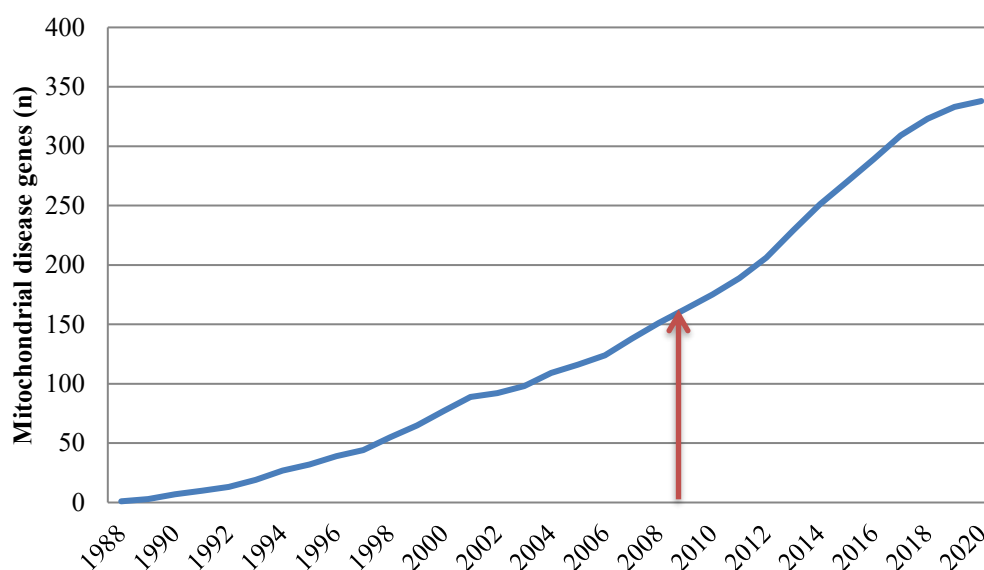


Figure 45. Number of mitochondrial disease genes over years. The red arrow indicates start of clinical implementation of NGS. Figure adjusted from Stenton and Prokisch (2020).

As shown in Figure 46, NGS has in a short time become a leading method of novel disease gene identification. Although the implementation of NGS is still hampered worldwide mostly due to lack of available funding, in many diagnostic centers it is a standard diagnostic tool and even the first-tier tool for Mendelian disorders. In addition, the development of patient registries and collaborative networks aiming to promote global communication and collaborations between clinicians and researchers such as GENOMIT (www.genomit.eu) and mitoNET (www.mitonet.org) promises a more comprehensive approach to the disease, up to a global scale. Having in mind that “just” ~450 out of about 1,500 genes encoding mitochondrial proteins are known to cause mitochondrial disease, it is doubtless that more are still awaiting recognition. In line with this, the biological role of most genes in the human genome remains unknown, with just 20% of protein-coding genes being associated with one or more phenotypes to date. Bamshad et al. (2019) calculated that 4,450 novel genes could underlie a Mendelian disease based on the constraint metrics and Mouse Genome Database phenotypes. However, the effect of loss-of-function variants in yet undescribed genes needs to be carefully assessed. Some are embryonically lethal in a heterozygous state, whereas others are benign at homozygosity, and there is a wide spectrum in between. Although new computational approaches, technologies (e.g., CRISPR screens, multiplexed assays) and experimental models offer an unprecedented opportunity to assess the variant and genome function in a high-throughput fashion, investigating natural genome variation in humans still provides the most straightforward path to establish a phenotype-genotype association (Karczewski et al., 2020). Large scale human exome and genome projects, as well as collection of diagnostic cohorts, enable cataloguing of human variants and classification of genes according to inactivation tolerance. Indeed, NGS performed in the patient cohorts or trios enabled the identification of highly deleterious variants and disease genes, while the updated gnomAD reported 1,815 genes likely to be tolerant to biallelic inactivation

(Karczewski et al., 2020). It also enables the detection of potential therapeutic targets and improves functional annotation of the human genome (Minikel et al., 2020; Karczewski et al., 2020). Finally, although NGS will inevitably lead to a plateau effect concerning single-gene discoveries in the foreseeable future, this could lead to a definite shift in the field of human genetics towards improved variant interpretation, understanding of the non-coding genome, and more complex pathomechanisms.

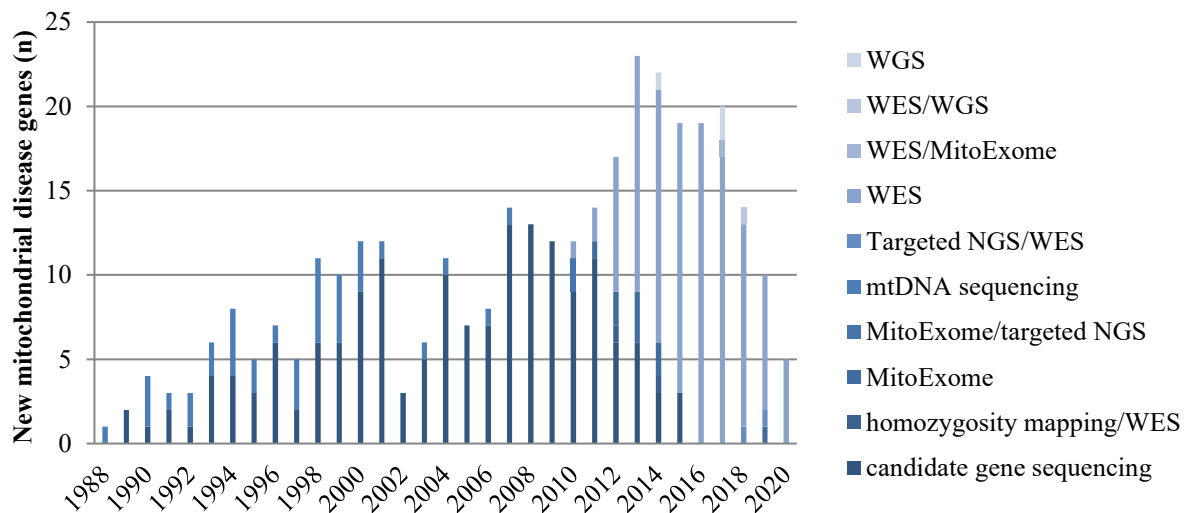


Figure 46. Yearly discovery rate of the mitochondrial-disease genes with the method of identification.

Figure adjusted from Stenton and Prokisch (2020).

The *MDH2* and *UQCRCF1* studies highlight how a “novel” gene can be associated with a disease. In both projects, I have taken advantage of a great amount of information available in our in-house database, as well as in publicly available datasets, especially gnomAD. These datasets have provided me with invaluable information on the variant frequency and the pLI score of the gene, especially important when trying to assess if the depletion of a still unreported disease gene could indeed be disease-causing. In addition, I used in silico tools such as PolyPhen-2, SIFT and CADD to estimate the deleteriousness of identified variants. Knowledge of gene’s (predicted) function obtained from the literature search is also valuable for the initial association with the observed phenotype. These lines of evidence, although not sufficient to establish the diagnosis, provide a strong indication for genes pathogenicity and urge for definite functional validation. Having access to WES and all these lines of evidence has enabled me to pinpoint *MDH2* and *UQCRCF1* as strong candidate genes in two unrelated patients. Both genes harbored rare, predicted probably deleterious variants. Bearing in mind that both patients exhibited a phenotype highly suggestive of mitochondrial disease and that both genes are involved in essential mitochondrial processes—the Krebs cycle and OXPHOS, my assumption of causality was well-grounded. Next, searching for the additional carriers of biallelic potentially damaging variants did not yield any additional cases in our in-house WES database or gnomAD, arguing for the selection against LoF. This finding goes both in my favor and against it, as with just one patient for each gene it is difficult to confidentially prove the genotype-phenotype association.

Fortunately, in a collaborative effort, I was soon connected to other researchers, via GENOMIT for *MDH2* and via mitoNET for *UQCRFS1*. This enabled identification of other unrelated individuals with a similar phenotype to my cases and biallelic variants in *MDH2* and *UQCRFS1*. Finding these cases strengthened the genotype-phenotype correlation based evidence. In addition, it highlighted the importance of cooperation and sharing findings, especially in the rare disease field. To start, I performed a segregation study by Sanger sequencing of patient and parental blood samples, to verify the variants and that their inheritance follows the Mendelian laws. This confirmed that the *MDH2* and *UQCRFS1* variants are indeed present and biallelic homozygous/compound heterozygous (Figure 14 and 21). In the *MDH2* Family 3, biallelic inheritance was a combination of a maternally inherited variant and a *de novo* variant. In the consanguineous *UQCRFS1* Family I, each parent carried the variant in a heterozygous manner. The next step was to functionally validate the variants. For this purpose, I used skin-derived patient fibroblasts. Easily clinically accessible through the non-invasive skin biopsy, although not absolutely representing the affected tissue, fibroblasts have become a standard and valuable cell line for inspecting the variants causality (Danhauser et al., 2011). The first step was to assess the effect of the variants on the gene product itself. For *MDH2*, I did so by performing a Western blot and enzymatic assay, which confirmed the absence of the *MDH2* protein and, consequently, no *MDH2* enzymatic activity (Figure 17 and 18). For *UQCRFS1*, I first performed an RT-PCR and Sanger sequencing, as well as RNA-seq, to validate the effect of the variant in P1 on the transcript. Indeed, the alteration of the direct splice site disrupted the transcript, in turn leading to its downregulation and skipping of 30 nucleotides, causing an in-frame deletion of 10 conserved amino acids (Figure 23 and 24). However, as this deletion remained in frame, it was still possible that the truncated protein was functional. To assess this, and also the effect of missense variants detected in P2 on the protein and CIII as a whole, immunostainings, Western blot, and BN-PAGE were performed (Figure 27, 28 and 29). These assays revealed a significant decrease in the levels of *UQCRFS1* and CIII, more pronounced in P1. On the biochemical level, these results demonstrated that the identified variants are indeed protein-deleterious and that the depletion of *UQCRFS1* affects the CIII. This proves that the presence of *UQCRFS1*, though the last subunit to be incorporated, is essential for the assembly and function of CIII. Next, I performed microscale respirometry to measure the mitochondrial respiration in patients' fibroblasts. This showed that loss of *MDH2* does not affect mitochondrial respiration (Figure 19). On the other hand, mitochondrial respiration was expectedly significantly affected by the depletion of CIII (Figure 25 and 26C). I used this measurable cellular phenotype as the basis for my rescue experiment, in which I overexpressed the wild-type *UQCRFS1* by lentiviral transduction. Though the transduced P1 cells did not grow well enough to assess the rescue in all experiments, I was fortunately able to see the rescue of CIII and respiration levels in both patients' cell lines (Figure 26, 27, 28 and 29). These experiments provided functional evidence for the effect of variants on the RNA and protein level, and their causality for the observed phenotype in cells. Together with the clinical and genetic information, these studies lead to the establishment of these genes as disease-causing. These observations are now included in the OMIM database, where *MDH2*

is connected to the entry #154100 and *UQCRCF1* to the entry #191327. The pathogenic variants described in both studies have been submitted to ClinVar.

5.1.1. *MDH2* as a novel mitochondrial disease gene

Considering the crucial role of the Krebs cycle in metabolism, it may be surprising to find the loss-of-function variants in its genes, although pathogenic, still compatible with life. Yet, patients with genetic defects affecting nine genes involved in the Krebs cycle have been reported. A possible explanation could be the functional splitting of the Krebs cycle in complementary mini-cycles, as suggested by Yudkoff et al. (1994). They questioned the unity of cycle and suggested that it is actually comprised of two independent cycles, allowing different fluxes, extending from α -KG to oxaloacetate and vice versa, respectively, consuming or producing glutamate and aspartate, and requiring only catalytic amounts of α -KG and oxaloacetate. As seen in Figure 47, the first mini-cycle (A) would allow conversion of pyruvate up to α -KG, even when the second mini-cycle (B) is defective (as in *MDH2* deficiency). This could explain the observed urinary excretion of α -KG in patients presenting with a defect of α -KG, SDH or fumarase activity. Likewise, it could still produce reduced equivalents in amounts that sustain OXPHOS activity. Indeed, normal oxygen uptake in circulating lymphocytes or fibroblasts from these patients was measured (Rustin et al., 1997). In agreement with this, I did not observe any defect in the oxygen consumption by the Seahorse measurements in P3 fibroblasts (Figure 19). In addition, only signs of OXPHOS deficiency or mitochondrial dysfunction were observed in fibroblasts from P3 and liver of P1, where reduced activities of CI and CV were measured, respectively (Table 6). However, this could also be associated with the well-known problems of the OXPHOS analyses, such as primary and secondary effects, tissue variation, or artefacts.

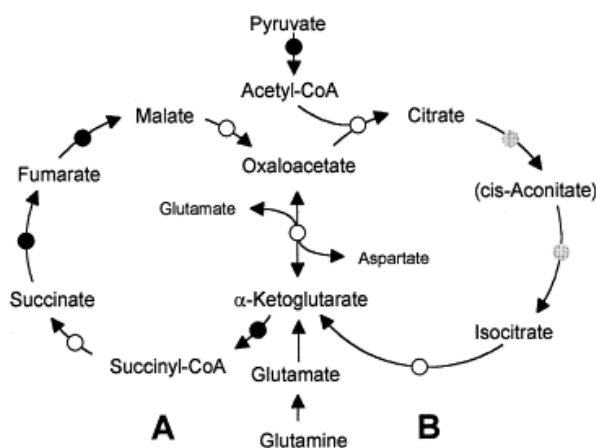


Figure 47. Proposed model for a functional splitting of the Krebs cycle into two complementary mini-cycles. Figure taken from Rustin et al. (1997).

Discussing the patients' phenotype, three described patients exhibited severe early-onset neurological presentations such as generalized hypotonia, psychomotor delay, refractory epilepsy, and elevated lactate in blood and cerebrospinal fluid. An autosomal recessive mode of inheritance, early-onset, neurological manifestations and lactic acidosis all

go in line with the previously reported Krebs cycle deficiencies. Furthermore, epileptic seizures are well associated with inborn errors of metabolism (Rahman et al., 2013). In 2017, Chen et al. provided a direct link between MDH2 and epileptogenesis, as they report MDH2 as an RNA-binding protein that binds and negatively regulates *SCN1A* in the hippocampus of a mouse model upon seizure-induced oxidation. As the downregulation of *SCN1A* is a well-known cause of epilepsy (MIM: 182389), this work proposes MDH2 as a possible effective therapeutic target for seizure control. Unfortunately, there are no curative treatments for the three patients, but a ketogenic diet trial has appeared to decrease the frequency of seizures in two children, as already reported in other cases of refractory epilepsy (Kang et al., 2007; Pasca et al., 2016). After our publication, Broeks et al. (2019) reported two patients from an extended consanguineous family with a pathogenic homozygous missense variant in the cytosolic isoenzyme of MDH (*MDH1*), presenting a neurological phenotype with global developmental delay, epilepsy and progressive microcephaly. The similar phenotype makes also sense as both MDH1 and MDH2 take part in the malate-aspartate shuttle, and goes in line with other malate-aspartate shuttle deficiencies (Broeks et al., 2019).

Identifying *MDH2* deficiency is difficult without genetic analysis. Observed phenotypes and increased lactate were highly suggestive of a mitochondrial disease, but brain MRI findings were non-specific. In all three patients levels of urinary malate and fumarate were normal or moderately increased. As observed in patients' fibroblasts, there was no increase in urinary succinate, except during an episode of metabolic decompensation in P2, highlighting the need for urinary organic acid reanalysis during such events.

Finally, disruption of the Krebs cycle can impact the pathogenesis of some cancers (Sajani et al., 2017; Anderson et al., 2018), as it can lead to abnormal accumulation of oncometabolites and promote the shift from OXPHOS to anaerobic glycolysis that favors cancer cells survival and proliferation. Dominant defects connected to tumor development have been reported in *SDHA*, *SDHB*, *FH*, *IDH1* and *IDH2* (Baysal et al., 2000; Tomlinson et al., 2002; Yan et al., 2009). *MDH2* has also been identified as a pheochromocytoma and paraganglioma susceptibility gene (Cascon et al., 2015; Calsina et al., 2018). Calsina et al. (2018) reported five heterozygous variants with potential pathogenic involvement, but incomplete penetrance, and argue that *MDH2* variants could be responsible for 0.6% of pheochromocytoma/paraganglioma cases. In Cascon et al. (2015), the single reported germline heterozygous variant was associated with loss of MDH2 activity but did not lead to malate accumulation, whereas targeted knockdown of *MDH2* in HeLa cells did lead to accumulation of malate and fumarate. Two mechanisms connect these enzymes and tumorigenesis. In the same fashion as succinate and fumarate, malate stabilizes hypoxia inducible factor 1 alpha (HIF-1 α), triggering the hypoxia response transcriptional program (Raimundo et al., 2011). HIF-1 α activation has been described to promote the transcriptional upregulation of proteins involved in pyruvate metabolism, which eventually leads to PDH complex inactivation and the pyruvate accumulation in the cytosol, and a decrease in the production of NADH and FADH₂ (Jochmanova et al., 2015). This eventually leads to cellular lactic acidosis as PDH complex, Krebs cycle, or OXPHOS are impaired. The second mechanism involves global epigenetic changes. In detail, accumulation of succinate and

fumarate, caused by SDH and FH disruptions, respectively, and the accumulation of α -KG, connected with IDH1 and IDH2 alterations, lead to an inhibition of multiple α -KG-dependent dioxygenases (Xu et al., 2011; Xiao et al., 2012). This impairs histone demethylation and 5-mC hydroxylation, thus making genome-wide histone and DNA methylation perturbations. This creates a characteristic CpG island methylator phenotype (Letouze et al., 2013). The reported *MDH2*-mutated tumor had a global transcriptional profile and a CIMP-like signature consistent with known Krebs cycle disruptions (Cascon et al., 2015). As DNA methylation changes are hallmarks of human cancers (Flavahan et al., 2017), this provides a rationale for tumor development. Altogether, these reports argue that although the detected biallelic *MDH2* variants lead to a severe neurological presentation, the heterozygous carriers could be at an elevated risk of tumorigenesis. In line with that, parents of patients with *FH* recessive disorders often develop tumors (Bayley et al., 2008). Although there is no indication of cancer susceptibility in three described families, surveillance and screening of the heterozygous family members are recommended.

In conclusion, I described biallelic variants in *MDH2* as causes of mitochondrial disease. They are loss-of-function variants and lead to a severe early-onset neurological phenotype. There are no specific clinical presentations or biomarkers to enable the rapid diagnosis of *MDH2* deficiency, so genetic analysis remains as an important tool for diagnosing additional cases, and identification of heterozygous carriers that could harbor a risk of tumorigenesis.

5.1.2. *UQCRFS1* as a novel mitochondrial disease gene

Mitochondrial diseases characterized by isolated CIII deficiency are exceptionally rare. They can be overlooked due to the general technical difficulties in measuring CIII activity (Chretien et al., 2004), or due to clinically relevant yet unremarkable or statistically insignificant reduction of CIII activity in tissues of investigation. The increasing application of NGS early in the diagnostics setting could provide a better, less biased understanding of the frequency of CIII deficiencies in the future. Indeed, with the help of WES, variants in ten genes encoding CIII subunits or assembly factors have been described as a cause of mitochondrial disease. The last subunit to be incorporated into CIII, *UQCRFS1*, is crucial for CIII maturity and enzymatic function. This argument has been confirmed in mice, as homozygous *Uqcrfs1* LoF variants are embryonically lethal (Hughes et al., 2011). Yet, despite their presumed lethality, WES analysis at our institute and in Charité-Universitätsmedizin Berlin discovered the first patients carrying biallelic potentially pathogenic variants in *UQCRFS1*. Upon these findings, a number of experiments were performed to provide the functional evidence for the variants pathogenicity.

In line with Diaz et al. (2012b), the depletion of *UQCRFS1* severely affected CIII activity and its structure on BN-PAGE. Under the conditions used, we were unfortunately unable to detect gel mobility differences between fully assembled CIII and pre-CIII by BN-PAGE. Concerning the CI/CIV deficiencies often following primary CIII deficiency, these have been traditionally explained due to their instability in the absence of their CIII partner in supercomplexes (Acin-Perez et al., 2004). However, recent work in a homoplasmic *MT-CYB*

null mutant human cell line (Protasoni et al., 2020) challenges this assumption, revealing that the loss of CIII leads to the complete loss of the supercomplexes containing CIII₂ and CI, followed by the accumulation of an inactive pre-CI lacking the catalytic N-module and the decrease of fully assembled CIV and, to a lesser extent, CII. This makes supercomplexes the platforms for the assembly of CI, CIV, and possibly CII, with CIII₂ as a central player.

Alongside the CIII deficiency, the two patients described with *UQCRFS1* pathogenic biallelic variants had fetal bradycardia, lactic acidosis, thrombocytopenia, hypertrophic cardiomyopathy, and *alopecia totalis* in common as clinical features. Cardiomyopathy occurs in 20–40% of children with mitochondrial diseases, with hypertrophic cardiomyopathy as the most frequently observed subtype (Scaglia et al., 2004; Yaplito-Lee et al., 2007). In reported CIII disease genes, hypertrophic cardiomyopathy was observed in two patients with *BCSIL* variants (Al-Owain et al., 2013), and hypertrophic, dilated, and histiocytoid cardiomyopathies have been described in three individuals with the *MT-CYB* variants (Valnot et al., 1999; Schuelke et al., 2002; Hagen et al., 2013). The reported *MT-CYB* variant m.15243G>A (p.Gly166Glu) from one of these cases (Valnot et al., 1999) disrupted the interaction between cytochrome b and UQCRFS1 in a yeast model of the disease (Fisher et al., 2004). Early-onset *alopecia totalis* is seldom associated with the mitochondrial disease. Previously, in a cohort of 140 children with mitochondrial disease, 14 were described with specific hair and skin disorders, of which three demonstrated CIII deficiency (Bodemer et al., 1999). Interestingly, *pili torti* and secondary *alopecia* are frequent hair abnormalities in *BSCIL*-caused Björnstad syndrome (Hinson et al., 2007). On top of the above-mentioned features, P2 suffers from chronic anemia, and both patients suffered from thrombocytopenia, a known symptom of mitochondrial diseases, especially in Pearson syndrome (Finsterer, 2007). Recent work in mice underlined the role of the *Uqcrrf1* and CIII in hematopoiesis and immune response. Cell-specific loss of *Uqcrrf1* in mice fetal hematopoietic stem cells impaired their differentiation, leading to anemia and prenatal death (Anso et al., 2017). Moreover, cell-specific *Uqcrrf1* knockout (KO) in mice regulatory T cells resulted in an early-onset fatal inflammatory disease without affecting their number, indicating a role of CIII in the modulation of proteins important for regulation and suppression of immune responses (Weinberg et al., 2019).

Apart from its crucial role for the CIII, several reports highlighted the additional role of UQCRFS1 in ROS production. Neuron-specific *Uqcrrf1* knockout mice presented with lower nocturnal movements and shorter life spans, dying at around 3 months of age. Neuronal death was preceded by OXPHOS disruption and early and severe ROS damage, arguing for additional pathomechanism besides ATP depletion (Diaz et al., 2012a). The same authors created *Uqcrrf1* KO in mouse lung fibroblasts, upon which they observed the reduced ROS generation, still assembled CIII pre-complex, but decreased CI, CIV, and supercomplexes assembly/stability, together with decreased CI, III and IV activity (Diaz et al., 2012b). Moreover, UQCRFS1 has been pinpointed as an important factor in ROS generation originating from the mitochondria under hypoxic conditions in the pulmonary artery smooth muscle cells (PASMCS), as its knockdown led to the abrogation of ROS generation. The authors proposed UQCRFS1 as a primary ROS generating factor during hypoxia, leading to

an increase in Ca^{2+} and eventually hypoxic pulmonary vasoconstriction, a major contributor to pulmonary hypertension (Korde et al., 2011; Ward and McMurtry, 2009). This has been confirmed both *in vitro* and *in vivo* by Waypa et al. (2013), who showed that *Uqcrrf1* (encoding for RISP) KO in mice prevents acute hypoxia-induced pulmonary hypertension (Figure 48). These studies demonstrate that RISP has an important role in pulmonary arteries, which could make it a promising therapeutic target in pulmonary vascular and heart diseases. Although the connection of ROS and UQCRFS1 was not inspected in this study, the previously published work connects CIII, especially UQCRFS1, with cardiac function whose disruption was noted in the patients I described. Altogether, this could draw attention to a more specific role of CIII in cardiac function in the future.

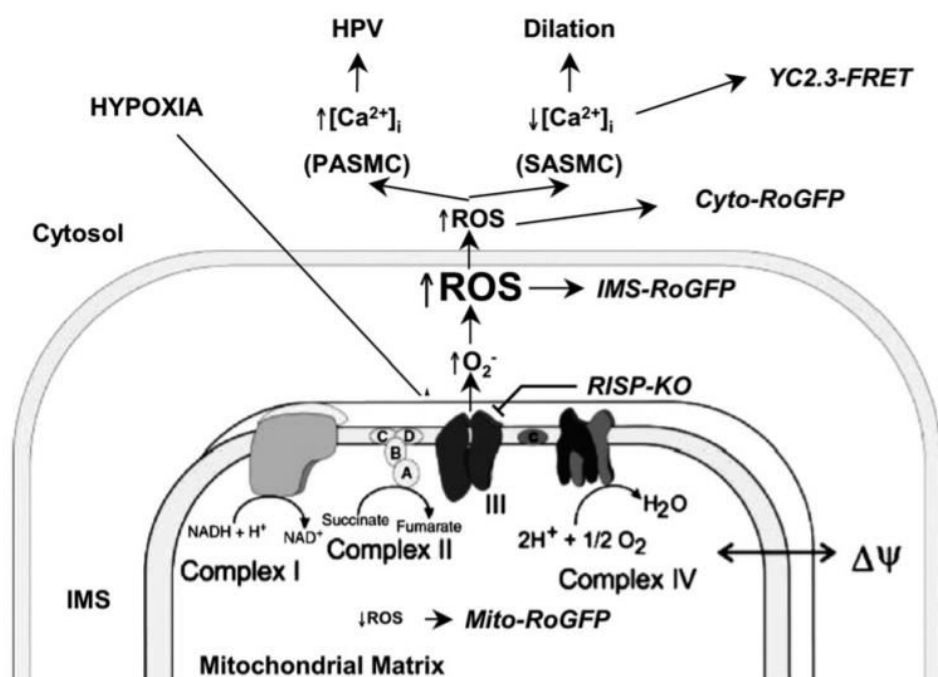


Figure 48. A model depicting CIII as the O_2 sensor underlying the HPV response. In short, CIII is required for hypoxia-induced ROS generation and consequent signaling that triggers the acute increase in cytosolic Ca^{2+} and contraction in the pulmonary circulation. Abbreviations: PAMC, pulmonary arterial smooth muscle cells; SASMC, systemic arterial smooth muscle cells. Figure taken from Waypa et al. (2013).

In conclusion, I expanded the clinical and genetic spectrum of CIII deficiencies by reporting biallelic pathogenic variants in *UQCRFS1* as a cause of mitochondrial disease in two unrelated individuals.

5.2. Genetic diagnosis of mitochondrial diseases: state-of-the-art

5.2.1. Whole exome sequencing- benefits and drawbacks

WES dominates the NGS-based diagnostics due to lower cost and easier data storage, processing, and analysis. In agreement with published cohorts, WES analysis in our institute leads to a genetic diagnosis in approximately 50% of patients with clinically suspected mitochondrial disease, thus leaving the remaining patients in a diagnostic odyssey. As discussed in Chapter 1.3.5, there are several reasons for inconclusive WES result, which can

be grouped into issues with complex pathomechanisms, variant detection, and variant prioritization (Figure 49).

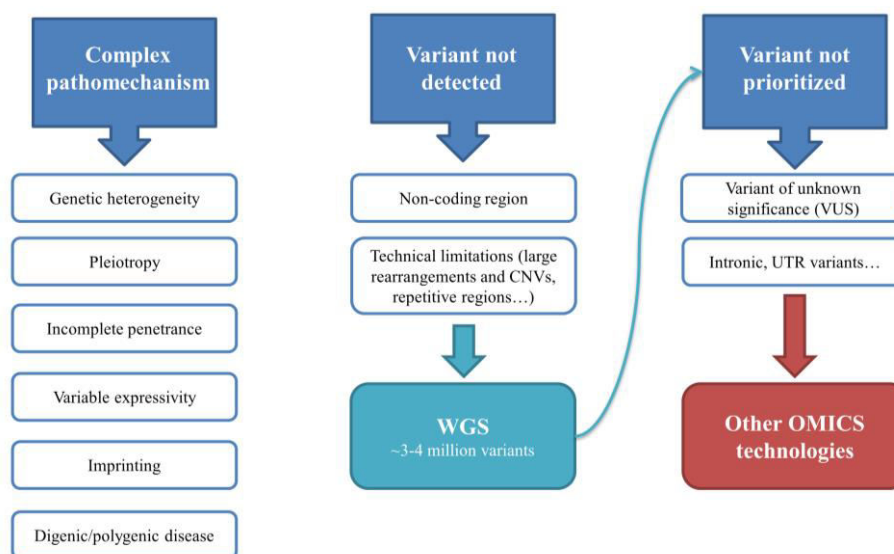


Figure 49. Summary of reasons behind inconclusive WES, with complementary approaches.

Mitochondrial diseases are amongst the most complex metabolic diseases, with a vast genetic and phenotypic variation. The lack of clear pathognomonic features and biomarkers for the majority hampers the generation of well-defined cohorts and genetic analysis. Indeed, in our cohort of suspected mitochondrial disease cases, 15-20% of cases were diagnosed with a non-mitochondrial disease. In terms of novel candidate genes, the usual focus is on mitochondria-localized proteins or proteins somewhat involved in mitochondrial function, such as *MDH2* and *UQCRC1*. However, one could hypothesize that a certain number of unsolved cases are caused by genes not encoding mitochondrial proteins. In addition, a patient may suffer from a disease with a more complex pathomechanism than initially assumed. The phenomena of genetic heterogeneity, pleiotropy, incomplete penetrance and variable expressivity complicate variant interpretation and the ability to establishing a genotype-phenotype association. For example, Ardisson et al. (2015) reported a case of reduced penetrance associated with an autosomal recessive mitochondrial disorder, where a patient with leukoencephalopathy and complex II deficiency and her healthy sister both carried a homozygous missense *SDHB* variant. Imprinting and other epigenetic factors can also underlie the disease. Finally, the suspected disease need not be monogenic at all, but caused by digenic or oligogenic factors. Such factors are difficult to interpret during the usual diagnostic analyses, and are thus seldom considered. Single example reported was an adult patient with progressive external ophthalmoplegia (PEO), where authors argue that the combined effect of *de novo ANTI* and homozygous pathogenic *POLG1* lead to more severe neurological presentations (Galassi et al., 2008).

The genetic diagnosis could remain elusive due to variant not being detected by WES. For example, pathogenic variants may reside within the non-coding regions that are per definition mostly not captured by WES. Though it is widely reported that 85% of the

pathogenic variants reside within the protein-coding regions, this estimation is likely to be biased due to the standard use of WES, and the lack of knowledge about the roles of the non-coding genome. Human genome annotation is not definite and we still do not even fully understand the coding regions. As WES enrichment kits employ baits that target only those regions already identified as exonic, it is clear that some exonic regions awaiting identification are overlooked. The WES methodology has technical limitations and biases in the detection of certain variants. Varying efficiency of bait capture and PCR amplification of problematic regions (such as GC-rich ones) result in uninform coverage. Barbitoff et al. (2020) report that most of the bias in WES comes from mapping limitations and exome probe design rather than sequence composition, affecting ~2000 exons of more than 500 genes. Thus, though the performance of WES is constantly being improved given market demand, one should still be aware that its implementation comes with certain drawbacks.

Next, WES exhibits issues in the detection of SVs. The latest gnomAD release reports that SVs account for more than 25% of all rare protein-truncating events, highlighting the need for their routine detection (Collins et al., 2020). 27,622 SVs are reportedly identified per genome (Chaisson et al., 2019). Although challenging to detect and interpret, they are well-known causes of sporadic Mendelian syndromes, and are becoming increasingly associated with complex diseases such as autism, schizophrenia, rheumatoid arthritis, and diabetes (reviewed in Weischenfeldt et al., 2013). Intragenic CNVs accounted for 7.7% of causative variants in a paediatric and rare disease cohort (Truty et al., 2019). CNV calling is mostly based on detecting changes in read coverage per position. However, as outlined earlier, the read coverage tends to fluctuate in WES data even in the absence of a CNV. Apart from ExomeDepth implemented in our analysis, several other computational tools for WES-based CNV detection have been developed, but there are still no widely accepted recommendations for their application in diagnostics (Zhao et al., 2020).

It is important to mention that in this study we did not perform trio sequencing, but opted for single exome sequencing. This subsequently disables the prioritization of *de novo* variants. Such variants can be more harmful than inherited ones as they have not been subject to natural selection (Veltman and Brunner, 2012). WGS-based studies estimate that each individual carries on average 44 to 82 *de novo* SNVs, of which 1-2 fall in a coding region (Acuna-Hidalgo et al., 2016). Damaging *de novo* SNVs and indels have been associated with both rare and common genetic disorders, most strikingly the malformation syndromes, neurodevelopmental disorders and congenital heart disease (Hoischen et al., 2010; de Ligt et al., 2012; Iossifov et al., 2014; Jin et al., 2017; Heyne et al., 2018). Trio sequencing has been shown to be substantially more effective in various patient populations than singleton one (Retterer et al., 2016; Rockowitz et al., 2020). For example, studies of 410 and 2000 patients with genetic disorders revealed a causative *de novo* variant in half of the diagnosed cases (Lee et al., 2014; Yang et al., 2014). This indicates that a considerable amount of inconclusive WES could be attributed to *de novo* mutations. Similar to the *de novo* variant detection, lack of genetic data of other family members hampers the prioritization of dominant variants. 36 mitochondrial disease genes are reported to have dominant inheritance and are frequent causes in adult-onset disease (Figure 5; Stenton and Prokisch, 2020). Thus,

we can hypothesize that yet undescribed dominant variants may be the cause of disease in some molecularly undefined adult patients. On the other side, their role in early-onset patients can be negated when the parents are healthy, in the absence of incomplete penetrance.

5.2.2. Whole genome sequencing- future challenges

As emphasized above, genetic diagnosis via WES analysis can be inconclusive due to the causative variant not being detected or prioritized. The problem of variant detection can to some degree be overcome by performing WGS, in which coverage spans the entire genome. Apart from the detection of all non-coding variants, WGS also shows an advantage in the coding regions. As it does not depend on the enrichment step and PCR, it does not show a bias in the GC content. In addition, it offers more uniform coverage, making it more suitable for CNV detection. Diagnostically relevant, comparison of the two methods revealed that WGS covered 100% of HGMD pathogenic variants, while WES did 98.22% (Meienberg et al., 2016). Although the advantages of WGS, together with the gradual reduction in the price, predict that it will replace WES in the near future, to date WGS has failed to be widely implemented in the diagnostic setting. Though WGS detects all sequence variation, which is in theory desirable, we still lack knowledge on how to interpret the vast majority of detected variants, and the main focus remains on the coding regions and CNVs. Finally, short-read WGS is only able to capture a subset of SVs (Chaisson et al., 2019). Thus, it is clear that the future steps in WGS-based diagnostics should be focused on improving the variant annotation, especially in the non-coding regions, but to also shift towards long-read sequencing to achieve maximal benefit. Indeed, an analysis of over 40 genomes from inconclusive WES cases in our institute did not help diagnose additional cases.

5.2.3. Towards the OMICS era: complementary tools to DNA-analysis

5.2.3.1. Towards the routine application of RNA-seq in diagnostics

Development of OMICS technologies, in the context of the present diagnostic gap, is slowly but surely shifting our attention beyond the one dimensional DNA analysis offered by WES and WGS to the incorporation of high-throughput functional validation. Starting with Kremer et al. and Cummings et al. (2017), RNA-seq has emerged as a complementary tool to DNA sequencing, increasing diagnostic yield, providing functional evidence for variant interpretation and re(prioritization), and finally enabling the detection of pathogenic variants invisible to WES. The strategy for RNA-seq-based diagnostic studies is a systematic search for outliers in expression, splicing and allelic balance, by comparing a single case to the rest. Several studies developed computational tools to systematically pinpoint outliers in a diagnostic setting (Brechtmann et al., 2018; Mertes et al. 2021; Mohammadi et al., 2019), assessed the importance of large control datasets (Fresard et al., 2019), and investigated the issue of adequate RNA-seq source material (Fresard et al., 2019; Gonorazky et al., 2019; Aicher et al., 2020). During my doctoral studies, our RNA-seq cohort was expanded, and I systematically analysed it with the support of the new computational tools that were implemented. This has led to a large compendium of RNA-seq data in diagnostics and a new, faster pipeline for the aberrant events detection (Figure 42). Focusing on the use of RNA-seq

in a diagnostic setting, I observed 28 aberrant transcript events per case (Figure 30). This presents a manageable number for manual investigation, similar to the average number of biallelic rare non-synonymous variants detected by WES, thus adding no significant additional workload. On average it took two hours to analyse one case in detail. The consequent analysis helped establish genetic diagnosis in 16% of WES-inconclusive cases (34 out of 214; Table 9). This number falls within the range of previously published cohorts (8-36%), but presents a significant increase over the initially reported one from our institute. The lower rate compared to Cummings et al. (2017) and Gonorazky et al. (2019) could be attributed to the criteria for inclusion into the cohort. Neuromuscular diseases, investigated by both, are much more clearly clinically defined than mitochondrial diseases. In the line with this, Fresard et al. (2019) reported a diagnostic yield of 8% in a cohort with over 16 diverse rare disease categories. Taken together, I hypothesize that precise clinical phenotyping prior to RNA-sequencing heavily influences the diagnostic yield. Returning to the original cohort, we were able not only to replicate the findings by Kremer et al. (2017), but to also establish the genetic diagnosis in additional four cases (two *NDUFAF5*, *LPINI* and *TAZ* cases; Table 9). These cases were initially left inconclusive for various reasons. The *NDUFAF5* cases were not found to be significant with the previously implemented splice detection algorithm LeafCutter. The causative variant in *LPINI* (a 1,759 bp deletion), though not called by WES, was later detected via WGS. Finally, in the *TAZ* case the causative synonymous variant c.348C>T, creating the new upstream donor site, was not indexed by ClinVar at the time of detection (Mertes et al., 2021). These examples highlight the importance of data reanalysis upon method improvement, but also regular updates on new findings, and databases entries. Noteworthy, our RNA-seq analysis pipeline is available online and publicly as part of the DROP computational workflow (Yepez et al., 2021). We hope that the reduction in time and automated steps to obtaining results using DROP will further encourage clinicians and researchers to implement RNA-seq in routine diagnostics.

Apart from helping in genetic diagnosis, RNA-seq provides functional validation of variants, leading to their (re)prioritization, or shedding light on variants in the non-coding regions as regulators of gene expression. The method aids not only the interpretation of splicing variants, but can also help to judge regulatory variants and PTVs. As summarized in Figure 50, transcriptome analysis pinpointed causative variants that ultimately lead to genetic diagnosis in 34 cases whose initial WES was inconclusive (Table 9). Although we almost exclusively focus on boosting diagnostic yield, it is equally important that RNA-seq can prevent false assignments of variant pathogenicity, especially for the variants predicted to alter splicing, exemplified in the *BUB1* case (Figure 40). These cases altogether highlight the benefits of RNA analysis for the WES cases where no candidate gene was detected, but also cases in which the variants pathogenicity has not been previously reported and demands functional confirmation (Figure 50).

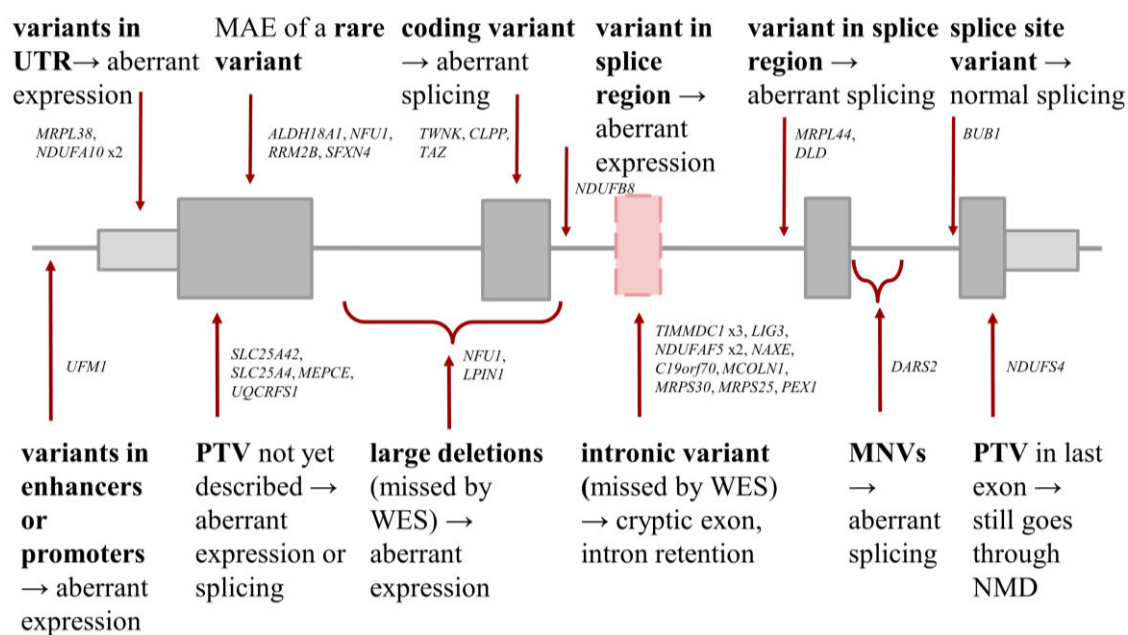


Figure 50. Possible effects of variants assessed by RNA-seq. Summary of variants and their effect on transcript across 35 cases from RNA-seq cohort, where RNA-seq was decisive in establishing genetic diagnosis in 34 of them, but also prevented false diagnosis in one, highlighting the value of transcriptome analysis in genetic diagnostics.

Nevertheless, although praised, the use of RNA-seq should still be considered with caution. To start with, it will not provide functional evidence for all variants, notably missense variants that mostly do not affect the transcript. Moreover, although variant calling in RNA-seq provides an added value to WES, it still fails in most of the intergenic and intronic regions, as well as in genes that are not expressed.

The most discussed limitation of RNA-seq is the issue of tissue-specific gene expression and splicing. Before opting for RNA-seq, clinicians and researchers can investigate in advance if and at which level are the genes of their interest expressed in a certain tissue by looking at the public datasets such as GTEx and the Expression Atlas (Papatheodorou et al., 2018). In addition, they can search for the sets of genes in PAGE (Gonorazky et al., 2019). Regarding differential splicing across tissues, MAJIQ-CAT tests similarity in splicing events of a group of genes in CAT compared to the others (Aicher et al., 2020). Even if the gene is expressed in the probed tissue, its variant might not have an effect there (Rivas et al., 2015). For our cohort, we used skin-derived patients' fibroblasts, easily obtainable during skin biopsy. We were able to detect 90% of mitochondrial disease genes, 5% more than in blood (Table 8). This percentage, together with tissue-specific splicing analysis from Aicher et al. (2020), argues in favor of using this CAT as a model of RNA-based investigation in the case of suspected mitochondrial disease. Nevertheless, our choice of tissue is simplified due to the essential role of mitochondria throughout cells and tissues. For some diseases, such as neuromuscular ones, this is not the case and the use of affected tissue is recommended (Gonorazky et al., 2019). An alternative approach could be transdifferentiation of available cells into induced pluripotent stem cells (iPSCs). With their embryonic-like state, iPSCs can provide insights into genetic regulation that occurs during

early development, and present a reliable tool for disease modelling (Sternecker et al., 2014). Indeed, they express as many as 27,046 genes, providing greater power over other tissues (Bonder et al., 2019). Furthermore, primary cells and iPSCs can be transdifferentiated into cells or even organoids that more accurately mirror tissue-specific pathology (Rowe and Daley, 2019). Gonorazky et al. (2019) generated t-myotubes from fibroblasts and showed that their transcriptome reflects the one from skeletal muscle. However, such approaches are more laborious, time-consuming, and limited to the cell types for which protocols are established. Also, they abolish the contribution of any somatically acquired mutation in the affected tissue to the disease pathology, and, on the contrary, can propagate somatic mutations accumulated during reprogramming, thus overestimating their relevance for the disease (Gore et al., 2011).

In summary, this work confirmed the value of RNA-seq as a complementary tool to WES in diagnostics. Apart from detecting aberrant transcript events, RNA-seq can be used for the discovery of variants elusive to WES. The methods used improved the specificity and sensitivity of outlier detection, and the established pipeline accelerated the preprocessing of raw data while ensuring its reproducibility, robustness, and scalability. With these improvements, implementation of RNA-seq has been simplified and improved our diagnosis rate by 16%. Thus, with decreasing costs of sequencing and such established and straightforward computational workflow, we hope to see RNA-seq as a standard companion to DNA analysis in future diagnostics.

5.2.3.2. Proteomics

Although RNA-seq enables functional validation of variant effect, it is uninformative if a variant does not affect the transcript, for example missense and gain-of-function variants. Missense variants are by far the most commonly described pathogenic variant in mitochondrial disease, and encompass 21% of all ClinVar pathogenic variants. Only frame-shift variants are more frequently reported as pathogenic (Figure 8). The effect of missense variant can vary, from none at all to the perturbation of molecular interactions and disrupted protein folding and/or stability (Sahni et al., 2015). For the global overview of the cellular proteome, proteomics can be implemented (Stenton et al., 2020). Based on mass spectrometry, it allows quantification of protein levels and specific protein signatures in a single assay, providing functional evidence of variant effect and removing the need for laborious and candidate-focused Western blot analyses. As opposed to RNA-seq, proteomics probes the final result of gene expression and is thus more closely related to a phenotype. This is important as protein levels are more robust against functionally irrelevant variance on the mRNA level. Also, RNA-seq analysis alone underestimates the effect of post-transcriptional and post-translational regulation on protein abundance. On the other hand, the detection power of mass-spectrometry is smaller than that of the sequencing methods. As an example, in patient-derived fibroblast cell lines approximately 5,000–8,000 proteins are quantified, including over 65% of mitochondrial disease proteins (Kremer et al., 2017; Stenton et al., 2020). Similarly as in RNA-seq, proteomics can detect expression outliers, slowly but surely paving the way for its implementation in the diagnostic setting (Kremer et al., 2017). It can also provide insights into protein complexes, exemplified by a study of the

role of 31 accessory subunits of CI for its assembly and function (Stroud et al., 2016). To date, proteomics has not been systematically applied to any rare disease cohort. However, it was used for functional validation of a VUS in *TIMMDC1* and its effect on CI (Kremer et al., 2017), as well as assessment of consequence of ablated MRPS34 on mitoribosome and mitochondrial translation (Lake et al., 2017). It was also used in my study for the validation of *NFUI*, *MRPL44* and *DARS2* gene defects (Figure 34 and 38). The main limitation of proteomics lays in its still incomplete coverage of cellular proteins and tissue-specific gene expression, so it does not yet allow systematic detection of all possible causal proteins. Another major limitation of proteomics arises in its detection of only aberrant expression, being fruitless for the variants that do not lead to protein degradation. Indeed, a handful of studies that systematically evaluated the impact of missense variants report that the majority disrupts specific protein interactions without affecting protein folding or stability, which may go undetected by proteomics (Wei et al., 2014; Sahni et al., 2015; Fragoza et al., 2019).

5.2.3.3. Metabolomics

Based on mass spectrometry, untargeted metabolomics offers the systematic capture and quantification of up to thousands of small molecule metabolites within tissue in a single assay (Zhang et al., 2016). Metabolomics can enable biochemical diagnosis, facilitate variant interpretation (Kremer et al., 2017; Chao de la Barca et al., 2020), reveal the up- and downstream effects of gene and mitochondrial dysfunction on cellular processes and identify novel biomarkers. However, steps in metabolomics are complex, from data generation due to the large differences in metabolites properties, to data interpretation that requires advanced enrichment and pathway analyses (Stenton et al., 2020). Finally, metabolic data are influenced not only by the primary genetic defect, but also by the genetic background and environment. As with proteomics, a large systematic metabolomics study in a diagnostic setting is yet to be reported. Still, metabolomics has been used for biochemical diagnostics of inborn errors of metabolism, identifying key diagnostic metabolites in 91% of samples and providing functional evidence for VUS interpretation (Coene et al., 2018).

To summarize, current efforts in genetics and diagnostics are focused on the implementation of OMICS technologies for a global insight into biological systems (Figure 51). It is important to emphasize that each method comes with its benefits and limitations. Thus, integrating OMICS data would provide a more comprehensive understanding of gene expression. This highlights the importance of further implementation of artificial intelligence for the interpretation of such large biological data (Dias and Torkamani, 2019; Azodi et al., 2020). With this in mind, hopefully in the near future more disease cohorts with integrated OMICS data will emerge, providing an increase in diagnostic yield and a better understanding of biological perturbations.

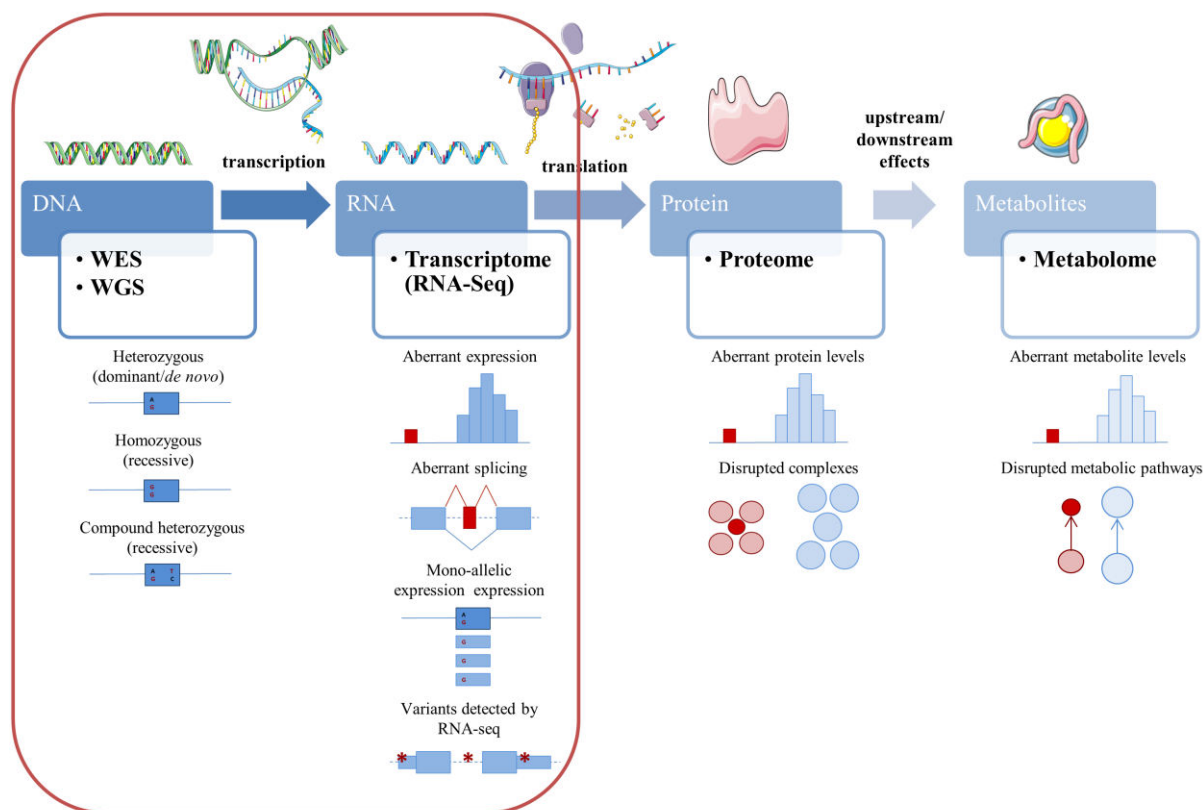


Figure 51. Proposed OMICS approach for genetic diagnostics based on the central dogma of molecular biology. Boxes provide the flow of genetic information (upper one) and the OMICS technologies used to investigate each level of information (lower one). Below the boxes possible observations are listed that can be obtained by each technology. The red box highlights the approaches used in this thesis.

6. Future perspectives

6.1. New sequencing technologies

6.1.1. Long read sequencing

Although WGS overcomes important limitations of WES with regards to variant detection, both methods are based on short-read sequencing, which is challenging for the detection of SVs, variants complicated by pseudogenes or homologous regions, and repetitive regions. Despite the constant improvements of bioinformatics algorithms, some regions remain impossible to be accurately mapped and assembled (Ashley, 2016). In addition, the variant phasing information is often not available. The development of long-read sequencing (LRS) in recent years promises to overcome these limitations. Pacific Biosciences' (PacBio) single-molecule real-time (SMRT) and Oxford Nanopore Technologies' (ONT) nanopore sequencing (Clarke et al., 2009; Eid et al., 2009) dominate the LRS market (Amarasinghe et al., 2020). These third-generation sequencing technologies generate reads >10 kb in average length from the single native DNA or RNA molecules (van Dijk et al., 2018). As LRS is PCR-free and its sequencing occurs in real-time, amplification biases are eliminated, and base modifications, such as methylation, are preserved in principle (Laszlo et al., 2013). These benefits of LRS promise better *de novo* assembly, mapping, identification of transcript isoforms, and detection of SVs. Clinical utility of LRS has been shown in several studies where they enabled SV detection and investigation of tandem repeat expansion loci, resolved variant phasing, and distinguished a gene of interest from pseudogenes (reviewed in Mantere et al., 2019). Although LRS is unquestionably the future of NGS, its higher error rate compared to short-read sequencing, still underdeveloped data analysis tools, and higher price are hampering their clinical implementation (Sedlazeck et al., 2018; Watson et al., 2019). Nevertheless, constant improvements in these aspects promise their role in medical genetics in foreseeable future.

6.1.2. Single-cell sequencing

The majority of current sequencing experiments are based on the assumption that cells within the same tissue are homogeneous. However, the sequencing of the bulk population underestimates the importance of cell-to-cell variability, loses the information on cellular heterogeneity, and is unable to analyse a small number of cells. Indeed, line of evidence suggests that gene expression differs even within the same cell types (Huang, 2009; Li and Clevers, 2010), enabling cellular coordination and differentiation (Eldar and Elowitz, 2010). Improvements in microfluidics and nanotechnology have enabled the rise of single-cell technologies, promising a better understanding of the behaviour of individual cells and its physiological consequences. Several types of single-cell OMICS technologies have been developed, and single-cell sequencing technologies refer to the sequencing of a single-cell genome or transcriptome. Single-cell genomic sequencing enables the detection of cell-specific DNA mutations. It revealed unexpectedly large genomic diversity arising from germ

cells (Wang et al., 2012; Hou et al., 2013), as well as somatic mutations, especially large CNVs in neurons (Cai et al., 2014). Additionally, the somatic mutations acquired by an individual cell can be used to construct cell lineage trees (Woodworth et al., 2017). Single-cell RNA-seq (scRNA-seq), first reported in 2009 (Tang et al., 2009), is currently the most widely used single-cell technology. This technique is based on sequencing of cDNA of an individual cell that was exponentially amplified by PCR before being subjected to library preparation (Tang et al., 2009). Detecting up to ~6,500 genes per cell (Baran-Gale et al., 2018), it allows dissecting cellular heterogeneity and transcriptomic changes in individual cells. This is very helpful for the understanding of the complexity of the immune system, where scRNA-seq led to the identification of specific immune cell subtypes (Chen et al., 2019), as well as of nerve system and its development (Lake et al., 2016; Zhong et al., 2018). In a clinical setting, scRNA-seq provides valuable information for the identification of disease-related cell populations or tracking the course of a disease. This has been widely used in cancer research, where it helped understand the molecular pathways behind its progression and guided the development of new diagnostic and treatment methods (Gonzalez-Silva et al., 2020). In addition, scRNA-seq has been used for investigation of common disorders, such as autism, Alzheimer's, cardiac diseases, as well as immune responses (Velmeshev et al., 2019; Mathys et al., 2019; Chaudhry et al., 2019; Szabo et al., 2019). Current limitations to wider application of single-cell sequencing are higher costs of equipment and reagents, and more complicated sample preparation and computational data analysis due to their high sensitivity (reviewed in Nguyen et al., 2018). However, constant improvements promise its wider application to basic research, as well as clinical diagnosis and treatments. Recently, single-cell sequencing has been used to detect mtDNA mutations, which were used for clone and lineage tracing of cells (Ludwig et al., 2019). This could be of future use in quantification and tracking of heteroplasmy levels in patients with pathogenic mtDNA variants.

6.2. Emerging non-coding functionality

Historically, the main focus of human genetics was on the protein-coding regions, while the remaining 98% of the genome was often referred to as “dark matter” or “junk DNA”. Its function is less clear and interpretation of its variants difficult. During the last decade, development of high-throughput OMICS technologies has made a systematic analysis of different stages of gene expression possible. The ENCODE project, aiming to complete the functional annotation of the genome, reported that 80% of the human genome is functional, taking part in at least one biochemical RNA- and/or chromatin-associated event (Consortium, 2012). Transcriptome analysis revealed that up to 83% of the genome is transcribed (Djebali et al., 2012). Ribosomal profiling showed translation is widespread and pervasive, with ~40% of lncRNAs, non-coding RNAs longer than 200 nt, interacting with ribosomes (Ingolia et al., 2011; Zeng et al., 2018). These results indicated the existence of different protein isoforms and regulatory upstream open reading frames ORFs (uORFs) originating from the mRNAs, but also new mechanisms of translational regulation, as well as micropeptides, peptides shorter than 100 AA, encoded by lncRNAs and uORFs. Nevertheless, one must always be aware that ribosomal occupancy of transcripts need not

necessarily mean the production of functional polypeptides, and that further evidence is necessary (Guttman et al., 2013). Mass-spectrometry has been implemented as a validation tool, and indeed detected hundreds of peptides encoded by previously non-coding regions (Wilhelm et al., 2014; Wang et al., 2019; van Heesch et al., 2019; Chen et al., 2020). Yet, to date these novel peptides remain characterized just on the level of reported detection, and further studies are needed in order to elucidate their function and omit the possibility of false positive findings. Nevertheless, OMICS-based studies stimulated the identification and functional characterization of non-coding regions and transcripts; especially the lncRNAs, in order to better understand their regulatory roles (Morris and Mattick, 2014; Kopp and Mendel, 2018). As reviewed in Gusic and Prokisch (2020), numerous ncRNAs have been functionally associated with mitochondria, reported to affect mitochondrial function indirectly via acting in the cytosol, or directly in mitochondria, as some seem to be encoded by mtDNA. However, the presence of ncRNAs in mitochondria, especially the issue of RNA transport to mitochondria, is still a subject of debates and more experimental evidence is needed. In addition, more challenges emerge in distinguishing between mRNAs and ncRNAs that indicate the limitations of automatic gene annotation. In the line with that, reports have emerged of hidden mitochondrial micropeptides. Humanin, MOTS-c, and SHLPs are encoded by mtDNA and involved in mitochondrial bioenergetics and regulation of metabolism (Kim et al., 2017), and mitoregulin, previously annotated as *LINC00116*, has been connected to the formation of supercomplexes, fatty acid oxidation, and Ca^{2+} dynamics (Stein et al., 2018; Makarewich et al., 2018; Lin et al., 2019). With 5% of the mitochondrial proteome consisting of micropeptides (Calvo et al., 2016), and emerging reports of micropeptides hidden in the non-coding regions that localize in mitochondria (van Heesch et al., 2019; Chen et al., 2020), it is expected that so far unknown mitochondrial proteins will be described in the near future, which will improve the genome annotation, but could also lead to a novel disease gene discovery. Currently, large-scale projects, particularly FANTOM (<https://fantom.gsc.riken.jp/>), ENCODE (<https://www.encodeproject.org/>), and GTEx (<https://www.gtexportal.org/home/>) are constantly improving the functional annotation of the genome and understanding of complexity behind gene regulation. This accumulating data indicate that at least a fraction of missing heritability could fall within the non-coding, but regulatory regions, such as promoters, enhancers, structural elements, and regions encoding for ncRNAs (Gloss and Dinger, 2018). So far, these regulatory regions have been seldom described as a cause of (mitochondrial) disease, for example in *UFMI* and *NDUFA10* cases with deletions in promoter and 5'UTR, respectively, within our RNA-seq cohort (Table 9). Further elucidation of function of these regions may enable more such discoveries.

6.3. Beyond Mendelian inheritance

With the exception of diseases caused by variants in the mtDNA, the usual focus of diagnostics in suspected mitochondrial diseases is on full-penetrant rare variants that segregate within a family. Such focus treats Mendelian traits and genetic loci as distinct entities, and thus underestimates the contribution of variants of different effect size as well as

environmental influences. Although inarguably the majority of variants we harbor are without consequence, some variants are harmful, and a considerable amount of variants could have an intermediate effect on a particular phenotype. These “beyond Mendelian inheritance” factors include genetic modifiers and digenic, oligogenic, and polygenic inheritance, as well as tissue specificity and gene-environment interactions.

Slowly but steadily, the concept of genetic modification is being introduced to the genetics of Mendelian diseases, where one allele - the “modifier”- modulates the effect of the second- disease-causing event (Kousi and Katsanis, 2015). So far, a handful of genetic modifiers in mitochondrial diseases have been described, where nuclear genes influenced the phenotypic effect of mtDNA variants. This is exemplified by a start-loss variant in *SSBP1* co-segregating with hearing loss in carriers of the homoplasmic m.1555A>G variant (Kullar et al., 2018). In addition, two studies reported the additive effect of nuclear alleles on LHON phenotype, where c.157C>T in *PRICKLE3* and c.572G>T in *YARS2* in interaction with m.11778G>A lead to a visual failure (Jiang et al., 2016; Yu et al., 2020). Such X-linked modifiers could explain the variations in the penetrance and sex bias observed in LHON. Next, the same variant in *YARS2* causes deafness in combination with m.7511A>G variant (Fan et al., 2019). It is noteworthy that these studies were based on genetic and segregation analysis of a small number of large families, and may not be widely applicable. A more systematic, singleton approach is yet to be reported. Finally, in cases of digenic, oligogenic, or polygenic inheritance, the combined consequences of multiple alleles ultimately cause the disease. Such cases are yet to be reported in mitochondrial diseases. Compared to the discovery of single disease genes, dissection of these phenomena is lagging heavily behind. The limiting factors to their discovery include reduced power of traditional genetic analyses, insufficiently detailed clinical evaluation, lack of large pedigrees, and unclear contribution of non-genetic factors (Kousi and Katsanis, 2015). Recent approaches, utilizing polygenic risk scores calculated by large GWAS studies, argued for the polygenic contribution of common variants to the variable expression when applied in a rare disease (Khera et al., 2018; Oetjens et al., 2019). Nevertheless, with the constant accumulation and integration of OMICS datasets, and improvements of computational methods, novel mechanisms that cause and shape a disease will emerge. This will altogether improve our understanding of both pathomechanisms and gene regulation.

7. References

- Acin-Perez, R., Fernandez-Silva, P., Peleato, M. L., Perez-Martos, A., & Enriquez, J. A. (2008). Respiratory active mitochondrial supercomplexes. *Mol Cell*, *32*(4), 529-539. doi:10.1016/j.molcel.2008.10.021
- Acuna-Hidalgo, R., Veltman, J. A., & Hoischen, A. (2016). New insights into the generation and role of de novo mutations in health and disease. *Genome Biol*, *17*(1), 241. doi:10.1186/s13059-016-1110-1
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., . . . Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods*, *7*(4), 248-249. doi:10.1038/nmeth0410-248
- Aicher, J. K., Jewell, P., Vaquero-Garcia, J., Barash, Y., & Bhoj, E. J. (2020). Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genetics in Medicine*. doi:10.1038/s41436-020-0780-y
- Ait-El-Mkadem, S., Dayem-Quere, M., Gusic, M., Chausseot, A., Bannwarth, S., Francois, B., . . . Paquis-Flucklinger, V. (2017). Mutations in MDH2, Encoding a Krebs Cycle Enzyme, Cause Early-Onset Severe Encephalopathy. *Am J Hum Genet*, *100*(1), 151-159. doi:10.1016/j.ajhg.2016.11.014
- Al-Owain, M., Colak, D., Albakheet, A., Al-Younes, B., Al-Humaidi, Z., Al-Sayed, M., . . . Kaya, N. (2013). Clinical and biochemical features associated with BCS1L mutation. *J Inherit Metab Dis*, *36*(5), 813-820. doi:10.1007/s10545-012-9536-4
- Albers, C. A., Paul, D. S., Schulze, H., Freson, K., Stephens, J. C., Smethurst, P. A., . . . Ghevaert, C. (2012). Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet*, *44*(4), 435-439, S431-432. doi:10.1038/ng.1083
- Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat Rev Genet*, *16*(4), 197-212. doi:10.1038/nrg3891
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J. D. (1994). Molecular biology of the cell, 4th edition. *New York: Garland Science*.
- Alfares, A., Aloraini, T., Al Subaie, L., Alissa, A., Al Qudsi, A., Alahmad, A., . . . Alfadhel, M. (2018). Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genetics in Medicine*, *20*(11), 1328-1333. doi:10.1038/gim.2018.41
- Allegretti, M., Klusch, N., Mills, D. J., Vonck, J., Kuhlbrandt, W., & Davies, K. M. (2015). Horizontal membrane-intrinsic alpha-helices in the stator a-subunit of an F-type ATP synthase. *Nature*, *521*(7551), 237-240. doi:10.1038/nature14185
- Alston, C. L., Davison, J. E., Meloni, F., van der Westhuizen, F. H., He, L., Hornig-Do, H. T., . . . Taylor, R. W. (2012). Recessive germline SDHA and SDHB mutations causing leukodystrophy and isolated mitochondrial complex II deficiency. *J Med Genet*, *49*(9), 569-577. doi:10.1136/jmedgenet-2012-101146
- Alston, C. L., Rocha, M. C., Lax, N. Z., Turnbull, D. M., & Taylor, R. W. (2017). The genetics and pathology of mitochondrial disease. *Journal of Pathology*, *241*(2), 236-250. doi:10.1002/path.4809
- Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. *21*(1):30. doi: 10.1186/s13059-020-1935-5.

- Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., & Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res*, *43*(Database issue), D789-798. doi:10.1093/nar/gku1205
- Anderson, N. M., Mucka, P., Kern, J. G., & Feng, H. (2018). The emerging role and targetability of the TCA cycle in cancer metabolism. *Protein Cell*, *9*(2), 216-237. doi:10.1007/s13238-017-0451-1
- Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., . . . Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, *290*(5806), 457-465. doi:10.1038/290457a0
- Andreu, A. L., Hanna, M. G., Reichmann, H., Bruno, C., Penn, A. S., Tanji, K., . . . DiMauro, S. (1999). Exercise intolerance due to mutations in the cytochrome b gene of mitochondrial DNA. *N Engl J Med*, *341*(14), 1037-1044. doi:10.1056/NEJM199909303411404
- Andreux, P. A., Blanco-Bose, W., Ryu, D., Burdet, F., Ibberson, M., Aebischer, P., . . . Rinsch, C. (2019). The mitophagy activator urolithin A is safe and induces a molecular signature of improved mitochondrial and cellular health in humans. *Nat Metab*, *1*(6), 595-603. doi:10.1038/s42255-019-0073-4
- Anna, A., & Monika, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*, *59*(3), 253-268. doi:10.1007/s13353-018-0444-7
- Anso, E., Weinberg, S. E., Diebold, L. P., Thompson, B. J., Malinge, S., Schumacker, P. T., . . . Chandel, N. S. (2017). The mitochondrial respiratory chain is essential for haematopoietic stem cell function. *Nat Cell Biol*, *19*(6), 614-625. doi:10.1038/ncb3529
- Ardissone, A., Invernizzi, F., Nasca, A., Moroni, I., Farina, L., & Ghezzi, D. (2015). Mitochondrial leukoencephalopathy and complex II deficiency associated with a recessive SDHB mutation with reduced penetrance. *Mol Genet Metab Rep*. *5*:51-54. doi:10.1016/j.ymgmr.2015.10.006
- Ashley, E. A. (2016). Towards precision medicine. *Nat Rev Genet*, *17*(9), 507-522. doi:10.1038/nrg.2016.86
- Azodi, C. B., Tang, J., & Shiu, S. H. (2020). Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends Genet*, *36*(6), 442-455. doi:10.1016/j.tig.2020.03.005
- Bacman, S. R., Kauppila, J. H. K., Pereira, C. V., Nissanka, N., Miranda, M., Pinto, M., . . . Moraes, C. T. (2018). MitoTALEN reduces mutant mtDNA load and restores tRNA(Ala) levels in a mouse model of heteroplasmic mtDNA mutation. *Nat Med*, *24*(11), 1696-1700. doi:10.1038/s41591-018-0166-8
- Bacman, S. R., Williams, S. L., Pinto, M., Peralta, S., & Moraes, C. T. (2013). Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs. *Nat Med*, *19*(9), 1111-1113. doi:10.1038/nm.3261
- Baldovino, S., Moliner, A. M., Taruscio, D., Daina, E., & Roccatello, D. (2016). Rare Diseases in Europe: from a Wide to a Local Perspective. *Isr Med Assoc J*, *18*(6), 359-363.
- Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., & Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, *12*(11), 745-755. doi:10.1038/nrg3031

- Bamshad, M. J., Nickerson, D. A., & Chong, J. X. (2019). Mendelian Gene Discovery: Fast and Furious with No End in Sight. *Am J Hum Genet*, *105*(3), 448-455. doi:10.1016/j.ajhg.2019.07.011
- Bannwarth, S., Ait-El-Mkadem, S., Chaussonot, A., Genin, E. C., Lacas-Gervais, S., Fragaki, K., . . . Paquis-Flucklinger, V. (2014). A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain*, *137*(Pt 8), 2329-2345. doi:10.1093/brain/awu138
- Baralle, F. E., & Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol*, *18*(7), 437-451. doi:10.1038/nrm.2017.27
- Baran-Gale, J., Chandra, T., & Kirschner, K. (2018). Experimental design for single-cell RNA sequencing. *Brief Funct Genomics*, *17*(4), 233-239. doi:10.1093/bfgp/elx035
- Barbitoff, Y. A., Polev, D. E., Glotov, A. S., Serebryakova, E. A., Shcherbakova, I. V., Kiselev, A. M., . . . Predeus, A. V. (2020). Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. *Sci Rep*, *10*(1), 2057. doi:10.1038/s41598-020-59026-y
- Barel, O., Shorer, Z., Flusser, H., Ofir, R., Narkis, G., Finer, G., . . . Birk, O. S. (2008). Mitochondrial complex III deficiency associated with a homozygous mutation in UQCRCQ. *Am J Hum Genet*, *82*(5), 1211-1216. doi:10.1016/j.ajhg.2008.03.020
- Baur, J. A., Pearson, K. J., Price, N. L., Jamieson, H. A., Lerin, C., Kalra, A., . . . Sinclair, D. A. (2006). Resveratrol improves health and survival of mice on a high-calorie diet. *Nature*, *444*(7117), 337-342. doi:10.1038/nature05354
- Bayley, J. P., Launonen, V., & Tomlinson, I. P. (2008). The FH mutation database: an online database of fumarate hydratase mutations involved in the MCUL (HLRCC) tumor syndrome and congenital fumarase deficiency. *BMC Med Genet*, *9*, 20. doi:10.1186/1471-2350-9-20
- Baynam, G., Walters, M., Claes, P., Kung, S., LeSouef, P., Dawkins, H., . . . Goldblatt, J. (2015). Phenotyping: targeting genotype's rich cousin for diagnosis. *J Paediatr Child Health*, *51*(4), 381-386. doi:10.1111/jpc.12705
- Baysal, B. E., Ferrell, R. E., Willett-Brozick, J. E., Lawrence, E. C., Myssiorek, D., Bosch, A., . . . Devlin, B. (2000). Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. *Science*, *287*(5454), 848-851. doi:10.1126/science.287.5454.848
- Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., Antipenko, A., . . . Abel, L. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5473-5478. doi:10.1073/pnas.1418631112
- Bell, C. J., Dinwiddie, D. L., Miller, N. A., Hateley, S. L., Ganusova, E. E., Mudge, J., . . . Kingsmore, S. F. (2011). Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing. *Science Translational Medicine*, *3*(65). doi:10.1126/scitranslmed.3001756
- Benit, P., Lebon, S., & Rustin, P. (2009). Respiratory-chain diseases related to complex III deficiency. *Biochim Biophys Acta*, *1793*(1), 181-185. doi:10.1016/j.bbamcr.2008.06.004
- Bianco, A., Martinez-Romero, I., Bisceglia, L., D'Agruma, L., Favia, P., Ruiz-Pesini, E., . . . Petruzzella, V. (2016). Mitochondrial DNA copy number differentiates the Leber's hereditary optic neuropathy affected individuals from the unaffected mutation carriers. *Brain*, *139*(Pt 1), e1. doi:10.1093/brain/awv216

- Biesecker, L. G., & Green, R. C. (2014). Diagnostic clinical genome and exome sequencing. *N Engl J Med*, *371*(12), 1170. doi:10.1056/NEJMc1408914
- Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. *Cell*, *126*(1), 37-47. doi:10.1016/j.cell.2006.06.023
- Bodemer, C., Rotig, A., Rustin, P., Cormier, V., Niaudet, P., Saudubray, J. M., . . . de Prost, Y. (1999). Hair and skin disorders as signs of mitochondrial disease. *Pediatrics*, *103*(2), 428-433. doi:10.1542/peds.103.2.428
- Bohovych, I., & Khalimonchuk, O. (2016). Sending Out an SOS: Mitochondria as a Signaling Hub. *Front Cell Dev Biol*, *4*, 109. doi:10.3389/fcell.2016.00109
- Bonder, M. J., Smail, C., Gloudemans, M. J., Frésard, L., Jakubosky, D., D'Antonio, M., . . . Stegle, O. (2019). Systematic assessment of regulatory effects of human disease variants in pluripotent cells. *bioRxiv*. doi: <https://doi.org/10.1101/784967>
- Bonnefont, J. P., Chretien, D., Rustin, P., Robinson, B., Vassault, A., Aupetit, J., . . . Munnich, A. (1992). Alpha-ketoglutarate dehydrogenase deficiency presenting as congenital lactic acidosis. *J Pediatr*, *121*(2), 255-258. doi:10.1016/s0022-3476(05)81199-0
- Bottani, E., Cerutti, R., Harbour, M. E., Ravaglia, S., Dogan, S. A., Giordano, C., . . . Zeviani, M. (2017). TTC19 Plays a Husbandry Role on UQCRFS1 Turnover in the Biogenesis of Mitochondrial Respiratory Complex III. *Mol Cell*, *67*(1), 96-105 e104. doi:10.1016/j.molcel.2017.06.001
- Bourgeron, T., Chretien, D., Poggi-Bach, J., Doonan, S., Rabier, D., Letouze, P., . . . Rustin, P. (1994). Mutation of the fumarase gene in two siblings with progressive encephalopathy and fumarase deficiency. *J Clin Invest*, *93*(6), 2514-2518. doi:10.1172/JCI117261
- Bourgeron, T., Rustin, P., Chretien, D., Birch-Machin, M., Bourgeois, M., Viegas-Pequignot, E., . . . Rotig, A. (1995). Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. *Nat Genet*, *11*(2), 144-149. doi:10.1038/ng1095-144
- Boycott, K. M., Rath, A., Chong, J. X., Hartley, T., Alkuraya, F. S., Baynam, G., . . . Lochmuller, H. (2017). International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet*, *100*(5), 695-705. doi:10.1016/j.ajhg.2017.04.003
- Brand, M. D., & Nicholls, D. G. (2011). Assessing mitochondrial dysfunction in cells (vol 435, pg 297, 2011). *Biochemical Journal*, *437*, 575-575.
- Brandt, U., Yu, L., Yu, C. A., & Trumpower, B. L. (1993). The mitochondrial targeting presequence of the Rieske iron-sulfur protein is processed in a single step after insertion into the cytochrome bc1 complex in mammals and retained as a subunit in the complex. *J Biol Chem*, *268*(12), 8387-8390.
- Brechtmann, F., Mertes, C., Matuszewska, A., Yezpez, V. A., Avsec, Z., Herzog, M., . . . Gagneur, J. (2018). OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data. *Am J Hum Genet*, *103*(6), 907-917. doi:10.1016/j.ajhg.2018.10.025
- Briere, J. J., Favier, J., Gimenez-Roqueplo, A. P., & Rustin, P. (2006). Tricarboxylic acid cycle dysfunction as a cause of human diseases and tumor formation. *American Journal of Physiology-Cell Physiology*, *291*(6), C1114-C1120. doi:10.1152/ajpcell.00216.2006
- Broeks, M. H., Shamseldin, H. E., Alhashem, A., Hashem, M., Abdulwahab, F., Alshedi, T., . . . Alkuraya, F. S. (2019). MDH1 deficiency is a metabolic disorder of the malate-

- aspartate shuttle associated with early onset severe encephalopathy. *Hum Genet*, 138(11-12), 1247-1257. doi:10.1007/s00439-019-02063-z
- Buskin, A., Zhu, L., Chichagova, V., Basu, B., Mozaffari-Jovin, S., Dolan, D., . . . Lako, M. (2018). Disrupted alternative splicing for genes implicated in splicing and ciliogenesis causes PRPF31 retinitis pigmentosa. *Nature Communications*, 9(1), 4234. doi:10.1038/s41467-018-06448-y
- Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., & Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nature Reviews Genetics*, 17(5), 257-271. doi:10.1038/nrg.2016.10
- Cai, X., Evrony, G. D., Lehmann, H. S., Elhosary, P. C., Mehta, B. K., Poduri, A., & Walsh, C. A. (2014). Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep*, 8(5), 1280-1289. doi:10.1016/j.celrep.2014.07.043
- Calsina, B., Curras-Freixes, M., Buffet, A., Pons, T., Contreras, L., Leton, R., . . . Robledo, M. (2018). Role of MDH2 pathogenic variant in pheochromocytoma and paraganglioma patients. *Genetics in Medicine*, 20(12), 1652-1662. doi:10.1038/s41436-018-0068-7
- Calvo, S. E., Clauser, K. R., & Mootha, V. K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res*, 44(D1), D1251-1257. doi:10.1093/nar/gkv1003
- Calvo, S. E., Compton, A. G., Hershman, S. G., Lim, S. C., Lieber, D. S., Tucker, E. J., . . . Mootha, V. K. (2012). Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Science Translational Medicine*, 4(118), 118ra110. doi:10.1126/scitranslmed.3003310
- Cameron, J. M., Janer, A., Levandovskiy, V., Mackay, N., Rouault, T. A., Tong, W. H., . . . Robinson, B. H. (2011). Mutations in iron-sulfur cluster scaffold genes NFU1 and BOLA3 cause a fatal deficiency of multiple respiratory chain and 2-oxoacid dehydrogenase enzymes. *Am J Hum Genet*, 89(4), 486-495. doi:10.1016/j.ajhg.2011.08.011
- Carossa, V., Ghelli, A., Tropeano, C. V., Valentino, M. L., Iommarini, L., Maresca, A., . . . Carelli, V. (2014). A novel in-frame 18-bp microdeletion in MT-CYB causes a multisystem disorder with prominent exercise intolerance. *Hum Mutat*, 35(8), 954-958. doi:10.1002/humu.22596
- Carroll, C. J., Isohanni, P., Poyhonen, R., Euro, L., Richter, U., Brillhante, V., . . . Suomalainen, A. (2013). Whole-exome sequencing identifies a mutation in the mitochondrial ribosome protein MRPL44 to underlie mitochondrial infantile cardiomyopathy. *J Med Genet*, 50(3), 151-159. doi:10.1136/jmedgenet-2012-101375
- Carrozzo, R., Dionisi-Vici, C., Steuerwald, U., Luciola, S., Deodato, F., Di Giandomenico, S., . . . Wevers, R. A. (2007). SUCLA2 mutations are associated with mild methylmalonic aciduria, Leigh-like encephalomyopathy, dystonia and deafness. *Brain*, 130(Pt 3), 862-874. doi:10.1093/brain/awl389
- Cartegni, L., Chew, S. L., & Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet*, 3(4), 285-298. doi:10.1038/nrg775
- Cascon, A., Comino-Mendez, I., Curras-Freixes, M., de Cubas, A. A., Contreras, L., Richter, S., . . . Robledo, M. (2015). Whole-exome sequencing identifies MDH2 as a new familial paraganglioma gene. *J Natl Cancer Inst*, 107(5). doi:10.1093/jnci/djv053

- Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., & Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biol*, *16*, 195. doi:10.1186/s13059-015-0762-6
- Castresana, J., & Saraste, M. (1995). Evolution of energetic metabolism: the respiration-early hypothesis. *Trends Biochem Sci*, *20*(11), 443-448. doi:10.1016/s0968-0004(00)89098-2
- Cavelier, L., Jazin, E., Jalonen, P., & Gyllensten, U. (2000). MtDNA substitution rate and segregation of heteroplasmy in coding and noncoding regions. *Hum Genet*, *107*(1), 45-50. doi:10.1007/s004390000305
- Chaisson, M. J. P., Sanders, A. D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., . . . Lee, C. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nature Communications*, *10*(1), 1784. doi:10.1038/s41467-018-08148-z
- Chang, W. H., Niu, D. M., Lu, C. Y., Lin, S. Y., Liu, T. C., & Chang, J. G. (2017). Modulation the alternative splicing of GLA (IVS4+919G>A) in Fabry disease. *PLoS One*, *12*(4), e0175929. doi:10.1371/journal.pone.0175929
- Chao de la Barca, J. M., Fogazza, M., Rugolo, M., Chupin, S., Del Dotto, V., Ghelli, A. M., . . . Zanna, C. (2020). Metabolomics hallmarks OPA1 variants correlating with their in vitro phenotype and predicting clinical severity. *Hum Mol Genet*, *29*(8), 1319-1329. doi:10.1093/hmg/ddaa047
- Chaudhry, F., Isherwood, J., Bawa, T., Patel, D., Gurdziel, K., Lanfear, D. E., . . . Levy, P. D. (2019). Single-Cell RNA Sequencing of the Cardiovascular System: New Looks for Old Diseases. *Front Cardiovasc Med*, *6*, 173. doi:10.3389/fcvm.2019.00173
- Chen, H., Ye, F., & Guo, G. (2019). Revolutionizing immunology with single-cell RNA sequencing. *Cell Mol Immunol*, *16*(3), 242-249. doi:10.1038/s41423-019-0214-4
- Chen, J., Brunner, A. D., Cogan, J. Z., Nunez, J. K., Fields, A. P., Adamson, B., . . . Weissman, J. S. (2020). Pervasive functional translation of noncanonical human open reading frames. *Science*, *367*(6482), 1140-1146. doi:10.1126/science.aay0262
- Chen, L., Tovar-Corona, J. M., & Urrutia, A. O. (2012). Alternative splicing: a potential source of functional innovation in the eukaryotic genome. *Int J Evol Biol*, *2012*, 596274. doi:10.1155/2012/596274
- Chen, Y. H., Liu, S. J., Gao, M. M., Zeng, T., Lin, G. W., Tan, N. N., . . . Long, Y. S. (2017). MDH2 is an RNA binding protein involved in downregulation of sodium channel *Scn1a* expression under seizure condition. *Biochim Biophys Acta Mol Basis Dis*, *1863*(6), 1492-1499. doi:10.1016/j.bbadis.2017.04.018
- Chess, A. (2012). Mechanisms and consequences of widespread random monoallelic expression. *Nat Rev Genet*, *13*(6), 421-428. doi:10.1038/nrg3239
- Chinnery, P. F. (1993). Mitochondrial Disorders Overview. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, & A. Amemiya (Eds.), *GeneReviews((R))*. Seattle (WA).
- Chinnery, P. F. (2014) Mitochondrial Disorders Overview. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, & A. Amemiya (Eds.), *GeneReviews((R))*. Seattle (WA).
- Choi, M., Scholl, U. I., Ji, W. Z., Liu, T. W., Tikhonova, I. R., Zumbo, P., . . . Lifton, R. P. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 19096-19101. doi:10.1073/pnas.0910672106

- Chong, J. X., Buckingham, K. J., Jhangiani, S. N., Boehm, C., Sobreira, N., Smith, J. D., . . . Genomics, C. M. (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *American Journal of Human Genetics*, 97(2), 199-215. doi:10.1016/j.ajhg.2015.06.009
- Chretien, D., Slama, A., Briere, J. J., Munnich, A., Rotig, A., & Rustin, P. (2004). Revisiting pitfalls, problems and tentative solutions for assaying mitochondrial respiratory chain complex III in human samples. *Curr Med Chem*, 11(2), 233-239. doi:10.2174/0929867043456151
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). 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)*, 6(2), 80-92. doi:10.4161/fly.19695
- Clark, M. M., Hildreth, A., Batalov, S., Ding, Y., Chowdhury, S., Watkins, K., . . . Kingsmore, S. F. (2019). Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci Transl Med*, 11(489). doi:10.1126/scitranslmed.aat6177
- Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S., & Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*, 4(4), 265-270. doi:10.1038/nnano.2009.12
- Claussnitzer, M., Cho, J. H., Collins, R., Cox, N. J., Dermitzakis, E. T., Hurler, M. E., . . . McCarthy, M. I. (2020). A brief history of human disease genetics. *Nature*, 577(7789), 179-189. doi:10.1038/s41586-019-1879-7
- Coene, K. L. M., Kluijtmans, L. A. J., van der Heeft, E., Engelke, U. F. H., de Boer, S., Hoegen, B., . . . Wevers, R. A. (2018). Next-generation metabolic screening: targeted and untargeted metabolomics for the diagnosis of inborn errors of metabolism in individual patients. *J Inherit Metab Dis*, 41(3), 337-353. doi:10.1007/s10545-017-0131-6
- Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alfoldi, J., Francioli, L. C., . . . Talkowski, M. E. (2020). A structural variation reference for medical and population genetics. *Nature*, 581(7809), 444-451. doi:10.1038/s41586-020-2287-8
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., . . . Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol*, 17, 13. doi:10.1186/s13059-016-0881-8
- Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57-74. doi:10.1038/nature11247
- Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., & Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet*, 132(10), 1077-1130. doi:10.1007/s00439-013-1331-2
- Cooper, G. M., & Shendure, J. (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet*, 12(9), 628-640. doi:10.1038/nrg3046
- Crofts, A. R., Holland, J. T., Victoria, D., Kolling, D. R., Dikanov, S. A., Gilbreth, R., . . . Kuras, M. G. (2008). The Q-cycle reviewed: How well does a monomeric mechanism of the bc(1) complex account for the function of a dimeric complex? *Biochim Biophys Acta*, 1777(7-8), 1001-1019. doi:10.1016/j.bbabo.2008.04.037
- Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., . . . Cinsort, G.-T. E. (2017). Improving genetic diagnosis in Mendelian disease with

- transcriptome sequencing. *Science Translational Medicine*, 9(386). doi:10.1126/scitranslmed.aal5209
- D'Souza, A. R., & Minczuk, M. (2018). Mitochondrial transcription and translation: overview. *Essays Biochem*, 62(3), 309-320. doi:10.1042/EBC20170102
- Dahal, S., Dubey, S., & Raghavan, S. C. (2018). Homologous recombination-mediated repair of DNA double-strand breaks operates in mammalian mitochondria. *Cell Mol Life Sci*, 75(9), 1641-1655. doi:10.1007/s00018-017-2702-y
- Danhauser, K., Iuso, A., Haack, T. B., Freisinger, P., Brockmann, K., Mayr, J. A., . . . Prokisch, H. (2011). Cellular rescue-assay aids verification of causative DNA-variants in mitochondrial complex I deficiency. *Mol Genet Metab*. 103(2):161-6. doi: 10.1016/j.ymgme.2011.03.004
- DaRe, J. T., Vasta, V., Penn, J., Tran, N. T., & Hahn, S. H. (2013). Targeted exome sequencing for mitochondrial disorders reveals high genetic heterogeneity. *BMC Med Genet*, 14, 118. doi:10.1186/1471-2350-14-118
- de Ligt, J., Willemsen, M. H., van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., . . . Vissers, L. E. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med*, 367(20), 1921-1929. doi:10.1056/NEJMoa1206524
- de Lonlay, P., Valnot, I., Barrientos, A., Gorbatyuk, M., Tzagoloff, A., Taanman, J. W., . . . Rotig, A. (2001). A mutant mitochondrial respiratory chain assembly protein causes complex III deficiency in patients with tubulopathy, encephalopathy and liver failure. *Nat Genet*, 29(1), 57-60. doi:10.1038/ng706
- Desmet, F. O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., & Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*, 37(9), e67. doi:10.1093/nar/gkp215
- Dewey, F. E., Grove, M. E., Pan, C. P., Goldstein, B. A., Bernstein, J. A., Chaib, H., . . . Quertermous, T. (2014). Clinical Interpretation and Implications of Whole-Genome Sequencing. *Jama-Journal of the American Medical Association*, 311(10), 1035-1044. doi:10.1001/jama.2014.1717
- Dias, R., & Torkamani, A. (2019). Artificial intelligence in clinical and genomic diagnostics. *Genome Med*, 11(1), 70. doi:10.1186/s13073-019-0689-8
- Diaz, F., Enriquez, J. A., & Moraes, C. T. (2012a). Cells lacking Rieske iron-sulfur protein have a reactive oxygen species-associated decrease in respiratory complexes I and IV. *Mol Cell Biol*, 32(2), 415-429. doi:10.1128/MCB.06051-11
- Diaz, F., Garcia, S., Padgett, K. R., & Moraes, C. T. (2012b). A defect in the mitochondrial complex III, but not complex IV, triggers early ROS-dependent damage in defined brain regions. *Hum Mol Genet*, 21(23), 5066-5077. doi:10.1093/hmg/dds350
- Dillon, L. M., Hida, A., Garcia, S., Prolla, T. A., & Moraes, C. T. (2012). Long-term bezafibrate treatment improves skin and spleen phenotypes of the mtDNA mutator mouse. *PLoS One*, 7(9), e44335. doi:10.1371/journal.pone.0044335
- DiMauro, S., Schon, E. A., Carelli, V., & Hirano, M. (2013). The clinical maze of mitochondrial neurology. *Nat Rev Neurol*, 9(8), 429-444. doi:10.1038/nrneurol.2013.126
- Distelmaier, F., Haack, T. B., Catarino, C. B., Gallenmuller, C., Rodenburg, R. J., Strom, T. M., . . . Klopstock, T. (2015). MRPL44 mutations cause a slowly progressive multisystem disease with childhood-onset hypertrophic cardiomyopathy. *Neurogenetics*, 16(4), 319-323. doi:10.1007/s10048-015-0444-2

- Distelmaier, F., Haack, T. B., Wortmann, S. B., Mayr, J. A., & Prokisch, H. (2017). Treatable mitochondrial diseases: cofactor metabolism and beyond. *Brain*, *140*(2), e11. doi:10.1093/brain/aww303
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., . . . Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101-108. doi:10.1038/nature11233
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. doi:10.1093/bioinformatics/bts635
- Dominguez-Gonzalez, C., Madruga-Garrido, M., Mavillard, F., Garone, C., Aguirre-Rodriguez, F. J., Donati, M. A., . . . Hirano, M. (2019). Deoxynucleoside Therapy for Thymidine Kinase 2-Deficient Myopathy. *Ann Neurol*, *86*(2), 293-303. doi:10.1002/ana.25506
- Ebbert, M. T. W., Jensen, T. D., Jansen-West, K., Sens, J. P., Reddy, J. S., Ridge, P. G., . . . Fryer, J. D. (2019). Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol*, *20*(1), 97. doi:10.1186/s13059-019-1707-2
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., . . . Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*(5910), 133-138. doi:10.1126/science.1162986
- Eilbeck, K., Quinlan, A., & Yandell, M. (2017). Settling the score: variant prioritization and Mendelian disease. *Nat Rev Genet*, *18*(10), 599-612. doi:10.1038/nrg.2017.52
- Eisner, V., Picard, M., & Hajnocyky, G. (2018). Mitochondrial dynamics in adaptive and maladaptive cellular stress responses. *Nat Cell Biol*, *20*(7), 755-765. doi:10.1038/s41556-018-0133-0
- El-Hattab, A. W., Zarante, A. M., Almannai, M., & Scaglia, F. (2017). Therapies for mitochondrial diseases and current clinical trials. *Mol Genet Metab*, *122*(3), 1-9. doi:10.1016/j.ymgme.2017.09.009
- Eldar, A., & Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature*, *467*(7312), 167-173. doi:10.1038/nature09326
- Elpeleg, O., Miller, C., Hershkovitz, E., Bitner-Glindzicz, M., Bondi-Rubenstein, G., Rahman, S., . . . Saada, A. (2005). Deficiency of the ADP-forming succinyl-CoA synthase activity is associated with encephalomyopathy and mitochondrial DNA depletion. *American Journal of Human Genetics*, *76*(6), 1081-1086. doi:10.1086/430843
- Elstner, M., Andreoli, C., Klopstock, T., Meitinger, T., & Prokisch, H. (2009). The mitochondrial proteome database: MitoP2. *Methods Enzymol*, *457*, 3-20. doi:10.1016/S0076-6879(09)05001-0
- Emanuelsson, O., Brunak, S., von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*, *2*(4), 953-971. doi:10.1038/nprot.2007.131
- Endo, H., Hasegawa, K., Narisawa, K., Tada, K., Kagawa, Y., & Ohta, S. (1989). Defective gene in lactic acidosis: abnormal pyruvate dehydrogenase E1 alpha-subunit caused by a frame shift. *Am J Hum Genet*, *44*(3), 358-364.
- Ernster, L., Ikkos, D., & Luft, R. (1959). Enzymic activities of human skeletal muscle mitochondria: a tool in clinical metabolic research. *Nature*, *184*, 1851-1854. doi:10.1038/1841851a0

- Fan, W., Zheng, J., Kong, W., Cui, L., Aishanjiang, M., Yi, Q., . . . Guan, M. X. (2019). Contribution of a mitochondrial tyrosyl-tRNA synthetase mutation to the phenotypic expression of the deafness-associated tRNA(Ser(UCN)) 7511A>G mutation. *J Biol Chem*, 294(50), 19292-19305. doi:10.1074/jbc.RA119.010598
- Fang, H., Wu, Y. Y., Narzisi, G., O'Rawe, J. A., Barron, L. T. J., Rosenbaum, J., . . . Lyon, G. J. (2014). Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Medicine*, 6. doi:10.1186/s13073-014-0089-z
- Feichtinger, R. G., Brunner-Krainz, M., Alhaddad, B., Wortmann, S. B., Kovacs-Nagy, R., Stojakovic, T., . . . Mayr, J. A. (2017). Combined Respiratory Chain Deficiency and UQCC2 Mutations in Neonatal Encephalomyopathy: Defective Supercomplex Assembly in Complex III Deficiencies. *Oxid Med Cell Longev*, 2017, 7202589. doi:10.1155/2017/7202589
- Feng, J., Bussiere, F., & Hekimi, S. (2001). Mitochondrial electron transport is a key determinant of life span in *Caenorhabditis elegans*. *Dev Cell*, 1(5), 633-644. doi:10.1016/s1534-5807(01)00071-5
- Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet*, 12(8), 565-575. doi:10.1038/nrg3032
- Fernandez-Vizarra, E., & Zeviani, M. (2015). Nuclear gene mutations as the cause of mitochondrial complex III deficiency. *Frontiers in Genetics*, 6, 134. doi:10.3389/fgene.2015.00134
- Fernandez-Vizarra, E., & Zeviani, M. (2018). Mitochondrial complex III Rieske Fe-S protein processing and assembly. *Cell Cycle*, 17(6), 681-687. doi:10.1080/15384101.2017.1417707
- Ferreira, C. R. (2019). The burden of rare diseases. *Am J Med Genet A*, 179(6), 885-892. doi:10.1002/ajmg.a.61124
- Finsterer, J. (2007). Hematological manifestations of primary mitochondrial disorders. *Acta Haematol*, 118(2), 88-98. doi:10.1159/000105676
- Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., . . . Carter, N. P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*, 84(4), 524-533. doi:10.1016/j.ajhg.2009.03.010
- Fisher, N., Bourges, I., Hill, P., Brasseur, G., & Meunier, B. (2004). Disruption of the interaction between the Rieske iron-sulfur protein and cytochrome b in the yeast bc1 complex owing to a human disease-associated mutation within cytochrome b. *Eur J Biochem*, 271(7), 1292-1298. doi:10.1111/j.1432-1033.2004.04036.x
- Flavahan, W. A., Gaskell, E., & Bernstein, B. E. (2017). Epigenetic plasticity and the hallmarks of cancer. *Science*, 357(6348). doi:10.1126/science.aal2380
- Fragoza, R., Das, J., Wierbowski, S. D., Liang, J., Tran, T. N., Liang, S., . . . Yu, H. (2019). Extensive disruption of protein interactions by genetic variants across the allele frequency spectrum in human populations. *Nature Communications*, 10(1), 4141. doi:10.1038/s41467-019-11959-3
- Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., . . . Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*, 47(D1), D766-D773. doi:10.1093/nar/gky955
- Frazier, A. E., Thorburn, D. R., & Compton, A. G. (2019). Mitochondrial energy generation disorders: genes, mechanisms, and clues to pathology. *J Biol Chem*, 294(14), 5386-5395. doi:10.1074/jbc.R117.809194

- Fresard, L., Smail, C., Ferraro, N. M., Teran, N. A., Li, X., Smith, K. S., . . . Montgomery, S. B. (2019). Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med*, 25(6), 911-919. doi:10.1038/s41591-019-0457-8
- Gabalton, T., & Huynen, M. A. (2004). Shaping the mitochondrial proteome. *Biochim Biophys Acta*, 1659(2-3), 212-220. doi:10.1016/j.bbabi.2004.07.011
- Gaignard, P., Menezes, M., Schiff, M., Bayot, A., Rak, M., Ogier de Baulny, H., . . . Rustin, P. (2013). Mutations in CYC1, encoding cytochrome c1 subunit of respiratory chain complex III, cause insulin-responsive hyperglycemia. *Am J Hum Genet*, 93(2), 384-389. doi:10.1016/j.ajhg.2013.06.015
- Galassi, G., Lamantea, E., Invernizzi, F., Tavani, F., Pisano, I., Ferrero, I., Palmieri, L., & Zeviani, M. (2008). Additive effects of POLG1 and ANT1 mutations in a complex encephalomyopathy. *Neuromuscul Disord*, 18(6):465-70. doi:10.1016/j.nmd.2008.03.013.
- Galluzzi, L., Kepp, O., & Kroemer, G. (2012). Mitochondria: master regulators of danger signalling. *Nat Rev Mol Cell Biol*, 13(12), 780-788. doi:10.1038/nrm3479
- Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., . . . Kroemer, G. (2018). Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ*, 25(3), 486-541. doi:10.1038/s41418-017-0012-4
- Gammage, P. A., Van Haute, L., & Minczuk, M. (2016). Engineered mtZFNs for Manipulation of Human Mitochondrial DNA Heteroplasmy. *Methods Mol Biol*, 1351, 145-162. doi:10.1007/978-1-4939-3040-1_11
- Gammage, P. A., Viscomi, C., Simard, M. L., Costa, A. S. H., Gaude, E., Powell, C. A., . . . Minczuk, M. (2018). Genome editing in mitochondria corrects a pathogenic mtDNA mutation in vivo. *Nat Med*, 24(11), 1691-1695. doi:10.1038/s41591-018-0165-9
- Ghezzi, D., Arzuffi, P., Zordan, M., Da Re, C., Lamperti, C., Benna, C., . . . Zeviani, M. (2011). Mutations in TTC19 cause mitochondrial complex III deficiency and neurological impairment in humans and flies. *Nat Genet*, 43(3), 259-263. doi:10.1038/ng.761
- Ghezzi, D., Goffrini, P., Uziel, G., Horvath, R., Klopstock, T., Lochmuller, H., . . . Zeviani, M. (2009). SDHAF1, encoding a LYR complex-II specific assembly factor, is mutated in SDH-defective infantile leukoencephalopathy. *Nature Genetics*, 41(6), 654-656. doi:10.1038/ng.378
- Gilissen, C., Hahir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W. M., Willemsen, M. H., . . . Veltman, J. A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature*, 511(7509), 344-+. doi:10.1038/nature13394
- Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., & Chess, A. (2007). Widespread monoallelic expression on human autosomes. *Science*, 318(5853), 1136-1140. doi:10.1126/science.1148910
- Gloss, B. S., & Dinger, M. E. (2018). Realizing the significance of noncoding functionality in clinical genomics. *Exp Mol Med*, 50(8), 97. doi:10.1038/s12276-018-0087-0
- Goldstein, D. B., Allen, A., Keebler, J., Margulies, E. H., Petrou, S., Petrovski, S., & Sunyaev, S. (2013). Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet*, 14(7), 460-470. doi:10.1038/nrg3455
- Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., . . . Dowling, J. J. (2019). Expanding the Boundaries of RNA Sequencing as a Diagnostic

- Tool for Rare Mendelian Disease. *Am J Hum Genet*, 104(5), 1007. doi:10.1016/j.ajhg.2019.04.004
- Gonzalez-Silva, L., Quevedo, L., & Varela, I. (2020). Tumor Functional Heterogeneity Unraveled by scRNA-seq Technologies. *Trends Cancer*, 6(1), 13-19. doi:10.1016/j.trecan.2019.11.010
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333-351. doi:10.1038/nrg.2016.49
- Gore, A., Li, Z., Fung, H. L., Young, J. E., Agarwal, S., Antosiewicz-Bourget, J., . . . Zhang, K. (2011). Somatic coding mutations in human induced pluripotent stem cells. *Nature*, 471(7336), 63-67. doi:10.1038/nature09805
- Gorman, G. S., Chinnery, P. F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., . . . Turnbull, D. M. (2016). Mitochondrial diseases. *Nat Rev Dis Primers*, 2, 16080. doi:10.1038/nrdp.2016.80
- Gorman, G. S., McFarland, R., Stewart, J., Feeney, C., & Turnbull, D. M. (2018). Mitochondrial donation: from test tube to clinic. *Lancet*, 392(10154), 1191-1192. doi:10.1016/S0140-6736(18)31868-3
- Gorman, G. S., Schaefer, A. M., Ng, Y., Gomez, N., Blakely, E. L., Alston, C. L., . . . McFarland, R. (2015). Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann Neurol*, 77(5), 753-759. doi:10.1002/ana.24362
- Goward, C. R., & Nicholls, D. J. (1994). Malate dehydrogenase: a model for structure, evolution, and catalysis. *Protein Sci*, 3(10), 1883-1888. doi:10.1002/pro.5560031027
- Grady, J. P., Pickett, S. J., Ng, Y. S., Alston, C. L., Blakely, E. L., Hardy, S. A., . . . McFarland, R. (2018). mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease. *EMBO Mol Med*, 10(6). doi:10.15252/emmm.201708262
- Greenfield, A., Braude, P., Flinter, F., Lovell-Badge, R., Ogilvie, C., & Perry, A. C. F. (2017). Assisted reproductive technologies to prevent human mitochondrial disease transmission. *Nat Biotechnol*, 35(11), 1059-1068. doi:10.1038/nbt.3997
- Greggio, C., Jha, P., Kulkarni, S. S., Lagarrigue, S., Broskey, N. T., Boutant, M., . . . Amati, F. (2017). Enhanced Respiratory Chain Supercomplex Formation in Response to Exercise in Human Skeletal Muscle. *Cell Metab*, 25(2), 301-311. doi:10.1016/j.cmet.2016.11.004
- Gruschke, S., Kehrein, K., Rompler, K., Grone, K., Israel, L., Imhof, A., . . . Ott, M. (2011). Cbp3-Cbp6 interacts with the yeast mitochondrial ribosomal tunnel exit and promotes cytochrome b synthesis and assembly. *J Cell Biol*, 193(6), 1101-1114. doi:10.1083/jcb.201103132
- Gruschke, S., Rompler, K., Hildenbeutel, M., Kehrein, K., Kuhl, I., Bonnefoy, N., & Ott, M. (2012). The Cbp3-Cbp6 complex coordinates cytochrome b synthesis with bc(1) complex assembly in yeast mitochondria. *J Cell Biol*, 199(1), 137-150. doi:10.1083/jcb.201206040
- Gu, J., Wu, M., Guo, R., Yan, K., Lei, J., Gao, N., & Yang, M. (2016). The architecture of the mammalian respirasome. *Nature*, 537(7622), 639-643. doi:10.1038/nature19359
- Guan, J., Yang, E., Yang, J., Zeng, Y., Ji, G., & Cai, J. J. (2016). Exploiting aberrant mRNA expression in autism for gene discovery and diagnosis. *Hum Genet*, 135(7), 797-811. doi:10.1007/s00439-016-1673-7

- Guffon, N., Lopez-Mediavilla, C., Dumoulin, R., Mousson, B., Godinot, C., Carrier, H., . . . Guibaud, P. (1993). 2-Ketoglutarate dehydrogenase deficiency, a rare cause of primary hyperlactataemia: report of a new case. *J Inherit Metab Dis*, *16*(5), 821-830. doi:10.1007/bf00714273
- Guo, R., Zong, S., Wu, M., Gu, J., & Yang, M. (2017). Architecture of Human Mitochondrial Respiratory Megacomplex I2III2IV2. *Cell*, *170*(6), 1247-1257 e1212. doi:10.1016/j.cell.2017.07.050
- Gusic, M., & Prokisch, H. (2020). ncRNAs: New Players in Mitochondrial Health and Disease? *Frontiers in Genetics*, *11*, 95. doi:10.3389/fgene.2020.00095
- Gusic, M., Schottmann, G., Feichtinger, R. G., Du, C., Scholz, C., Wagner, M., . . . Schuelke, M. (2020). Bi-Allelic UQCRFS1 Variants Are Associated with Mitochondrial Complex III Deficiency, Cardiomyopathy, and Alopecia Totalis. *Am J Hum Genet*, *106*(1), 102-111. doi:10.1016/j.ajhg.2019.12.005
- Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., & Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*, *154*(1), 240-251. doi:10.1016/j.cell.2013.06.009
- Guy, J., Feuer, W. J., Davis, J. L., Porciatti, V., Gonzalez, P. J., Koilkonda, R. D., . . . Lam, B. L. (2017). Gene Therapy for Leber Hereditary Optic Neuropathy: Low- and Medium-Dose Visual Results. *Ophthalmology*, *124*(11), 1621-1634. doi:10.1016/j.ophtha.2017.05.016
- Haack, T. B., Danhauser, K., Haberberger, B., Hoser, J., Strecker, V., Boehm, D., . . . Prokisch, H. (2010). Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nature Genetics*, *42*(12), 1131-+. doi:10.1038/ng.706
- Haack, T. B., Haberberger, B., Frisch, E. M., Wieland, T., Iuso, A., Gorza, M., . . . Prokisch, H. (2012a). Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *J Med Genet*, *49*(4), 277-283. doi:10.1136/jmedgenet-2012-100846
- Haack, T. B., Madignier, F., Herzer, M., Lamantea, E., Danhauser, K., Invernizzi, F., . . . Prokisch, H. (2012b). Mutation screening of 75 candidate genes in 152 complex I deficiency cases identifies pathogenic variants in 16 genes including NDUFB9. *J Med Genet*, *49*(2), 83-89. doi:10.1136/jmedgenet-2011-100577
- Haas, R. H., Parikh, S., Falk, M. J., Saneto, R. P., Wolf, N. I., Darin, N., & Cohen, B. H. (2007). Mitochondrial disease: a practical approach for primary care physicians. *Pediatrics*, *120*(6), 1326-1333. doi:10.1542/peds.2007-0391
- Habara, Y., Takeshima, Y., Awano, H., Okizuka, Y., Zhang, Z., Saiki, K., . . . Matsuo, M. (2009). In vitro splicing analysis showed that availability of a cryptic splice site is not a determinant for alternative splicing patterns caused by +1G-->A mutations in introns of the dystrophin gene. *J Med Genet*, *46*(8), 542-547. doi:10.1136/jmg.2008.061259
- Hagen, C. M., Aidt, F. H., Havndrup, O., Hedley, P. L., Jespersgaard, C., Jensen, M., . . . Christiansen, M. (2013). MT-CYB mutations in hypertrophic cardiomyopathy. *Mol Genet Genomic Med*, *1*(1), 54-65. doi:10.1002/mgg3.5
- Haller, T., Ortner, M., & Gnaiger, E. (1994). A respirometer for investigating oxidative cell metabolism: toward optimization of respiratory studies. *Anal Biochem*, *218*(2), 338-342. doi:10.1006/abio.1994.1188
- Hamilton, E. M. C., Bertini, E., Kalaydjieva, L., Morar, B., Dojcakova, D., Liu, J., . . . Recessive, H. A. B. C. R. G. (2017). UFM1 founder mutation in the Roma population causes recessive variant of H-ABC. *Neurology*, *89*(17), 1821-1828. doi:10.1212/WNL.0000000000004578

- Hatefi, Y. (1985). The mitochondrial electron transport and oxidative phosphorylation system. *Annu Rev Biochem*, 54, 1015-1069. doi:10.1146/annurev.bi.54.070185.005055
- Haut, S., Brivet, M., Touati, G., Rustin, P., Lebon, S., Garcia-Cazorla, A., . . . Slama, A. (2003). A deletion in the human QP-C gene causes a complex III deficiency resulting in hypoglycaemia and lactic acidosis. *Hum Genet*, 113(2), 118-122. doi:10.1007/s00439-003-0946-0
- Havens, M. A., Duelli, D. M., & Hastings, M. L. (2013). Targeting RNA splicing for disease therapy. *Wiley Interdiscip Rev RNA*, 4(3), 247-266. doi:10.1002/wrna.1158
- Henze, K., & Martin, W. (2003). Evolutionary biology: essence of mitochondria. *Nature*, 426(6963), 127-128. doi:10.1038/426127a
- Heyne, H. O., Singh, T., Stamberger, H., Abou Jamra, R., Caglayan, H., Craiu, D., . . . Lemke, J. R. (2018). De novo variants in neurodevelopmental disorders with epilepsy. *Nat Genet*, 50(7), 1048-1053. doi:10.1038/s41588-018-0143-7
- Hildenbeutel, M., Hegg, E. L., Stephan, K., Gruschke, S., Meunier, B., & Ott, M. (2014). Assembly factors monitor sequential hemylation of cytochrome b to regulate mitochondrial translation. *J Cell Biol*, 205(4), 511-524. doi:10.1083/jcb.201401009
- Hinson, J. T., Fantin, V. R., Schonberger, J., Breivik, N., Siem, G., McDonough, B., . . . Seidman, C. E. (2007). Missense mutations in the BCS1L gene as a cause of the Bjornstad syndrome. *N Engl J Med*, 356(8), 809-819. doi:10.1056/NEJMoa055262
- Hoischen, A., van Bon, B. W., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., . . . Veltman, J. A. (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet*, 42(6), 483-485. doi:10.1038/ng.581
- Hollegaard, M. V., Skogstrand, K., Thorsen, P., Norgaard-Pedersen, B., Hougaard, D. M., & Grove, J. (2013). Joint analysis of SNPs and proteins identifies regulatory IL18 gene variations decreasing the chance of spastic cerebral palsy. *Hum Mutat*, 34(1), 143-148. doi:10.1002/humu.22173
- Holt, I. J., Harding, A. E., & Morgan-Hughes, J. A. (1988). Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature*, 331(6158), 717-719. doi:10.1038/331717a0
- Horvath, R. (2012). Update on clinical aspects and treatment of selected vitamin-responsive disorders II (riboflavin and CoQ 10). *J Inherit Metab Dis*, 35(4), 679-687. doi:10.1007/s10545-011-9434-1
- Hou, Y., Fan, W., Yan, L., Li, R., Lian, Y., Huang, J., . . . Qiao, J. (2013). Genome analyses of single human oocytes. *Cell*, 155(7), 1492-1506. doi:10.1016/j.cell.2013.11.040
- Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136(23), 3853-3862. doi:10.1242/dev.035139
- Hughes, B. G., & Hekimi, S. (2011). A mild impairment of mitochondrial electron transport has sex-specific effects on lifespan and aging in mice. *PLoS One*, 6(10), e26116. doi:10.1371/journal.pone.0026116
- Hwang, S., Kim, E., Lee, I., & Marcotte, E. M. (2015). Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Scientific Reports*, 5. doi:10.1038/srep17875
- Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4), 789-802. doi:10.1016/j.cell.2011.10.002
- Invernizzi, F., Tigano, M., Dallabona, C., Donnini, C., Ferrero, I., Cremonese, M., . . . Zeviani, M. (2013). A homozygous mutation in LYRM7/MZM1L associated with early onset

- encephalopathy, lactic acidosis, and severe reduction of mitochondrial complex III activity. *Hum Mutat*, 34(12), 1619-1622. doi:10.1002/humu.22441
- Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., . . . Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515(7526), 216-221. doi:10.1038/nature13908
- Iwata, S., Lee, J. W., Okada, K., Lee, J. K., Iwata, M., Rasmussen, B., . . . Jap, B. K. (1998). Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. *Science*, 281(5373), 64-71. doi:10.1126/science.281.5373.64
- Jackson, C. B., Nuoffer, J. M., Hahn, D., Prokisch, H., Haberberger, B., Gautschi, M., . . . Schaller, A. (2014). Mutations in SDHD lead to autosomal recessive encephalomyopathy and isolated mitochondrial complex II deficiency. *J Med Genet*, 51(3), 170-175. doi:10.1136/jmedgenet-2013-101932
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., . . . Farh, K. K. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, 176(3), 535-548 e524. doi:10.1016/j.cell.2018.12.015
- Jeppesen, T. D., Schwartz, M., Olsen, D. B., Wibrand, F., Krag, T., Duno, M., . . . Vissing, J. (2006). Aerobic training is safe and improves exercise capacity in patients with mitochondrial myopathy. *Brain*, 129(Pt 12), 3402-3412. doi:10.1093/brain/awl149
- Jiang, P., Jin, X., Peng, Y., Wang, M., Liu, H., Liu, X., . . . Guan, M. X. (2016). The exome sequencing identified the mutation in YARS2 encoding the mitochondrial tyrosyl-tRNA synthetase as a nuclear modifier for the phenotypic manifestation of Leber's hereditary optic neuropathy-associated mitochondrial DNA mutation. *Hum Mol Genet*, 25(3), 584-596. doi:10.1093/hmg/ddv498
- Jiang, Y. H., Yuen, R. K. C., Wang, M. B., Jin, X., Chen, N., Wu, X. L., . . . Scherer, S. W. (2013). Detection of Clinically Relevant Genetic Variants in Autism Spectrum Disorder by Whole-Genome Sequencing. *American Journal of Human Genetics*, 93(2), 249-263. doi:10.1016/j.ajhg.2013.06.012
- Jin, S. C., Homsy, J., Zaidi, S., Lu, Q., Morton, S., DePalma, S. R., . . . Brueckner, M. (2017). Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet*, 49(11), 1593-1601. doi:10.1038/ng.3970
- Jochmanova, I., Zhuang, Z., & Pacak, K. (2015). Pheochromocytoma: Gasping for Air. *Horm Cancer*, 6(5-6), 191-205. doi:10.1007/s12672-015-0231-4
- Johns, D. R. (1996). The other human genome: mitochondrial DNA and disease. *Nat Med*, 2(10), 1065-1068. doi:10.1038/nm1096-1065
- Kang, H. C., Lee, Y. M., Kim, H. D., Lee, J. S., & Slama, A. (2007). Safe and effective use of the ketogenic diet in children with epilepsy and mitochondrial respiratory chain complex defects. *Epilepsia*, 48(1), 82-88. doi:10.1111/j.1528-1167.2006.00906.x
- Kapustin, Y., Chan, E., Sarkar, R., Wong, F., Vorechovsky, I., Winston, R. M., . . . Dibb, N. J. (2011). Cryptic splice sites and split genes. *Nucleic Acids Res*, 39(14), 5837-5844. doi:10.1093/nar/gkr203
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., . . . MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434-443. doi:10.1038/s41586-020-2308-7
- Katsyuba, E., Mottis, A., Zietak, M., De Franco, F., van der Velpen, V., Gariani, K., . . . Auwerx, J. (2018). De novo NAD(+) synthesis enhances mitochondrial function and improves health. *Nature*, 563(7731), 354-359. doi:10.1038/s41586-018-0645-6

- Kaupilla, J. H. K., Baines, H. L., Bratic, A., Simard, M. L., Freyer, C., Mourier, A., . . . Stewart, J. B. (2016). A Phenotype-Driven Approach to Generate Mouse Models with Pathogenic mtDNA Mutations Causing Mitochondrial Disease. *Cell Rep*, *16*(11), 2980-2990. doi:10.1016/j.celrep.2016.08.037
- Khan, N. A., Auranen, M., Paetau, I., Pirinen, E., Euro, L., Forsstrom, S., . . . Suomalainen, A. (2014). Effective treatment of mitochondrial myopathy by nicotinamide riboside, a vitamin B3. *EMBO Mol Med*, *6*(6), 721-731. doi:10.1002/emmm.201403943
- Khan, N. A., Nikkanen, J., Yatsuga, S., Jackson, C., Wang, L., Pradhan, S., . . . Suomalainen, A. (2017). mTORC1 Regulates Mitochondrial Integrated Stress Response and Mitochondrial Myopathy Progression. *Cell Metab*, *26*(2), 419-428 e415. doi:10.1016/j.cmet.2017.07.007
- Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., . . . Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*, *50*(9), 1219-1224. doi:10.1038/s41588-018-0183-z
- Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., . . . Pandey, A. (2014). A draft map of the human proteome. *Nature*, *509*(7502), 575-581. doi:10.1038/nature13302
- Kim, S. J., Xiao, J., Wan, J., Cohen, P., & Yen, K. (2017). Mitochondrially derived peptides as novel regulators of metabolism. *J Physiol*, *595*(21), 6613-6621. doi:10.1113/JP274472
- Kirby, D. M., Thorburn, D. R., Turnbull, D. M., & Taylor, R. W. (2007). Biochemical assays of respiratory chain complex activity. *Methods Cell Biol*, *80*, 93-119. doi:10.1016/S0091-679X(06)80004-X
- Klopstock, T., Metz, G., Yu-Wai-Man, P., Buchner, B., Gallenmuller, C., Bailie, M., . . . Chinnery, P. F. (2013). Persistence of the treatment effect of idebenone in Leber's hereditary optic neuropathy. *Brain*, *136*(Pt 2), e230. doi:10.1093/brain/aws279
- Knierim, E., Lucke, B., Schwarz, J. M., Schuelke, M., & Seelow, D. (2011). Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One*, *6*(11), e28240. doi:10.1371/journal.pone.0028240
- Kobayashi, Y., Yang, S., Nykamp, K., Garcia, J., Lincoln, S. E., & Topper, S. E. (2017). Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. *Genome Med*, *9*(1), 13. doi:10.1186/s13073-017-0403-7
- Koene, S., de Laat, P., van Tienoven, D. H., Vriens, D., Brandt, A. M., Sweep, F. C., . . . Smeitink, J. A. (2014). Serum FGF21 levels in adult m.3243A>G carriers: clinical implications. *Neurology*, *83*(2), 125-133. doi:10.1212/WNL.0000000000000578
- Koene, S., & Smeitink, J. (2011). Metabolic manipulators: a well founded strategy to combat mitochondrial dysfunction. *J Inherit Metab Dis*, *34*(2), 315-325. doi:10.1007/s10545-010-9162-y
- Koenig, M. K. (2008). Presentation and diagnosis of mitochondrial disorders in children. *Pediatr Neurol*, *38*(5), 305-313. doi:10.1016/j.pediatrneurol.2007.12.001
- Kohda, M., Tokuzawa, Y., Kishita, Y., Nyuzuki, H., Moriyama, Y., Mizuno, Y., . . . Okazaki, Y. (2016). A Comprehensive Genomic Analysis Reveals the Genetic Landscape of Mitochondrial Respiratory Chain Complex Deficiencies. *Plos Genetics*, *12*(1). doi:10.1371/journal.pgen.1005679
- Kopp, F., & Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*, *172*(3), 393-407. doi:10.1016/j.cell.2018.01.011

- Korde, A. S., Yadav, V. R., Zheng, Y. M., & Wang, Y. X. (2011). Primary role of mitochondrial Rieske iron-sulfur protein in hypoxic ROS production in pulmonary artery myocytes. *Free Radic Biol Med*, 50(8), 945-952. doi:10.1016/j.freeradbiomed.2011.01.010
- Kousi, M., & Katsanis, N. (2015). Genetic modifiers and oligogenic inheritance. *Cold Spring Harb Perspect Med*, 5(6). doi:10.1101/cshperspect.a017145
- Krebs, H. A., & Johnson, W. A. (1980). The role of citric acid in intermediate metabolism in animal tissues. *FEBS Lett*, 117 Suppl, K1-10. doi:10.4159/harvard.9780674366701.c143
- Kremer, L. S., Bader, D. M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., . . . Prokisch, H. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nature Communications*, 8. doi:10.1038/ncomms15824
- Kremer, L. S., & Prokisch, H. (2017). Identification of Disease-Causing Mutations by Functional Complementation of Patient-Derived Fibroblast Cell Lines. *Methods Mol Biol*, 1567, 391-406. doi:10.1007/978-1-4939-6824-4_24
- Kremer, L. S., Wortmann, S. B., & Prokisch, H. (2018). "Transcriptomics": molecular diagnosis of inborn errors of metabolism via RNA-sequencing. *Journal of Inherited Metabolic Disease*, 41(3), 525-532. doi:10.1007/s10545-017-0133-4
- Kullar, P. J., Gomez-Duran, A., Gammage, P. A., Garone, C., Minczuk, M., Golder, Z., . . . Chinnery, P. F. (2018). Heterozygous SSBP1 start loss mutation co-segregates with hearing loss and the m.1555A>G mtDNA variant in a large multigenerational family. *Brain*, 141(1), 55-62. doi:10.1093/brain/awx295
- Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7), 1073-1082. doi:10.1038/nprot.2009.86
- Lake, B. B., Ai, R., Kaeser, G. E., Salathia, N. S., Yung, Y. C., Liu, R., . . . Zhang, K. (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science*, 352(6293), 1586-1590. doi:10.1126/science.aaf1204
- Lake, N. J., Webb, B. D., Stroud, D. A., Richman, T. R., Ruzzenente, B., Compton, A. G., . . . Thorburn, D. R. (2017). Biallelic Mutations in MRPS34 Lead to Instability of the Small Mitochondrial Subunit and Leigh Syndrome. *Am J Hum Genet*, 101(2), 239-254. doi:10.1016/j.ajhg.2017.07.005
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., . . . Maglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46(D1), D1062-D1067. doi:10.1093/nar/gkx1153
- Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, 42(D1), D980-D985. doi:10.1093/nar/gkt1113
- Lane, R. K., Hilsabeck, T., & Rea, S. L. (2015). The role of mitochondrial dysfunction in age-related diseases. *Biochim Biophys Acta*, 1847(11), 1387-1400. doi:10.1016/j.bbabi.2015.05.021
- Laszlo, A. H., Derrington, I. M., Brinkerhoff, H., Langford, K. W., Nova, I. C., Samson, J. M., . . . Gundlach, J. H. (2013). Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A*, 110(47), 18904-18909. doi:10.1073/pnas.1310240110

- Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., . . . Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput Biol*, 9(8), e1003118. doi:10.1371/journal.pcbi.1003118
- Lee, C. F., Caudal, A., Abell, L., Nagana Gowda, G. A., & Tian, R. (2019). Targeting NAD(+) Metabolism as Interventions for Mitochondrial Disease. *Sci Rep*, 9(1), 3073. doi:10.1038/s41598-019-39419-4
- Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., . . . Nelson, S. F. (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA*, 312(18), 1880-1887. doi:10.1001/jama.2014.14604
- Lee, H., Huang, A. Y., Wang, L. K., Yoon, A. J., Renteria, G., Eskin, A., . . . Nelson, S. F. (2019). Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genetics in Medicine*. doi:10.1038/s41436-019-0672-1
- Lee, Y., & Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem*, 84, 291-323. doi:10.1146/annurev-biochem-060614-034316
- Legati, A., Reyes, A., Nasca, A., Invernizzi, F., Lamantea, E., Tiranti, V., . . . Zeviani, M. (2016). New genes and pathomechanisms in mitochondrial disorders unraveled by NGS technologies. *Biochim Biophys Acta*, 1857(8), 1326-1335. doi:10.1016/j.bbabi.2016.02.022
- Lehtinen, S. K., Hance, N., El Meziane, A., Juhola, M. K., Juhola, K. M., Karhu, R., . . . Jacobs, H. T. (2000). Genotypic stability, segregation and selection in heteroplasmic human cell lines containing np 3243 mutant mtDNA. *Genetics*, 154(1), 363-380.
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285-291. doi:10.1038/nature19057
- Lenaz, G., Fato, R., Formiggini, G., & Genova, M. L. (2007). The role of Coenzyme Q in mitochondrial electron transport. *Mitochondrion*, 7 Suppl, S8-33. doi:10.1016/j.mito.2007.03.009
- Letouze, E., Martinelli, C., Loriot, C., Burnichon, N., Abermil, N., Ottolenghi, C., . . . Favier, J. (2013). SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*, 23(6), 739-752. doi:10.1016/j.ccr.2013.04.018
- Letts, J. A., Fiedorczuk, K., & Sazanov, L. A. (2016). The architecture of respiratory supercomplexes. *Nature*, 537(7622), 644-648. doi:10.1038/nature19774
- Letts, J. A., & Sazanov, L. A. (2017). Clarifying the supercomplex: the higher-order organization of the mitochondrial electron transport chain. *Nat Struct Mol Biol*, 24(10), 800-808. doi:10.1038/nsmb.3460
- Levin, J. Z., Yassour, M., Adiconis, X. A., Nusbaum, C., Thompson, D. A., Friedman, N., . . . Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods*, 7(9), 709-U767. doi:10.1038/Nmeth.1491
- Levy-Lahad, E., Catane, R., Eisenberg, S., Kaufman, B., Hornreich, G., Lishinsky, E., . . . Halle, D. (1997). Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. *Am J Hum Genet*, 60(5), 1059-1067.
- Lewis, B. P., Green, R. E., & Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A*, 100(1), 189-192. doi:10.1073/pnas.0136770100

- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987-2993. doi:10.1093/bioinformatics/btr509
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. doi:10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. doi:10.1093/bioinformatics/btp352
- Li, H., Slone, J., Fei, L., & Huang, T. (2019). Mitochondrial DNA Variants and Common Diseases: A Mathematical Model for the Diversity of Age-Related mtDNA Mutations. *Cells*, 8(6). doi:10.3390/cells8060608
- Li, L., & Clevers, H. (2010). Coexistence of quiescent and active adult stem cells in mammals. *Science*, 327(5965), 542-545. doi:10.1126/science.1180794
- Li, X., Kim, Y., Tsang, E. K., Davis, J. R., Damani, F. N., Chiang, C., . . . Montgomery, S. B. (2017). The impact of rare variation on gene expression across tissues. *Nature*, 550(7675), 239-243. doi:10.1038/nature24267
- Li, Y. I., Knowles, D. A., Humphrey, J., Barbeira, A. N., Dickinson, S. P., Im, H. K., & Pritchard, J. K. (2018). Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet*, 50(1), 151-158. doi:10.1038/s41588-017-0004-9
- Lightowers, R. N., Taylor, R. W., & Turnbull, D. M. (2015). Mutations causing mitochondrial disease: What is new and what challenges remain? *Science*, 349(6255), 1494-1499. doi:10.1126/science.aac7516
- Lill, R. (2009). Function and biogenesis of iron-sulphur proteins. *Nature*, 460(7257), 831-838. doi:10.1038/nature08301
- Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J., & Fairbrother, W. G. (2011). Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A*, 108(27), 11093-11098. doi:10.1073/pnas.1101135108
- Lin, Y. F., Xiao, M. H., Chen, H. X., Meng, Y., Zhao, N., Yang, L., . . . Zhuang, S. M. (2019). A novel mitochondrial micropeptide MPM enhances mitochondrial respiratory activity and promotes myogenic differentiation. *Cell Death Dis*, 10(7), 528. doi:10.1038/s41419-019-1767-y
- Lindeboom, R. G., Supek, F., & Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet*, 48(10), 1112-1118. doi:10.1038/ng.3664
- Lionel, A. C., Costain, G., Monfared, N., Walker, S., Reuter, M. S., Hosseini, S. M., . . . Marshall, C. R. (2018). Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genetics in Medicine*, 20(4), 435-443. doi:10.1038/gim.2017.119
- Liu, H. Y., Zhou, L., Zheng, M. Y., Huang, J., Wan, S., Zhu, A., . . . Lu, Y. (2019). Diagnostic and clinical utility of whole genome sequencing in a cohort of undiagnosed Chinese families with rare diseases. *Sci Rep*, 9(1), 19365. doi:10.1038/s41598-019-55832-1
- Lott, M. T., Leipzig, J. N., Derbeneva, O., Xie, H. M., Chalkia, D., Sarmady, M., . . . Wallace, D. C. (2013). mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr Protoc Bioinformatics*, 44, 1-23. doi:10.1002/0471250953.bi0123s44

- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. doi:10.1186/s13059-014-0550-8
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS Comput Biol*, *13*(5), e1005457. doi:10.1371/journal.pcbi.1005457
- Ludwig, L. S., Lareau, C. A., Ulirsch, J. C., Christian, E., Muus, C., Li, L. H., . . . Sankaran, V. G. (2019). Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell*, *176*(6):1325-1339.e22. doi: 10.1016/j.cell.2019.01.022.
- Luft, R., Ikkos, D., Palmieri, G., Ernster, L., & Afzelius, B. (1962). 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*, *41*, 1776-1804. doi:10.1172/JCI104637
- Luo, S., Valencia, C. A., Zhang, J., Lee, N. C., Slone, J., Gui, B., . . . Huang, T. (2018). Biparental Inheritance of Mitochondrial DNA in Humans. *Proc Natl Acad Sci U S A*, *115*(51), 13039-13044. doi:10.1073/pnas.1810946115
- Lutsenko, S., & Cooper, M. J. (1998). Localization of the Wilson's disease protein product to mitochondria. *Proc Natl Acad Sci U S A*, *95*(11), 6004-6009. doi:10.1073/pnas.95.11.6004
- Lutz-Bonengel, S., & Parson, W. (2019). No further evidence for paternal leakage of mitochondrial DNA in humans yet. *Proc Natl Acad Sci U S A*, *116*(6), 1821-1822. doi:10.1073/pnas.1820533116
- Lykke-Andersen, S., & Jensen, T. H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol*, *16*(11), 665-677. doi:10.1038/nrm4063
- Lyons, T. W., Reinhard, C. T., & Planavsky, N. J. (2014). The rise of oxygen in Earth's early ocean and atmosphere. *Nature*, *506*(7488), 307-315. doi:10.1038/nature13068
- Ma, M., Ru, Y., Chuang, L. S., Hsu, N. Y., Shi, L. S., Hakenberg, J., . . . Chen, R. (2015). Disease-associated variants in different categories of disease located in distinct regulatory elements. *BMC Genomics*, *16*. doi:10.1186/1471-2164-16-S8-S3
- MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., . . . Tyler-Smith, C. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science*, *335*(6070), 823-828. doi:10.1126/science.1215040
- MacArthur, D. G., Manolio, T. A., Dimmock, D. P., Rehm, H. L., Shendure, J., Abecasis, G. R., . . . Gunter, C. (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature*, *508*(7497), 469-476. doi:10.1038/nature13127
- Maio, N., Kim, K. S., Singh, A., & Rouault, T. A. (2017). A Single Adaptable Cochaperone-Scaffold Complex Delivers Nascent Iron-Sulfur Clusters to Mammalian Respiratory Chain Complexes I-III. *Cell Metab*, *25*(4), 945-953 e946. doi:10.1016/j.cmet.2017.03.010
- Makarewich, C. A., Baskin, K. K., Munir, A. Z., Bezprozvannaya, S., Sharma, G., Khemtong, C., . . . Olson, E. N. (2018). MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid beta-Oxidation. *Cell Rep*, *23*(13), 3701-3709. doi:10.1016/j.celrep.2018.05.058
- Mantere, T., Kersten, S., & Hoischen, A. (2019). Long-Read Sequencing Emerging in Medical Genetics. *Frontiers in Genetics*, *10*. doi:10.3389/fgene.2019.00426
- Manwaring, N., Jones, M. M., Wang, J. J., Rochtchina, E., Howard, C., Mitchell, P., & Sue, C. M. (2007). Population prevalence of the MELAS A3243G mutation. *Mitochondrion*, *7*(3), 230-233. doi:10.1016/j.mito.2006.12.004

- Martinez-Reyes, I., & Chandel, N. S. (2020). Mitochondrial TCA cycle metabolites control physiology and disease. *Nature Communications*, *11*(1), 102. doi:10.1038/s41467-019-13668-3
- Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. *Nat Rev Mol Cell Biol*, *15*(2), 108-121. doi:10.1038/nrm3742
- Matharu, N., Rattanasopha, S., Tamura, S., Maliskova, L., Wang, Y., Bernard, A., . . . Ahituv, N. (2019). CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science*, *363*(6424). doi:10.1126/science.aau0629
- Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., . . . Tsai, L. H. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, *570*(7761), 332-337. doi:10.1038/s41586-019-1195-2
- Mayr, J. A., Haack, T. B., Freisinger, P., Karall, D., Makowski, C., Koch, J., . . . Sperl, W. (2015). Spectrum of combined respiratory chain defects. *J Inherit Metab Dis*, *38*(4), 629-640. doi:10.1007/s10545-015-9831-y
- McFarland, R., Taylor, R. W., & Turnbull, D. M. (2010). A neurological perspective on mitochondrial disease. *Lancet Neurol*, *9*(8), 829-840. doi:10.1016/S1474-4422(10)70116-2
- McGlinchy, N. J., & Smith, C. W. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem Sci*, *33*(8), 385-393. doi:10.1016/j.tibs.2008.06.001
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297-1303. doi:10.1101/gr.107524.110
- Mcrae, J. F., Clayton, S., Fitzgerald, T. W., Kaplanis, J., Prigmore, E., Rajan, D., . . . Hurles, M. E. (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature*, *542*(7642), 433-+. doi:10.1038/nature21062
- Medvedev, P., Stanciu, M., & Brudno, M. (2009). Computational methods for discovering structural variation with next-generation sequencing. *Nature Methods*, *6*(11), S13-S20. doi:10.1038/Nmeth.1374
- Meienberg, J., Bruggmann, R., Oexle, K., & Matyas, G. (2016). Clinical sequencing: is WGS the better WES? *Human Genetics*, *135*(3), 359-362. doi:10.1007/s00439-015-1631-9
- Mertes, C., Scheller, I., Yépez, V. A., Çelik, M. H., Liang, Y., Kremer, L., . . . Gagneur, Y. (2021). Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun*, *12*(1), 529. doi: 10.1038/s41467-020-20573-7
- Mignone, F., Gissi, C., Liuni, S., & Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biol*, *3*(3), REVIEWS0004. doi:10.1186/gb-2002-3-3-reviews0004
- Minikel, E. V., Karczewski, K. J., Martin, H. C., Cummings, B. B., Whiffin, N., Rhodes, D., . . . MacArthur, D. G. (2020). Evaluating drug targets through human loss-of-function genetic variation. *Nature*, *581*(7809), 459-464. doi:10.1038/s41586-020-2267-z
- Mishra, P., & Chan, D. C. (2014). Mitochondrial dynamics and inheritance during cell division, development and disease. *Nat Rev Mol Cell Biol*, *15*(10), 634-646. doi:10.1038/nrm3877
- Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, *191*, 144-148. doi:10.1038/191144a0
- Mitochondrial Medicine Society's Committee on, D., Haas, R. H., Parikh, S., Falk, M. J., Saneto, R. P., Wolf, N. I., . . . Naviaux, R. K. (2008). The in-depth evaluation of

- suspected mitochondrial disease. *Mol Genet Metab*, 94(1), 16-37. doi:10.1016/j.ymgme.2007.11.018
- MITOMAP: A Human Mitochondrial Genome Database (2008) <https://www.mitomap.org>
- Miyake, N., Yano, S., Sakai, C., Hatakeyama, H., Matsushima, Y., Shiina, M., . . . Matsumoto, N. (2013). Mitochondrial complex III deficiency caused by a homozygous UQCRC2 mutation presenting with neonatal-onset recurrent metabolic decompensation. *Hum Mutat*, 34(3), 446-452. doi:10.1002/humu.22257
- Mohammadi, P., Castel, S. E., Cummings, B. B., Einson, J., Sousa, C., Hoffman, P., . . . Lappalainen, T. (2019). Genetic regulatory variation in populations informs transcriptome analysis in rare disease. *Science*, 366(6463), 351-356. doi:10.1126/science.aay0256
- Montgomery, S. B., Lappalainen, T., Gutierrez-Arcelus, M., & Dermitzakis, E. T. (2011). Rare and common regulatory variation in population-scale sequenced human genomes. *PLoS Genet*, 7(7), e1002144. doi:10.1371/journal.pgen.1002144
- Moraes, C. T., Zeviani, M., Schon, E. A., Hickman, R. O., Vlcek, B. W., & DiMauro, S. (1991). Mitochondrial DNA deletion in a girl with manifestations of Kearns-Sayre and Lowe syndromes: an example of phenotypic mimicry? *Am J Med Genet*, 41(3), 301-305. doi:10.1002/ajmg.1320410308
- Moran, M., Marin-Buera, L., Gil-Borlado, M. C., Rivera, H., Blazquez, A., Seneca, S., . . . Ugalde, C. (2010). Cellular pathophysiological consequences of BCS1L mutations in mitochondrial complex III enzyme deficiency. *Hum Mutat*, 31(8), 930-941. doi:10.1002/humu.21294
- Morava, E., & Carozzo, R. (2014) Disorders of the Krebs Cycle. *Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases*, 313-322. doi:10.1007/978-3-642-40337-8_20
- Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nat Rev Genet*, 15(6), 423-437. doi:10.1038/nrg3722
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 5(7), 621-628. doi:10.1038/nmeth.1226
- Munnich, A., de Lonlay, P., Rotig, A., & Rustin, P. (2009). [Mitochondrial disorders]. *Bull Acad Natl Med*, 193(1), 19-41; discussion 41-13.
- Munnich, A., Rotig, A., Chretien, D., Cormier, V., Bourgeron, T., Bonnefont, J. P., . . . Rustin, P. (1996). Clinical presentation of mitochondrial disorders in childhood. *J Inherit Metab Dis*, 19(4), 521-527. doi:10.1007/bf01799112
- Munnich, A., & Rustin, P. (2001). Clinical spectrum and diagnosis of mitochondrial disorders. *Am J Med Genet*, 106(1), 4-17. doi:10.1002/ajmg.1391
- Nagel, R. L. (2005). Epistasis and the genetics of human diseases. *C R Biol*, 328(7), 606-615. doi:10.1016/j.crv.2005.05.003
- Nagy, E., & Maquat, L. E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci*, 23(6), 198-199. doi:10.1016/s0968-0004(98)01208-0
- Natarajan, P., Peloso, G. M., Zekavat, S. M., Montasser, M., Ganna, A., Chaffin, M., . . . Group, N. T. L. W. (2018). Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nature Communications*, 9(1), 3391. doi:10.1038/s41467-018-05747-8
- Nesbitt, V., Pitceathly, R. D., Turnbull, D. M., Taylor, R. W., Sweeney, M. G., Mudanohwo, E. E., . . . McFarland, R. (2013). 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*, 84(8), 936-938. doi:10.1136/jnnp-2012-303528
- Neveling, K., Feenstra, I., Gilissen, C., Hoefsloot, L. H., Kamsteeg, E. J., Mensenkamp, A. R., . . . Nelen, M. R. (2013). A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat*, 34(12), 1721-1726. doi:10.1002/humu.22450
- Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., . . . Bamshad, M. J. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 42(1), 30-35. doi:10.1038/ng.499
- Ng, Y. S., Grady, J. P., Lax, N. Z., Bourke, J. P., Alston, C. L., Hardy, S. A., . . . Gorman, G. S. (2016). Sudden adult death syndrome in m.3243A>G-related mitochondrial disease: an unrecognized clinical entity in young, asymptomatic adults. *Eur Heart J*, 37(32), 2552-2559. doi:10.1093/eurheartj/ehv306
- Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., . . . Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461(7261), 272-U153. doi:10.1038/nature08250
- Nguyen, Q. H., Pervolarakis, N., Nee, K., & Kessenbrock, K. (2018). Experimental Considerations for Single-Cell RNA Sequencing Approaches. *Front Cell Dev Biol*, 6, 108. doi:10.3389/fcell.2018.00108
- Nishino, I., Spinazzola, A., & Hirano, M. (1999). Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science*, 283(5402), 689-692. doi:10.1126/science.283.5402.689
- Norton, N., Robertson, P. D., Rieder, M. J., Zuchner, S., Rampersaud, E., Martin, E., . . . Ex, N. H. L. B. I. G. (2012). Evaluating Pathogenicity of Rare Variants From Dilated Cardiomyopathy in the Exome Era. *Circulation-Cardiovascular Genetics*, 5(2), 167-174. doi:10.1161/Circgenetics.111.961805
- Oetjens, M. T., Kelly, M. A., Sturm, A. C., Martin, C. L., & Ledbetter, D. H. (2019). Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. *Nature Communications*, 10(1), 4897. doi:10.1038/s41467-019-12869-0
- Ohtake, A., Murayama, K., Mori, M., Harashima, H., Yamazaki, T., Tamaru, S., . . . Okazaki, Y. (2014). Diagnosis and molecular basis of mitochondrial respiratory chain disorders: Exome sequencing for disease gene identification. *Biochimica Et Biophysica Acta-General Subjects*, 1840(4), 1355-1359. doi:10.1016/j.bbagen.2014.01.025
- Ojala, D., Montoya, J., & Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature*, 290(5806), 470-474. doi:10.1038/290470a0
- Olsen, R. K., Olpin, S. E., Andresen, B. S., Miedzybrodzka, Z. H., Pourfarzam, M., Merinero, B., . . . Morris, A. A. (2007). ETFDH mutations as a major cause of riboflavin-responsive multiple acyl-CoA dehydrogenation deficiency. *Brain*, 130(Pt 8), 2045-2054. doi:10.1093/brain/awm135
- Ostergaard, E., Christensen, E., Kristensen, E., Mogensen, B., Duno, M., Shoubridge, E. A., & Wibrand, F. (2007). Deficiency of the alpha subunit of succinate-coenzyme A ligase causes fatal infantile lactic acidosis with mitochondrial DNA depletion. *Am J Hum Genet*, 81(2), 383-387. doi:10.1086/519222
- Ottolenghi, C., Hubert, L., Allanore, Y., Brassier, A., Altuzarra, C., Mellot-Draznieks, C., . . . de Lonlay, P. (2011). Clinical and Biochemical Heterogeneity Associated with Fumarase Deficiency. *Human Mutation*, 32(9), 1046-1052. doi:10.1002/humu.21534

- Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*, *12*(2), 87-98. doi:10.1038/nrg2934
- Pagani, F., Raponi, M., & Baralle, F. E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc Natl Acad Sci U S A*, *102*(18), 6368-6372. doi:10.1073/pnas.0502288102
- Pakendorf, B., & Stoneking, M. (2005). Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet*, *6*, 165-183. doi:10.1146/annurev.genom.6.080604.162249
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, *40*(12), 1413-1415. doi:10.1038/ng.259
- Panneman, D. M., Smeitink, J. A., & Rodenburg, R. J. (2018). Mining for mitochondrial mechanisms: Linking known syndromes to mitochondrial function. *Clinical Genetics*, *93*(5), 943-951. doi:10.1111/cge.13094
- Papatheodorou, I., Fonseca, N. A., Keays, M., Tang, Y. A., Barrera, E., Bazant, W., . . . Petryszak, R. (2018). Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res*, *46*(D1), D246-D251. doi:10.1093/nar/gkx1158
- Parikh, S., Karaa, A., Goldstein, A., Ng, Y. S., Gorman, G., Feigenbaum, A., . . . Scaglia, F. (2016). Solid organ transplantation in primary mitochondrial disease: Proceed with caution. *Mol Genet Metab*, *118*(3), 178-184. doi:10.1016/j.ymgme.2016.04.009
- Pasca, L., De Giorgis, V., Macasaet, J. A., Trentani, C., Tagliabue, A., & Veggiotti, P. (2016). The changing face of dietary therapy for epilepsy. *Eur J Pediatr*, *175*(10), 1267-1276. doi:10.1007/s00431-016-2765-z
- Payne, B. A., & Chinnery, P. F. (2015). Mitochondrial dysfunction in aging: Much progress but many unresolved questions. *Biochim Biophys Acta*, *1847*(11), 1347-1353. doi:10.1016/j.bbabi.2015.05.022
- Payne, B. A., Wilson, I. J., Yu-Wai-Man, P., Coxhead, J., Deehan, D., Horvath, R., . . . Chinnery, P. F. (2013). Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet*, *22*(2), 384-390. doi:10.1093/hmg/ddc435
- Petersen, B. S., Fredrich, B., Hoepfner, M. P., Ellinghaus, D., & Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. *Bmc Genetics*, *18*. doi:10.1186/s12863-017-0479-5
- Pfanner, N., Warscheid, B., & Wiedemann, N. (2019). Mitochondrial proteins: from biogenesis to functional networks. *Nat Rev Mol Cell Biol*, *20*(5), 267-284. doi:10.1038/s41580-018-0092-0
- Pirinen, E., Auranen, M., Khan, N. A., Brillhante, V., Urho, N., Pessia, A., . . . Suomalainen, A. (2020). Niacin Cures Systemic NAD(+) Deficiency and Improves Muscle Performance in Adult-Onset Mitochondrial Myopathy. *Cell Metab*, *31*(6), 1078-1090 e1075. doi:10.1016/j.cmet.2020.04.008
- Plagnol, V., Curtis, J., Epstein, M., Mok, K. Y., Stebbings, E., Grigoriadou, S., . . . Nejentsev, S. (2012). A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*, *28*(21), 2747-2754. doi:10.1093/bioinformatics/bts526
- Popp, M. W., & Maquat, L. E. (2013). Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet*, *47*, 139-165. doi:10.1146/annurev-genet-111212-133424
- Pronicka, E., Piekutowska-Abramczuk, D., Ciara, E., Trubicka, J., Rokicki, D., Karkucinska-Wieckowska, A., . . . Ploski, R. (2016). New perspective in diagnostics of

- mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric centre. *Journal of Translational Medicine*, 14. doi:10.1186/s12967-016-0930-9
- Protasoni, M., Perez-Perez, R., Lobo-Jarne, T., Harbour, M. E., Ding, S., Penas, A., . . . Fernandez-Vizarra, E. (2020). Respiratory supercomplexes act as a platform for complex III-mediated maturation of human mitochondrial complexes I and IV. *EMBO J*, 39(3), e102817. doi:10.15252/embj.2019102817
- Puchalska, P., & Crawford, P. A. (2017). Multi-dimensional Roles of Ketone Bodies in Fuel Metabolism, Signaling, and Therapeutics. *Cell Metab*, 25(2), 262-284. doi:10.1016/j.cmet.2016.12.022
- Puusepp, S., Reinson, K., Pajusalu, S., Murumets, U., Oiglane-Shlik, E., Rein, R., . . . Ounap, K. (2018). Effectiveness of whole exome sequencing in unsolved patients with a clinical suspicion of a mitochondrial disorder in Estonia. *Mol Genet Metab Rep*, 15, 80-89. doi:10.1016/j.ymgmr.2018.03.004
- Rahman, J., Noronha, A., Thiele, I., & Rahman, S. (2017). Leigh map: A novel computational diagnostic resource for mitochondrial disease. *Ann Neurol*, 81(1), 9-16. doi:10.1002/ana.24835
- Rahman, S., Footitt, E. J., Varadkar, S., & Clayton, P. T. (2013). Inborn errors of metabolism causing epilepsy. *Dev Med Child Neurol*, 55(1), 23-36. doi:10.1111/j.1469-8749.2012.04406.x
- Raimundo, N., Baysal, B. E., & Shadel, G. S. (2011). Revisiting the TCA cycle: signaling to tumor formation. *Trends Mol Med*, 17(11), 641-649. doi:10.1016/j.molmed.2011.06.001
- Ramos, A., Santos, C., Mateiu, L., Gonzalez Mdel, M., Alvarez, L., Azevedo, L., . . . Aluja, M. P. (2013). Frequency and pattern of heteroplasmy in the complete human mitochondrial genome. *PLoS One*, 8(10), e74636. doi:10.1371/journal.pone.0074636
- Reese, M. G., Eeckman, F. H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in Genie. *J Comput Biol*, 4(3), 311-323. doi:10.1089/cmb.1997.4.311
- Rehm, H. L., Berg, J. S., Brooks, L. D., Bustamante, C. D., Evans, J. P., Landrum, M. J., . . . ClinGen. (2015). ClinGen--the Clinical Genome Resource. *N Engl J Med*, 372(23), 2235-2242. doi:10.1056/NEJMSr1406261
- Reinecke, F., Smeitink, J. A., & van der Westhuizen, F. H. (2009). OXPHOS gene expression and control in mitochondrial disorders. *Biochim Biophys Acta*, 1792(12), 1113-1121. doi:10.1016/j.bbadis.2009.04.003
- Reinius, B., & Sandberg, R. (2015). Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nat Rev Genet*, 16(11), 653-664. doi:10.1038/nrg3888
- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1), D886-D894. doi:10.1093/nar/gky1016
- Repp, B. M., Mastantuono, E., Alston, C. L., Schiff, M., Haack, T. B., Rotig, A., . . . Wortmann, S. (2018). Clinical, biochemical and genetic spectrum of 70 patients with ACAD9 deficiency: is riboflavin supplementation effective? *Orphanet J Rare Dis*, 13(1), 120. doi:10.1186/s13023-018-0784-8
- Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., . . . Bale, S. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genetics in Medicine*, 18(7), 696-704. doi:10.1038/gim.2015.148

- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405-424. doi:10.1038/gim.2015.30
- Riley, L. G., Cowley, M. J., Gayevskiy, V., Minoche, A. E., Puttick, C., Thorburn, D. R., . . . Christodoulou, J. (2020). The diagnostic utility of genome sequencing in a pediatric cohort with suspected mitochondrial disease. *Genetics in Medicine*. doi:10.1038/s41436-020-0793-6
- Rivas, M. A., Pirinen, M., Conrad, D. F., Lek, M., Tsang, E. K., Karczewski, K. J., . . . MacArthur, D. G. (2015). Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science*, *348*(6235), 666-669. doi:10.1126/science.1261877
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-140. doi:10.1093/bioinformatics/btp616
- Robinson, P. N., Kohler, S., Oellrich, A., Wang, K., Mungall, C. J., Lewis, S. E., . . . Project, S. M. G. (2014). Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Research*, *24*(2), 340-348. doi:10.1101/gr.160325.113
- Rocha, M. C., Grady, J. P., Grunewald, A., Vincent, A., Dobson, P. F., Taylor, R. W., . . . Rygiel, K. A. (2015). A novel immunofluorescent assay to investigate oxidative phosphorylation deficiency in mitochondrial myopathy: understanding mechanisms and improving diagnosis. *Sci Rep*, *5*, 15037. doi:10.1038/srep15037
- Rockowitz, S., LeCompte, N., Carmack, M., Quitadamo, A., Wang, L., Park, M., . . . Sliz, P. (2020). Children's rare disease cohorts: an integrative research and clinical genomics initiative. *NPJ Genom Med*, *5*, 29. doi:10.1038/s41525-020-0137-0
- Rodenburg, R. J. (2011). Biochemical diagnosis of mitochondrial disorders. *J Inherit Metab Dis*, *34*(2), 283-292. doi:10.1007/s10545-010-9081-y
- Rogers, G. W., Brand, M. D., Petrosyan, S., Ashok, D., Elorza, A. A., Ferrick, D. A., & Murphy, A. N. (2011). High throughput microplate respiratory measurements using minimal quantities of isolated mitochondria. *PLoS One*, *6*(7), e21746. doi:10.1371/journal.pone.0021746
- Rosignol, R., Faustin, B., Rocher, C., Malgat, M., Mazat, J. P., & Letellier, T. (2003). Mitochondrial threshold effects. *Biochem J*, *370*(Pt 3), 751-762. doi:10.1042/BJ20021594
- Rowe, R. G., & Daley, G. Q. (2019). Induced pluripotent stem cells in disease modelling and drug discovery. *Nat Rev Genet*, *20*(7), 377-388. doi:10.1038/s41576-019-0100-z
- Russell, O. M., Gorman, G. S., Lightowlers, R. N., & Turnbull, D. M. (2020). Mitochondrial Diseases: Hope for the Future. *Cell*, *181*(1), 168-188. doi:10.1016/j.cell.2020.02.051
- Rustin, P., Chretien, D., Bourgeron, T., Wucher, A., Saudubray, J. M., Rotig, A., & Munnich, A. (1991). Assessment of the mitochondrial respiratory chain. *Lancet*, *338*(8758), 60. doi:10.1016/0140-6736(91)90057-v
- Rustin, P., Bourgeron, T., Parfait, B., Chretien, D., Munnich, A., & Rotig, A. (1997). Inborn errors of the Krebs cycle: a group of unusual mitochondrial diseases in human. *Biochim Biophys Acta*, *1361*(2), 185-197. doi:10.1016/s0925-4439(97)00035-5
- Sagan, L. (1967). On the origin of mitosing cells. *J Theor Biol*, *14*(3), 255-274. doi:10.1016/0022-5193(67)90079-3

- Sahni, N., Yi, S., Taipale, M., Fuxman Bass, J. I., Coulombe-Huntington, J., Yang, F., . . . Vidal, M. (2015). Widespread macromolecular interaction perturbations in human genetic disorders. *Cell*, *161*(3), 647-660. doi:10.1016/j.cell.2015.04.013
- Sajjani, K., Islam, F., Smith, R. A., Gopalan, V., & Lam, A. K. (2017). Genetic alterations in Krebs cycle and its impact on cancer pathogenesis. *Biochimie*, *135*, 164-172. doi:10.1016/j.biochi.2017.02.008
- Sallevelt, S. C. E. H., de Die-Smulders, C. E. M., Hendrickx, A. T. M., Hellebrekers, D. M. E. I., de Coo, I. F. M., Alston, C. L., . . . Smeets, H. J. M. (2017). De novo mtDNA point mutations are common and have a low recurrence risk. *Journal of Medical Genetics*, *54*(2), 114-124. doi:10.1136/jmedgenet-2016-103876
- Sanchez, E., Lobo, T., Fox, J. L., Zeviani, M., Winge, D. R., & Fernandez-Vizarra, E. (2013). LYRM7/MZM1L is a UQCERS1 chaperone involved in the last steps of mitochondrial Complex III assembly in human cells. *Biochim Biophys Acta*, *1827*(3), 285-293. doi:10.1016/j.bbabi.2012.11.003
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, *74*(12), 5463-5467. doi:10.1073/pnas.74.12.5463
- Sato, M., & Sato, K. (2013). Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochim Biophys Acta*, *1833*(8), 1979-1984. doi:10.1016/j.bbamcr.2013.03.010
- Scaglia, F., Towbin, J. A., Craigen, W. J., Belmont, J. W., Smith, E. O., Neish, S. R., . . . Vogel, H. (2004). Clinical spectrum, morbidity, and mortality in 113 pediatric patients with mitochondrial disease. *Pediatrics*, *114*(4), 925-931. doi:10.1542/peds.2004-0718
- Schagger, H., & Pfeiffer, K. (2000). Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J*, *19*(8), 1777-1783. doi:10.1093/emboj/19.8.1777
- Schagger, H., & Pfeiffer, K. (2001). The ratio of oxidative phosphorylation complexes I-V in bovine heart mitochondria and the composition of respiratory chain supercomplexes. *J Biol Chem*, *276*(41), 37861-37867. doi:10.1074/jbc.M106474200
- Schagger, H., & von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem*, *199*(2), 223-231. doi:10.1016/0003-2697(91)90094-a
- Schmidt, O., Pfanner, N., & Meisinger, C. (2010). Mitochondrial protein import: from proteomics to functional mechanisms. *Nat Rev Mol Cell Biol*, *11*(9), 655-667. doi:10.1038/nrm2959
- Schneeberger, P. E., Bierhals, T., Neu, A., Hempel, M., & Kutsche, K. (2019). de novo MEPCE nonsense variant associated with a neurodevelopmental disorder causes disintegration of 7SK snRNP and enhanced RNA polymerase II activation. *Sci Rep*, *9*(1), 12516. doi:10.1038/s41598-019-49032-0
- Schon, E. A., DiMauro, S., & Hirano, M. (2012). Human mitochondrial DNA: roles of inherited and somatic mutations. *Nat Rev Genet*, *13*(12), 878-890. doi:10.1038/nrg3275
- Schuelke, M., Krude, H., Finckh, B., Mayatepek, E., Janssen, A., Schmelz, M., . . . Smeitink, J. (2002). Septo-optic dysplasia associated with a new mitochondrial cytochrome b mutation. *Ann Neurol*, *51*(3), 388-392. doi:10.1002/ana.10151
- Schwarz, J. M., Rodelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*, *7*(8), 575-576. doi:10.1038/nmeth0810-575

- Sciacco, M., & Bonilla, E. (1996). Cytochemistry and immunocytochemistry of mitochondria in tissue sections. *Methods Enzymol*, 264, 509-521. doi:10.1016/s0076-6879(96)64045-2
- Scocchia, A., Wigby, K. M., Masser-Frye, D., Del Campo, M., Galarreta, C. I., Thorpe, E., . . . Team, I. I. R. (2019). Clinical whole genome sequencing as a first-tier test at a resource-limited dysmorphology clinic in Mexico. *Npj Genomic Medicine*, 4. doi:10.1038/s41525-018-0076-1
- Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat Rev Genet*, 17(1), 19-32. doi:10.1038/nrg.2015.3
- Sedlazeck, F. J., Lee, H., Darby, C. A., & Schatz, M. C. (2018). Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Rev Genet*, 19(6), 329-346. doi:10.1038/s41576-018-0003-4
- Sena, L. A., & Chandel, N. S. (2012). Physiological roles of mitochondrial reactive oxygen species. *Mol Cell*, 48(2), 158-167. doi:10.1016/j.molcel.2012.09.025
- Servidei, S. (2001). Mitochondrial encephalomyopathies: gene mutation. *Neuromuscul Disord*, 11(2), 230-235. doi:10.1016/s0960-8966(01)00205-x
- Shaham, O., Slate, N. G., Goldberger, O., Xu, Q., Ramanathan, A., Souza, A. L., . . . Mootha, V. K. (2010). A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells. *Proc Natl Acad Sci U S A*, 107(4), 1571-1575. doi:10.1073/pnas.0906039107
- Shapiro, M. B., & Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res*, 15(17), 7155-7174. doi:10.1093/nar/15.17.7155
- Short, P. J., McRae, J. F., Gallone, G., Sifrim, A., Won, H., Geschwind, D. H., . . . Hurles, M. E. (2018). De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature*, 555(7698), 611-616. doi:10.1038/nature25983
- Signes, A., & Fernandez-Vizarra, E. (2018). Assembly of mammalian oxidative phosphorylation complexes I-V and supercomplexes. *Essays Biochem*, 62(3), 255-270. doi:10.1042/EBC20170098
- Simon, M. T., Eftekharian, S. S., Stover, A. E., Osborne, A. F., Braffman, B. H., Chang, R. C., . . . Abdenur, J. E. (2019). Novel mutations in the mitochondrial complex I assembly gene NDUFAF5 reveal heterogeneous phenotypes. *Mol Genet Metab*, 126(1), 53-63. doi:10.1016/j.ymgme.2018.11.001
- Singh, R. K., & Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med*, 18(8), 472-482. doi:10.1016/j.molmed.2012.06.006
- Skjorringe, T., Tumer, Z., & Moller, L. B. (2011). Splice site mutations in the ATP7A gene. *PLoS One*, 6(4), e18599. doi:10.1371/journal.pone.0018599
- Smeitink, J. A. (2003). Mitochondrial disorders: clinical presentation and diagnostic dilemmas. *J Inherit Metab Dis*, 26(2-3), 199-207. doi:10.1023/a:1024489218004
- Smeitink, J. A., Zeviani, M., Turnbull, D. M., & Jacobs, H. T. (2006). Mitochondrial medicine: a metabolic perspective on the pathology of oxidative phosphorylation disorders. *Cell Metab*, 3(1), 9-13. doi:10.1016/j.cmet.2005.12.001
- Smith, H. S., Swint, J. M., Lalani, S. R., Yamal, J. M., Otto, M. C. D., Castellanos, S., . . . Russell, H. V. (2019). Clinical Application of Genome and Exome Sequencing as a Diagnostic Tool for Pediatric Patients: a Scoping Review of the Literature. *Genetics in Medicine*, 21(1), 3-16. doi:10.1038/s41436-018-0024-6

- Smith, P. M., Fox, J. L., & Winge, D. R. (2012). Biogenesis of the cytochrome bc(1) complex and role of assembly factors. *Biochim Biophys Acta*, *1817*(2), 276-286. doi:10.1016/j.bbabi.2011.11.009
- Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., . . . Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet*, *84*(6), 740-759. doi:10.1016/j.ajhg.2009.05.001
- Sobreira, N., Schiettecatte, F., Valle, D., & Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat*, *36*(10), 928-930. doi:10.1002/humu.22844
- Soemedi, R., Cygan, K. J., Rhine, C. L., Wang, J., Bulacan, C., Yang, J., . . . Fairbrother, W. G. (2017). Pathogenic variants that alter protein code often disrupt splicing. *Nat Genet*, *49*(6), 848-855. doi:10.1038/ng.3837
- Sosnay, P. R., & Cutting, G. R. (2014). Interpretation of genetic variants. *Thorax*, *69*(3), 295-297. doi:10.1136/thoraxjnl-2013-204903
- Spiegel, K., Pines, O., Ta-Shma, A., Burak, E., Shaag, A., Halvardson, J., . . . Elpeleg, O. (2012). Infantile Cerebellar-Retinal Degeneration Associated with a Mutation in Mitochondrial Aconitase, ACO2. *American Journal of Human Genetics*, *90*(3), 518-523. doi:10.1016/j.ajhg.2012.01.009
- Spinazzi, M., Casarin, A., Pertegato, V., Salviati, L., & Angelini, C. (2012). Assessment of mitochondrial respiratory chain enzymatic activities on tissues and cultured cells. *Nature Protocols*, *7*(6), 1235-1246. doi:10.1038/nprot.2012.058
- Spinelli, J. B., & Haigis, M. C. (2018). The multifaceted contributions of mitochondria to cellular metabolism. *Nat Cell Biol*, *20*(7), 745-754. doi:10.1038/s41556-018-0124-1
- Stavropoulos, D. J., Merico, D., Jobling, R., Bowdin, S., Monfared, N., Thiruvahindrapuram, B., . . . Marshall, C. R. (2016). Whole Genome Sequencing Expands Diagnostic Utility and Improves Clinical Management in Pediatric Medicine. *NPJ Genom Med*, *1*. doi:10.1038/npjgenmed.2015.12
- Steffann, J., Gigarel, N., Corcos, J., Bonniere, M., Encha-Razavi, F., Sinico, M., . . . Bonnefont, J. P. (2007). Stability of the m.8993T->G mtDNA mutation load during human embryofetal development has implications for the feasibility of prenatal diagnosis in NARP syndrome. *J Med Genet*, *44*(10), 664-669. doi:10.1136/jmg.2006.048553
- Stein, C. S., Jadya, P., Zhang, X., McLendon, J. M., Abouassaly, G. M., Witmer, N. H., . . . Boudreau, R. L. (2018). Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. *Cell Rep*, *23*(13), 3710-3720 e3718. doi:10.1016/j.celrep.2018.06.002
- Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S., . . . Cooper, D. N. (2003). Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat*, *21*(6), 577-581. doi:10.1002/humu.10212
- Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., . . . Cooper, D. N. (2017). The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*, *136*(6), 665-677. doi:10.1007/s00439-017-1779-6
- Stenton, S. L., Kremer, L. S., Kopajtich, R., Ludwig, C., & Prokisch, H. (2020). The diagnosis of inborn errors of metabolism by an integrative "multi-omics" approach: A perspective encompassing genomics, transcriptomics, and proteomics. *J Inherit Metab Dis*, *43*(1), 25-35. doi:10.1002/jimd.12130

- Stenton, S. L., & Prokisch, H. (2018). Advancing genomic approaches to the molecular diagnosis of mitochondrial disease. *Essays Biochem*, 62(3), 399-408. doi:10.1042/EBC20170110
- Stenton, S. L., & Prokisch, H. (2020). Genetics of mitochondrial diseases: Identifying mutations to help diagnosis. *EBioMedicine*, 56, 102784. doi:10.1016/j.ebiom.2020.102784
- Sternecker, J. L., Reinhardt, P., & Scholer, H. R. (2014). Investigating human disease using stem cell models. *Nat Rev Genet*, 15(9), 625-639. doi:10.1038/nrg3764
- Stroud, D. A., Surgenor, E. E., Formosa, L. E., Reljic, B., Frazier, A. E., Dibley, M. G., . . . Ryan, M. T. (2016). Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature*, 538(7623), 123-126. doi:10.1038/nature19754
- Sue, C. M., Quigley, A., Katsabanis, S., Kapsa, R., Crimmins, D. S., Byrne, E., & Morris, J. G. (1998). Detection of MELAS A3243G point mutation in muscle, blood and hair follicles. *J Neurol Sci*, 161(1), 36-39. doi:10.1016/s0022-510x(98)00179-8
- Sulonen, A. M., Ellonen, P., Almusa, H., Lepisto, M., Eldfors, S., Hannula, S., . . . Saarela, J. (2011). Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol*, 12(9), R94. doi:10.1186/gb-2011-12-9-r94
- Suomalainen, A., & Battersby, B. J. (2018). Mitochondrial diseases: the contribution of organelle stress responses to pathology. *Nat Rev Mol Cell Biol*, 19(2), 77-92. doi:10.1038/nrm.2017.66
- Suomalainen, A., Elo, J. M., Pietilainen, K. H., Hakonen, A. H., Sevastianova, K., Korpela, M., . . . Tynismaa, H. (2011). FGF-21 as a biomarker for muscle-manifesting mitochondrial respiratory chain deficiencies: a diagnostic study. *Lancet Neurol*, 10(9), 806-818. doi:10.1016/S1474-4422(11)70155-7
- Surendran, S., Michals-Matalon, K., Krywawych, S., Qazi, Q. H., Tuchman, R., Rady, P. L., . . . Matalon, R. (2002). DOOR syndrome: deficiency of E1 component of the 2-oxoglutarate dehydrogenase complex. *Am J Med Genet*, 113(4), 371-374. doi:10.1002/ajmg.b.10804
- Szabo, P. A., Levitin, H. M., Miron, M., Snyder, M. E., Senda, T., Yuan, J., . . . Sims, P. A. (2019). Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications*, 10(1), 4706. doi:10.1038/s41467-019-12464-3
- Szeto, H. H. (2014). First-in-class cardiolipin-protective compound as a therapeutic agent to restore mitochondrial bioenergetics. *Br J Pharmacol*, 171(8), 2029-2050. doi:10.1111/bph.12461
- Tan, J., Wagner, M., Stenton, S. L., Strom, T. M., Wortmann, S. B., Prokisch, H., . . . Klopstock, T. (2020). Lifetime risk of autosomal recessive mitochondrial disorders calculated from genetic databases. *EBioMedicine*, 54, 102730. doi:10.1016/j.ebiom.2020.102730
- Tan, R. J., Wang, Y. D., Kleinstein, S. E., Liu, Y. Z., Zhu, X. L., Guo, H. Z., . . . Zhu, M. F. (2014). An Evaluation of Copy Number Variation Detection Tools from Whole-Exome Sequencing Data. *Human Mutation*, 35(7), 899-907. doi:10.1002/humu.22537
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., . . . Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*, 6(5), 377-382. doi:10.1038/nmeth.1315

- Tang, S., Wang, J., Zhang, V. W., Li, F. Y., Landsverk, M., Cui, H., . . . Wong, L. J. (2013). Transition to next generation analysis of the whole mitochondrial genome: a summary of molecular defects. *Hum Mutat*, *34*(6), 882-893. doi:10.1002/humu.22307
- Tarnopolsky, M. A. (2011). Creatine as a therapeutic strategy for myopathies. *Amino Acids*, *40*(5), 1397-1407. doi:10.1007/s00726-011-0876-4
- Taylor, J. C., Martin, H. C., Lise, S., Broxholme, J., Cazier, J. B., Rimmer, A., . . . McVean, G. (2015). Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nature Genetics*, *47*(7), 717+. doi:10.1038/ng.3304
- Taylor, R. W., Pyle, A., Griffin, H., Blakely, E. L., Duff, J., He, L. P., . . . Chinnery, P. F. (2014). Use of Whole-Exome Sequencing to Determine the Genetic Basis of Multiple Mitochondrial Respiratory Chain Complex Deficiencies. *Jama-Journal of the American Medical Association*, *312*(1), 68-77. doi:10.1001/jama.2014.7184
- Taylor, R. W., Schaefer, A. M., Barron, M. J., McFarland, R., & Turnbull, D. M. (2004). The diagnosis of mitochondrial muscle disease. *Neuromuscul Disord*, *14*(4), 237-245. doi:10.1016/j.nmd.2003.12.004
- Theunissen, T. E. J., Nguyen, M., Kamps, R., Hendrickx, A. T., Sallevelt, S., Gottschalk, R. W. H., . . . Smeets, H. J. M. (2018). Whole Exome Sequencing Is the Preferred Strategy to Identify the Genetic Defect in Patients With a Probable or Possible Mitochondrial Cause. *Frontiers in Genetics*, *9*, 400. doi:10.3389/fgene.2018.00400
- Thiffault, I., Farrow, E., Zellmer, L., Berrios, C., Miller, N., Gibson, M., . . . Saunders, C. (2019). Clinical genome sequencing in an unbiased pediatric cohort. *Genetics in Medicine*, *21*(2), 303-310. doi:10.1038/s41436-018-0075-8
- Thompson, K., Majd, H., Dallabona, C., Reinson, K., King, M. S., Alston, C. L., . . . Taylor, R. W. (2016). Recurrent De Novo Dominant Mutations in SLC25A4 Cause Severe Early-Onset Mitochondrial Disease and Loss of Mitochondrial DNA Copy Number (vol 99, pg 860, 2016). *American Journal of Human Genetics*, *99*(6), 1405-1405. doi:10.1016/j.ajhg.2016.11.001
- Thusberg, J., Olatubosun, A., & Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat*, *32*(4), 358-368. doi:10.1002/humu.21445
- Thusberg, J., & Vihinen, M. (2009). Pathogenic or Not? And If So, Then How? Studying the Effects of Missense Mutations Using Bioinformatics Methods. *Human Mutation*, *30*(5), 703-714. doi:10.1002/humu.20938
- Tilokani, L., Nagashima, S., Paupe, V., & Prudent, J. (2018). Mitochondrial dynamics: overview of molecular mechanisms. *Essays Biochem*, *62*(3), 341-360. doi:10.1042/EBC20170104
- Timmis, J. N., Ayliffe, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet*, *5*(2), 123-135. doi:10.1038/nrg1271
- Tiranti, V., Viscomi, C., Hildebrandt, T., Di Meo, I., Mineri, R., Tiveron, C., . . . Zeviani, M. (2009). Loss of ETHE1, a mitochondrial dioxygenase, causes fatal sulfide toxicity in ethylmalonic encephalopathy. *Nat Med*, *15*(2), 200-205. doi:10.1038/nm.1907
- Tomlinson, I. P., Alam, N. A., Rowan, A. J., Barclay, E., Jaeger, E. E., Kelsell, D., . . . Multiple Leiomyoma, C. (2002). Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat Genet*, *30*(4), 406-410. doi:10.1038/ng849
- Truty, R., Paul, J., Kennemer, M., Lincoln, S. E., Olivares, E., Nussbaum, R. L., & Aradhya, S. (2019). Prevalence and properties of intragenic copy-number variation in

- Mendelian disease genes. *Genetics in Medicine*, 21(1), 114-123. doi:10.1038/s41436-018-0033-5
- Tucker, E. J., Wanschers, B. F., Szklarczyk, R., Mountford, H. S., Wijeyeratne, X. W., van den Brand, M. A., . . . Thorburn, D. R. (2013). Mutations in the UQCC1-interacting protein, UQCC2, cause human complex III deficiency associated with perturbed cytochrome b protein expression. *PLoS Genet*, 9(12), e1004034. doi:10.1371/journal.pgen.1004034
- Turro, E., Astle, W. J., Megy, K., Graf, S., Greene, D., Shamardina, O., . . . Ouwehand, W. H. (2020). Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*, 583(7814), 96-102. doi:10.1038/s41586-020-2434-2
- Valnot, I., Kassis, J., Chretien, D., de Lonlay, P., Parfait, B., Munnich, A., . . . Rotig, A. (1999). A mitochondrial cytochrome b mutation but no mutations of nuclearly encoded subunits in ubiquinol cytochrome c reductase (complex III) deficiency. *Hum Genet*, 104(6), 460-466. doi:10.1007/s004390050988
- van Berge, L., Hamilton, E. M., Linnankivi, T., Uziel, G., Steenweg, M. E., Isohanni, P., . . . Group, L. R. (2014). Leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation: clinical and genetic characterization and target for therapy. *Brain*, 137(Pt 4), 1019-1029. doi:10.1093/brain/awu026
- van de Weijer, T., Phielix, E., Bilet, L., Williams, E. G., Ropelle, E. R., Bierwagen, A., . . . Schrauwen, P. (2015). Evidence for a direct effect of the NAD⁺ precursor acipimox on muscle mitochondrial function in humans. *Diabetes*, 64(4), 1193-1201. doi:10.2337/db14-0667
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), 418-426. doi:10.1016/j.tig.2014.07.001
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends Genet*, 34(9), 666-681. doi:10.1016/j.tig.2018.05.008
- van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J. F., Adami, E., Faber, A. B., . . . Hubner, N. (2019). The Translational Landscape of the Human Heart. *Cell*, 178(1), 242-260 e229. doi:10.1016/j.cell.2019.05.010
- Van Hout, C. V., Tachmazidou, I., Backman, J. D., Hoffman, J. X., Ye, B., Pandey, A. K., . . . Baras, A. (2019). Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv*. doi:https://doi.org/10.1101/572347
- Vaz-Drago, R., Custodio, N., & Carmo-Fonseca, M. (2017). Deep intronic mutations and human disease. *Hum Genet*, 136(9), 1093-1111. doi:10.1007/s00439-017-1809-4
- Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., . . . Kriegstein, A. R. (2019). Single-cell genomics identifies cell type-specific molecular changes in autism. *Science*, 364(6441), 685-689. doi:10.1126/science.aav8130
- Veltman, J. A., & Brunner, H. G. (2012). De novo mutations in human genetic disease. *Nat Rev Genet*, 13(8), 565-575. doi:10.1038/nrg3241
- Visapaa, I., Fellman, V., Vesa, J., Dasvarma, A., Hutton, J. L., Kumar, V., . . . Peltonen, L. (2002). GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. *Am J Hum Genet*, 71(4), 863-876. doi:10.1086/342773
- Volk, A. E., & Kubisch, C. (2017). The rapid evolution of molecular genetic diagnostics in neuromuscular diseases. *Current Opinion in Neurology*, 30(5), 523-528. doi:10.1097/Wco.0000000000000478

- von Heijne, G. (1989). The structure of signal peptides from bacterial lipoproteins. *Protein Eng*, 2(7), 531-534. doi:10.1093/protein/2.7.531
- Wagner, M., Berutti, R., Lorenz-Depiereux, B., Graf, E., Eckstein, G., Mayr, J. A., . . . Wortmann, S. B. (2019). Mitochondrial DNA mutation analysis from exome sequencing-A more holistic approach in diagnostics of suspected mitochondrial disease. *J Inherit Metab Dis*, 42(5), 909-917. doi:10.1002/jimd.12109
- Wai, H. A., Lord, J., Lyon, M., Gunning, A., Kelly, H., Cibin, P., . . . Baralle, D. (2020). Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genetics in Medicine*. doi:10.1038/s41436-020-0766-9
- Wallace, D. C. (1992). Diseases of the mitochondrial DNA. *Annu Rev Biochem*, 61, 1175-1212. doi:10.1146/annurev.bi.61.070192.005523
- Wallace, D. C., Singh, G., Lott, M. T., Hodge, J. A., Schurr, T. G., Lezza, A. M., . . . Nikoskelainen, E. K. (1988). Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science*, 242(4884), 1427-1430. doi:10.1126/science.3201231
- Wang, D., Eraslan, B., Wieland, T., Hallstrom, B., Hopf, T., Zolg, D. P., . . . Kuster, B. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol*, 15(2), e8503. doi:10.15252/msb.20188503
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., . . . Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), 470-476. doi:10.1038/nature07509
- Wang, G. S., & Cooper, T. A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet*, 8(10), 749-761. doi:10.1038/nrg2164
- Wang, J., Fan, H. C., Behr, B., & Quake, S. R. (2012). Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell*, 150(2), 402-412. doi:10.1016/j.cell.2012.06.030
- Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, 38(16), e164. doi:10.1093/nar/gkq603
- Wang, Q., Pierce-Hoffman, E., Cummings, B. B., Alfoldi, J., Francioli, L. C., Gauthier, L. D., . . . MacArthur, D. G. (2020). Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. *Nature Communications*, 11(1), 2539. doi:10.1038/s41467-019-12438-5
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10(1), 57-63. doi:10.1038/nrg2484
- Wang, Z., Liu, X., Yang, B. Z., & Gelernter, J. (2013). The role and challenges of exome sequencing in studies of human diseases. *Frontiers in Genetics*, 4, 160. doi:10.3389/fgene.2013.00160
- Wanschers, B. F., Szklarczyk, R., van den Brand, M. A., Jonckheere, A., Suijskens, J., Smeets, R., . . . Huynen, M. A. (2014). A mutation in the human CBP4 ortholog UQCC3 impairs complex III assembly, activity and cytochrome b stability. *Hum Mol Genet*, 23(23), 6356-6365. doi:10.1093/hmg/ddu357
- Ward, J. P., & McMurtry, I. F. (2009). Mechanisms of hypoxic pulmonary vasoconstriction and their roles in pulmonary hypertension: new findings for an old problem. *Curr Opin Pharmacol*, 9(3), 287-296. doi:10.1016/j.coph.2009.02.006
- Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., & Watson, M. (2015). Exome Sequencing: Current and Future Perspectives. *G3 (Bethesda)*, 5(8), 1543-1550. doi:10.1534/g3.115.018564

- Watson, M., & Warr, A. (2019). Errors in long-read assemblies can critically affect protein prediction. *Nat Biotechnol*, *37*(2), 124-126. doi:10.1038/s41587-018-0004-z
- Waypa, G. B., Marks, J. D., Guzy, R. D., Mungai, P. T., Schriewer, J. M., Dokic, D., . . . Schumacker, P. T. (2013). Superoxide generated at mitochondrial complex III triggers acute responses to hypoxia in the pulmonary circulation. *Am J Respir Crit Care Med*, *187*(4), 424-432. doi:10.1164/rccm.201207-1294OC
- Wei, W., Pagnamenta, A. T., Gleadall, N., Sanchis-Juan, A., Stephens, J., Broxholme, J., . . . Chinnery, P. F. (2020). Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans. *Nature Communications*, *11*(1), 1740. doi:10.1038/s41467-020-15336-3
- Wei, W., Tuna, S., Keogh, M. J., Smith, K. R., Aitman, T. J., Beales, P. L., . . . Chinnery, P. F. (2019). Germline selection shapes human mitochondrial DNA diversity. *Science*, *364*(6442). doi:10.1126/science.aau6520
- Wei, X., Das, J., Fragoza, R., Liang, J., Bastos de Oliveira, F. M., Lee, H. R., . . . Yu, H. (2014). A massively parallel pipeline to clone DNA variants and examine molecular phenotypes of human disease mutations. *PLoS Genet*, *10*(12), e1004819. doi:10.1371/journal.pgen.1004819
- Weinberg, S. E., Singer, B. D., Steinert, E. M., Martinez, C. A., Mehta, M. M., Martinez-Reyes, I., . . . Chandel, N. S. (2019). Mitochondrial complex III is essential for suppressive function of regulatory T cells. *Nature*, *565*(7740), 495-499. doi:10.1038/s41586-018-0846-z
- Weischenfeldt, J., Symmons, O., Spitz, F., & Korbel, J. O. (2013). Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet*, *14*(2), 125-138. doi:10.1038/nrg3373
- Wexler, I. D., Hemalatha, S. G., McConnell, J., Buist, N. R., Dahl, H. H., Berry, S. A., . . . Kerr, D. S. (1997). Outcome of pyruvate dehydrogenase deficiency treated with ketogenic diets. Studies in patients with identical mutations. *Neurology*, *49*(6), 1655-1661. doi:10.1212/wnl.49.6.1655
- Wiedemann, N., & Pfanner, N. (2017). Mitochondrial Machineries for Protein Import and Assembly. *Annu Rev Biochem*, *86*, 685-714. doi:10.1146/annurev-biochem-060815-014352
- Wieland, T. (2015) Next-Generation Sequencing Data Analysis. *Diss. Technische Universität München*: <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20151111-1244253-1-6>.
- Wiemerslage, L., & Lee, D. (2016). Quantification of mitochondrial morphology in neurites of dopaminergic neurons using multiple parameters. *J Neurosci Methods*, *262*, 56-65. doi:10.1016/j.jneumeth.2016.01.008
- Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., . . . Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, *509*(7502), 582-587. doi:10.1038/nature13319
- Witters, P., Saada, A., Honzik, T., Tesarova, M., Kleinle, S., Horvath, R., Goldstein, A., & Morava, E. (2018) Revisiting mitochondrial diagnostic criteria in the new era of genomics. *Genet Med*, *20*(4), 444-451. doi: 10.1038/gim.2017.125.
- Wittig, I., Braun, H. P., & Schagger, H. (2006). Blue native PAGE. *Nature Protocols*, *1*(1), 418-428. doi:10.1038/nprot.2006.62
- Wittig, I., Carozzo, R., Santorelli, F. M., & Schagger, H. (2007). Functional assays in high-resolution clear native gels to quantify mitochondrial complexes in human biopsies and cell lines. *Electrophoresis*, *28*(21), 3811-3820. doi:10.1002/elps.200700367

- Woodworth, M. B., Girsakis, K. M., & Walsh, C. A. (2017). Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat Rev Genet*, *18*(4), 230-244. doi:10.1038/nrg.2016.159
- Wortmann, S. B., Kluijtmans, L. A., Rodenburg, R. J., Sass, J. O., Nouws, J., van Kaauwen, E. P., . . . Wevers, R. A. (2013). 3-Methylglutaconic aciduria--lessons from 50 genes and 977 patients. *J Inherit Metab Dis*, *36*(6), 913-921. doi:10.1007/s10545-012-9579-6
- Wortmann, S. B., Koolen, D. A., Smeitink, J. A., van den Heuvel, L., & Rodenburg, R. J. (2015). Whole exome sequencing of suspected mitochondrial patients in clinical practice. *J Inherit Metab Dis*, *38*(3), 437-443. doi:10.1007/s10545-015-9823-y
- Wortmann, S. B., Mayr, J. A., Nuoffer, J. M., Prokisch, H., & Sperl, W. (2017). A Guideline for the Diagnosis of Pediatric Mitochondrial Disease: The Value of Muscle and Skin Biopsies in the Genetics Era. *Neuropediatrics*, *48*(4), 309-314. doi:10.1055/s-0037-1603776
- Wortmann, S. B., Vaz, F. M., Gardeitchik, T., Vissers, L. E., Renkema, G. H., Schuurs-Hoeijmakers, J. H., . . . de Brouwer, A. P. (2012). Mutations in the phospholipid remodeling gene SERAC1 impair mitochondrial function and intracellular cholesterol trafficking and cause dystonia and deafness. *Nat Genet*, *44*(7), 797-802. doi:10.1038/ng.2325
- Wright, C. F., FitzPatrick, D. R., & Firth, H. V. (2018a). Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet*, *19*(5), 325. doi:10.1038/nrg.2018.12
- Wright, C. F., McRae, J. F., Clayton, S., Gallone, G., Aitken, S., FitzGerald, T. W., . . . Study, D. D. D. (2018b). Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genetics in Medicine*, *20*(10), 1216-1223. doi:10.1038/gim.2017.246
- Wu, M., Gu, J., Guo, R., Huang, Y., & Yang, M. (2016). Structure of Mammalian Respiratory Supercomplex I1III2IV1. *Cell*, *167*(6), 1598-1609 e1510. doi:10.1016/j.cell.2016.11.012
- Wutz, A. (2011). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet*, *12*(8), 542-553. doi:10.1038/nrg3035
- Xia, D., Esser, L., Tang, W. K., Zhou, F., Zhou, Y., Yu, L., & Yu, C. A. (2013). Structural analysis of cytochrome bc1 complexes: implications to the mechanism of function. *Biochim Biophys Acta*, *1827*(11-12), 1278-1294. doi:10.1016/j.bbabo.2012.11.008
- Xiao, M., Yang, H., Xu, W., Ma, S., Lin, H., Zhu, H., . . . Guan, K. L. (2012). Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev*, *26*(12), 1326-1338. doi:10.1101/gad.191056.112
- Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., . . . Frey, B. J. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, *347*(6218), 1254806. doi:10.1126/science.1254806
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., . . . Xiong, Y. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell*, *19*(1), 17-30. doi:10.1016/j.ccr.2010.12.014
- Xue, Y. L., Chen, Y., Ayub, Q., Huang, N., Ball, E. V., Mort, M., . . . Consortium, G. P. (2012). Deleterious- and Disease-Allele Prevalence in Healthy Individuals: Insights from Current Predictions, Mutation Databases, and Population-Scale Resequencing.

- American Journal of Human Genetics*, 91(6), 1022-1032. doi:10.1016/j.ajhg.2012.10.015
- Yan, H., Parsons, D. W., Jin, G., McLendon, R., Rasheed, B. A., Yuan, W., . . . Bigner, D. D. (2009). IDH1 and IDH2 mutations in gliomas. *N Engl J Med*, 360(8), 765-773. doi:10.1056/NEJMoa0808710
- Yang, L., Duff, M. O., Graveley, B. R., Carmichael, G. G., & Chen, L. L. (2011). Genomewide characterization of non-polyadenylated RNAs. *Genome Biol*, 12(2), R16. doi:10.1186/gb-2011-12-2-r16
- Yang, W., & Hekimi, S. (2010). A mitochondrial superoxide signal triggers increased longevity in *Caenorhabditis elegans*. *PLoS Biol*, 8(12), e1000556. doi:10.1371/journal.pbio.1000556
- Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., . . . Eng, C. M. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*, 312(18), 1870-1879. doi:10.1001/jama.2014.14601
- Yao, R., Zhang, C., Yu, T. T., Li, N., Hu, X. Y., Wang, X. M., . . . Shen, Y. P. (2017). Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. *Molecular Cytogenetics*, 10. doi:10.1186/s13039-017-0333-5
- Yaplito-Lee, J., Weintraub, R., Jamsen, K., Chow, C. W., Thorburn, D. R., & Boneh, A. (2007). Cardiac manifestations in oxidative phosphorylation disorders of childhood. *J Pediatr*, 150(4), 407-411. doi:10.1016/j.jpeds.2006.12.047
- Yatsuga, S., Fujita, Y., Ishii, A., Fukumoto, Y., Arahata, H., Kakuma, T., . . . Koga, Y. (2015). Growth differentiation factor 15 as a useful biomarker for mitochondrial disorders. *Ann Neurol*, 78(5), 814-823. doi:10.1002/ana.24506
- Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*, 11(2-3), 377-394. doi:10.1089/1066527041410418
- Yeo, J. H., Skinner, J. P., Bird, M. J., Formosa, L. E., Zhang, J. G., Kluck, R. M., . . . Chong, M. M. (2015). A Role for the Mitochondrial Protein Mrpl44 in Maintaining OXPHOS Capacity. *PLoS One*, 10(7), e0134326. doi:10.1371/journal.pone.0134326
- Yepez, V. A., Kremer, L. S., Iuso, A., Gusic, M., Kopajtich, R., Konarikova, E., . . . Gagneur, J. (2018). OCR-Stats: Robust estimation and statistical testing of mitochondrial respiration activities using Seahorse XF Analyzer. *PLoS One*, 13(7). doi:10.1371/journal.pone.0199938
- Yepez, V. A., Mertes, C., Mueller, M. F., Klaproth-Andrade, D. S., Wachutka, L., Frésard, L., . . . Gagneur, J. (2021). Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc*, 16(2), 1276-1296. doi: 10.1038/s41596-020-00462-5
- Yu-Wai-Man, P., Turnbull, D. M., & Chinnery, P. F. (2002). Leber hereditary optic neuropathy. *J Med Genet*, 39(3), 162-169. doi:10.1136/jmg.39.3.162
- Yu, J., Liang, X., Ji, Y., Ai, C., Liu, J., Zhu, L., . . . Guan, M. X. (2020). PRICKLE3 linked to ATPase biogenesis manifested Leber's hereditary optic neuropathy. *J Clin Invest*. doi:10.1172/JCI134965
- Yudkoff, M., Nelson, D., Daikhin, Y., & Erecinska, M. (1994). Tricarboxylic acid cycle in rat brain synaptosomes. Fluxes and interactions with aspartate aminotransferase and malate/aspartate shuttle. *J Biol Chem*, 269(44), 27414-27420.
- Zecha, J., Satpathy, S., Kanashova, T., Avanesian, S. C., Kane, M. H., Clauser, K. R., . . . Kuster, B. (2019). TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol Cell Proteomics*, 18(7), 1468-1478. doi:10.1074/mcp.TIR119.001385

- Zeng, C., Fukunaga, T., & Hamada, M. (2018). Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. *BMC Genomics*, *19*(1), 414. doi:10.1186/s12864-018-4765-z
- Zeng, Y., Wang, G., Yang, E., Ji, G., Brinkmeyer-Langford, C. L., & Cai, J. J. (2015). Aberrant gene expression in humans. *PLoS Genet*, *11*(1), e1004942. doi:10.1371/journal.pgen.1004942
- Zeviani, M., Servidei, S., Gellera, C., Bertini, E., DiMauro, S., & DiDonato, S. (1989). An autosomal dominant disorder with multiple deletions of mitochondrial DNA starting at the D-loop region. *Nature*, *339*(6222), 309-311. doi:10.1038/339309a0
- Zhang, A., Sun, H., Yan, G., Wang, P., & Wang, X. (2016). Mass spectrometry-based metabolomics: applications to biomarker and metabolic pathway research. *Biomed Chromatogr*, *30*(1), 7-12. doi:10.1002/bmc.3453
- Zhao, J., Akinsanmi, I., Arafat, D., Cradick, T. J., Lee, C. M., Banskota, S., . . . Gibson, G. (2016). A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood. *Am J Hum Genet*, *98*(2), 299-309. doi:10.1016/j.ajhg.2015.12.023
- Zhao, L., Liu, H., Yuan, X., Gao, K., & Duan, J. (2020). Comparative study of whole exome sequencing-based copy number variation detection tools. *BMC Bioinformatics*, *21*(1), 97. doi:10.1186/s12859-020-3421-1
- Zhao, S. R., Zhang, Y., Gamini, R., Zhang, B. H., & von Schack, D. (2018). Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA plus selection versus rRNA depletion. *Scientific Reports*, *8*. doi:10.1038/s41598-018-23226-4
- Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., . . . Wang, X. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature*, *555*(7697), 524-528. doi:10.1038/nature25980

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Holger Prokisch for the opportunity to conduct research in his group. His guidance, support, patience, and trust, as well as creative ideas, have enabled me to feel deeply involved and integrated into the mitochondrial disease research community, as well as become an independent researcher. My deep gratitude also goes to Prof. Thomas Meitinger, my Doktorvater, for the privilege to work at his institute, and to learn from the pioneers in the field. A special thanks to Prof. Julien Gagneur, my collaborator and external advisor, for valuable advice and suggestions, especially concerning the basic genetics questions.

I could not imagine my Ph.D. journey without my girls and Ph.D. comrades, Dr. Sarah Stenton, Enrica Zanuttigh, and Agnieszka Nadel. Thank you for your immeasurable support, understanding, and care, and teaching me lessons in both lab and life. We empower each other every day. I'm deeply indebted to you for your friendship. My gratitude extends to all of my group members, former and present (Dr. Laura Kremer, Dr. Eliska Konarikova, Robert Kopajtich, Dr. Arcangela Iuso, Dr. Silvia Vidali, Dmitrii Smirnov, Dewi Schlieben, Marieta Borzes, Caterina Terrile, Michael Färberböck) and other members of the Institute of Human Genetics and Neurogenomics (namely Dr. Riccardo Berutti, Dr. Thomas Schwarzmayer, Dr. Ana Antic) who made me feel quickly integrated, inside and outside the lab. Your support, advice, help, and comfort have been invaluable during this rollercoaster. We created an enjoyable working environment I can only hope to keep encountering throughout my career. I'd also like to recognize the administrative help of Martina Kuhnert.

A big thank you goes to my bioinformatics colleagues from TUM for our fruitful collaboration, shaping my ideas into code, making me acknowledge the importance of interdisciplinarity, and fun times outside work. Separate gratitude goes to my partner in crime, Dr. Vicente Yopez, my dear friend and constant collaborator throughout my Ph.D. I am also thankful to Dr. Basak Eraslan for her admiring work ethic, feminism, and our endless philosophical discussions. Gratitude also goes to the RNA-seq pioneers, Dr. Daniel Bader and Christian Mertes. Thank you for all our "work hard, play hard" moments.

My work would not have been possible without the support of the DZHG (German Centre for Cardiovascular Research) and the BMBF (German Federal Ministry for Education and Research), who funded me throughout my Ph.D. In addition, I thank my numerous collaborators all over the world for all the research opportunities, improving my projects, appreciating my opinion, and making my years here so rewarding. I also had the great pleasure being a member of DINI and Helmholtz Juniors.

My biggest thanks go to my sister and eternal cheerleader, Milica, for sharing the same passion for knowledge and never-ending curiosity for all aspects of life. Let's never stop learning. My gratitude also goes to my parents, my brother, and my extended family, for their unconditional love, support, and trust. Thank you for accepting my inquisitive mind and giving me the ultimate freedom of choices, as well as teaching me responsibility, ambition, and value of hard work. Finally, I am blessed and grateful to my friends from Serbia and all over the world for listening, encouraging, and supporting me along the way.