Márcia Joana Nascimento Teixeira

Investigation of Genes Associated to Mitochondrial Import and Post Translational Processing In LHON

UNIVERSIDADE D COIMBRA

UNIVERSIDADE D
COIMBRA

Márcia Joana Nascimento Teixeira

Investigation of Genes Associated to Mitochondrial Import and Post Translational Processing In LHON

Dissertação no âmbito do Mestrado em Bioquímica orientada pela Professora Doutora Maria Manuela Monteiro Grazina e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Agosto de 2018

FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Investigation of Genes Associated to Mitochondrial Import and Post Translational Processing In LHON

Márcia Joana Nascimento Teixeira

Dissertação apresentada à Faculdade de Ciências e Tecnologia da Universidade de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioquímica, realizada sob a orientação científica da Professora Doutora Manuela Grazina (Professora Auxiliar e Directora de Qualidade – Faculdade de Medicina da Universidade de Coimbra; Investigadora e Diretora do Laboratório de Biomedicina Mitocondrial e Teranóstica, Centro de Neurociências e Biologia Celular, Universidade de Coimbra) e sob orientação interna do Professor Doutor António Portugal (Faculdade de Ciências e Tecnologia da Universidade de Coimbra).

Agosto 2018

Financial support

This work was financed by FEDER funds through the Operational Competiveness Program – COMPETE 2020 and by national funds through FCT – Portuguese Foundation for Science and Technology through the strategic plan: COMPETE: POCI-01-0145-FEDER-007440.

ACKNOWLEDGMENTS

Quero agradecer a todos que, desde sempre, fizeram parte da minha vida e que contribuíram para me tornar na pessoa que sou hoje.

Em primeiro lugar, quero agradecer à Professora Doutora Manuela Grazina pela oportunidade de realizar este projeto, que me fez crescer quer a nível científico, quer a nível pessoal. Obrigada por todos os ensinamentos, por me mostrar que todas as dificuldades e situações inesperadas são oportunidades para nos pormos à prova, para melhorar e para ganhar novas competências. Obrigada pela preocupação e carinho sempre demonstrados.

Ao orientador interno, Professor Doutor António Portugal pela disponibilidade e apoio sempre demostrados.

À Doutora Salomé Pires e à Professora Doutora Filomena Botelho do Instituto de Biofísica da Faculdade de Medicina da Universidade de Coimbra pela cedência do lisado total de células WiDr.

À Mestre Catarina Neves pertencente ao grupo de Disfunção Retinal e Neuroinflamação, liderado pelo Doutor Francisco Ambrósio, pela cedência dos lisados totais de testículo e córtex de ratinho.

A todos os membros do Laboratório de Biomedicina Mitocondrial e Teranóstica, especialmente à Maria João Santos, Mafalda Bacalhau e Marta Simões, que foram essenciais para tornar este projeto possível. Obrigada por me terem integrado na equipa, por me terem acompanhado na parte experimental, por me terem ajudado a encontrar alternativas para melhorar este trabalho e por me terem dado força e alento para ultrapassar as complicações que surgiram. Quero também agradecer à Carolina Ribeiro por me ter ajudado no início deste trabalho e, mesmo já não estando no laboratório, por se ter mostrado disponível por continuar a ajudar no que fosse preciso. Como a vida não é só trabalho, quero agradecer-vos, incluindo à Fernanda Castela e à Célia Gomes pelos momentos de descontração e lazer que me ajudaram a aliviar o stress e a manter a energia positiva.

A todos os meus amigos que sempre me acompanharam em todos os momentos, que me ajudaram e me deram força. Obrigada por terem aturado o meu mau feitio que, por vezes, é necessário uma paciência de santo para se aturar. Obrigada pela vossa amizade!

A toda a minha família, especialmente aos meus pais, por todo o amor, paciência, carinho e por todos os sacrifícios que sempre fizeram para ter uma vida melhor e para realizar

todos os meus sonhos e desejos. À minha maninha, por me ter acompanhado desde pequenina, por toda a paciência, por toda força, por todos os conselhos, por todo o carinho e amor.

Obrigada por tudo!

INDEX

FIGURES AND TABLE INDEX..... **XV**

ABBREVIATIONS..... **XVIX**

ABSTRACT..... **XXV**

RESUMO..... **XXIX**

1. INTRODUCTION..... **1**

1.1. Mitochondria – structure and function..... 3

1.2. Nuclear-encoded proteins requiring mitochondrial import..... 3

1.3. Precursor targeting for the diverse mitochondrial subcompartments..... 4

1.4. Mitochondrial protein import machineries..... 4

1.4.1. TOM complex..... 7

1.4.2. TIM23 complex..... 10

1.4.2.1. Import motor associated to TIM23 complex..... 12

1.4.3. TIM22 complex..... 13

1.4.4. MIA complex..... 15

1.4.5. SAM complex..... 16

1.5. Post-translational modification of mitochondrial precursor proteins..... 18

1.6. Leber's Hereditary Optic Neuropathy (LHON)..... 21

1.6.1. Clinical presentation..... 21

1.6.2. Genetic context..... 21

1.6.2.1. mtDNA..... 21

1.6.2.2. nDNA..... 22

1.6.3. Biochemical context..... 23

1.6.4. "Incomplete penetrance" and LHON..... 23

2. AIMS OF THE STUDY..... 25

3. MATERIALS AND METHODS..... 29

3.1. Samples and controls..... 31

3.2. Whole exome sequencing (WES)..... 31

3.3. Search of nuclear gene variants..... 32

3.4. *In silico* analysis..... 32

3.5. Identification of the variants with functional impact using the Exomeloupe® software..... 34

3.6. Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP)…… 34

3.6.1. PCR…… 35

3.6.2. Agarose gel electrophoresis…… 36

3.6.3. RFLP…… 37

3.7. Sanger sequencing…… 38

3.8. Western Blot…… 39

4. RESULTS …… 43

4.1. Identification of genetic variants and *in silico* analysis for prediction of pathogenicity…… 45

4.2. Identification of variants with the functional impact using the ExomeLoupe®…… 46

4.3. Comparison of Tom20L sequences in the presence of the different variants…… 46

4.4. Validation of the presence of *MIPEP* and *TOMM20L* variants…… 50

4.5. Detection and analysis of the proteins affected by the variants…… 52

5. DISCUSSION …… 53

6. CONCLUSIONS AND FUTURE PERSPECTIVES…… 61

7. REFERENCES …… 65

FIGURES AND TABLES INDEX

FIGURES INDEX

Figure 1. Mitochondrial systems responsible for import of diverse precursor proteins into the diverse mitochondrial sub compartments……. 5

Figure 2. The majority of precursor proteins are initially imported through the TOM complex..…… 8

Figure 3. Cooperation between TOM and TIM23 complexes and import motor is necessary for translocation and insertion of precursors. …… 11

Figure 4. Import of hydrophobic metabolite carrier proteins of the inner membrane mediated by TOM and TIM22 complexes..…… 14

Figure 5. Import of precursors containing cysteine motifs is mediated by TOM and disulfide relay system…….…… 16

Figure 6. Insertion of β-barrel proteins into MOM is mediated by SAM complex…… 17

Figure 7. Genetic variants located in genes encoding proteins involved in the mitochondrial protein import and processing.…… 45

Figure 8. Possible effects of *TOMM20L* gene variants in the protein sequence…… 47

Figure 9. Bands' patterns from RFLP……. 50

Figure 10. Electropherograms comparing the fragments sequences in a control individual and the LHON patient…… 51

Figure 11. Assessment of the presence and/or quantity of MIP, COX4 and Tom20L proteins from human and mouse cells…….. 52

TABLES INDEX

Table 1. Components of mitochondrial protein import systems……. 6

Table 2. Human genes encoding the various subunits and proteins of mitochondrial proteins import systems and post-translation processing of precursor proteins…… 32

Table 3. Primer sequences (designed using the Primer3 software) for amplification of fragments from *MIPEP* and *TOMM20L*…… 36

Table 4. Cycling conditions for amplification of *MIPEP* and *TOMM20L* fragments……. 36

Table 5. Restriction enzymes and expected band patterns.…… 38

Table 6. Cycling conditions used in sequencing PCR…… 39

Table 7. Variants identified with a predictive functional impact in genes encoding proteins belonging to mitochondrial import systems and processing of precursors……. 48

Table 8. Variants present in *MIPEP* and *TOMM20L* identified in sample 22 using the ExomeLoupe® software…….. 49

ABBREVIATIONS

Δψ	Electrochemical potential of the inner mitochondrial membrane
α-MPP	Mitochondrial-processing peptidase subunit-α
β-MPP	Mitochondrial-processing peptidase subunit-β
ALR	Augmented Liver Regeneration/ Flavin Adenine Dinucleotide (FAD)-linked sulfhydryl oxidase
AIF	Apoptosis-inducing factor
AIP	Arylhydrocarbon receptor interacting protein
AP-1	Activated by protein-1
ATP	Adenosine Triphosphate
BioMiT Lab	Laboratory of Mitochondrial Biomedicine and Theranostics
BSA	bovine serum albumin
COX4	Cytochrome c oxidase subunit 4
Cyt.c	Cytochrome c
CHCH4	Mitochondrial intermembrane space import and assembly protein 40
DMSO	Dimethylsulfoxide
dNTPs	Deoxyribonucleotide triphosphates
ddNTPs	Dideoxyribonucleotide triphosphates
e^-	electron
ERV1	Mitochondrial Flavin Adenine Dinucleotide (FAD)-linked sulfhydryl oxidase
ExAc	Exome Aggregation Consortium
GABP	GA-binding protein
GWAS	Genome-wide association study
Hsc70	Heat shock cognate 71 kDa protein
Hsp90	Heat shock protein 90-alpha
CIV	Cytochrome c oxidase
IMS	Intermembrane space
IMP	Inner membrane peptidase
IMMP1L/Imp1	Mitochondrial inner membrane protease subunit 1
IMMP2L/Imp2	Mitochondrial inner membrane protease subunit 2
LBG	Laboratory of Biochemical Genetics

LHON	Leber's hereditary optic neuropathy
MAF	Minor allele frequence
MIM	Mitochondrial inner membrane
MIA	Mitochondrial intermembrane space import and assembly machinery
Mia40	Mitochondrial intermembrane space import and assembly protein 40
MIB	Mitochondrial intermembrane space bridging complex
MITRAC	Mitochondrial translation regulation assembly intermediate of cytochrome c oxidase
MIP	Mitochondrial Intermediate Peptidase
MOM	Mitochondrial outer membrane
MPP	Mitochondrial Processing Peptidase
MRC	Mitochondrial respiratory chain
mtDNA	Mitochondrial DNA
MT-ND1	Mitochondrial NADH dehydrogenase subunit 1
MT-ND4	Mitochondrial NADH dehydrogenase subunit 4
MT-ND6	Mitochondrial NADH dehydrogenase subunit 6
Mtx 1/3	Metaxin 1/3
Mtx 2	Metaxin 2
nDNA	Nuclear DNA
NADH	Nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance
NRF1	Nuclear respiratory factor 1
NF-kB	Nuclear factor kappa-B
OXPHOS	Oxidative phosphorylation
PAM	Presequence translocase-associated motor
PARL	Presenilins-associated rhomboid-like
PCR-RFLP	Polymerase Chain Reaction – Restriction Fragment Length Polymorphism
PDB	Protein Data Bank
PMPCA	Mitochondrial processing peptidase subunit alpha

PMPCB	Mitochondrial processing peptidase subunit beta
PROVEAN®	Protein Variation Effect Analyzer®
Q	Glutamine
R	Arginine
RGCs	Retinal ganglion cells
rs	reference cluster ID
SAM	Sorting and assembly complex
Sam50	Sorting and assembly subunit 50
Sdh3	Succinate dehydrogenase cytochrome b subunit 3
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIFT®	Sorting Intolerant From Tolerant®
T	Threonine
TCA	Tricarboxylic acid cycle
TIM23	Inner membrane translocase
TIM22	Carrier translocase
Tim17A	Presequence translocase subunit 17a
Tim17B1	Presequence translocase subunit 17b1
Tim17B2	Presequence translocase subunit 17b2
Tim18	Carrier translocase subunit 18
Tim19	Carrier translocase subunit 19
Tim23	Presequence translocase subunit 23
Tim23B1	Presequence translocase subunit 23b1
Tim23B2	Presequence translocase subunit 23b2
Tim44	Presequence translocase subunit 44
Tim50	Presequence translocase subunit 50
Tim54	Carrier translocase subunit 54
TOM	Outer membrane translocase
Tom5	Outer membrane translocase subunit 5
Tom6	Outer membrane translocase subunit 6

Tom7	Outer membrane translocase subunit 7
Tom20	Outer membrane translocase subunit 20
Tom22	Outer membrane translocase subunit 22
Tom40	Outer membrane translocase subunit 40
Tom70	Outer membrane translocase subunit 70
UTR	Untranslated region
WES	Whole-Exome Sequencing
WiDr	Colon adenocarcinoma cell line
XPNPEP3	X-Pro aminopeptidase
YARS2	Mitochondrial tyrosyl-tRNA synthetase

ABSTRACT

The Leber's Hereditary Optic Neuropathy (LHON) is a rare mitochondrial disorder characterized by the loss of retinal ganglion cells (RGCs), which are responsible for the conduction of the visual information to the cerebral cortex. This loss leads to the degeneration of the optic nerve resulting in a sudden loss of vision.

The major confirmed genetic causes of this disorder are mitochondrial DNA (mtDNA) mutations in genes encoding subunits belonging to the complex I of the mitochondrial respiratory chain (MRC), namely *MT-ND1* (m.3460G>A), *MT-ND4* (m.11778G>A) and *MT-ND6* (m.14484T>C). The presence of these mutations do not totally explain the disease phenotype, since the existence of individuals carrying homoplasmic mtDNA mutations without disease manifestation have been reported.

There are several factors including nuclear genetic modifiers that have been suggested to be influencers of the phenotype manifestation.

The presence of nuclear gene variants in subunits and proteins involved in the mitochondrial protein import and processing of imported precursor proteins may contribute as genetic modifiers in LHON. To assess this possibility, a search for the mentioned genetic variants in whole-exome sequencing data was performed and the prediction of functional impact of the relevant variants was evaluated using bioinformatics' tools. This analysis resulted in the identification of the promising c.280C>T and c.170delA/c.172_176delGGCAC variants from *MIPEP* and *TOMM20L* gene, respectively, in a LHON individual with the m.14484T>C mutation. These variants were confirmed using two additional methods, namely Sanger sequencing and PCR-RFLP.

The implications at the protein level have been investigated in a preliminary study. Limitations at the level of tissue specificity of Tom20L protein expression did not allow the evaluation of the impact of the genetic variants identified in the protein expression. Further experimental validation of the genetic variants is required for the clarification of its pathogenicity, but the results suggest that the alteration c.280C>T of *MIPEP* gene may have an additional effect on the pathogenicity of the variant m.14484T>C.

Keywords: LHON; nuclear gene variants; mitochondrial protein import; precursor processing; *MIPEP*; *TOMM20L*

RESUMO

A Neuropatia Ótica Hereditária de Leber (LHON) é uma doença mitocondrial rara caracterizada pela perda de células ganglionares da retina, responsáveis pelo envio da informação visual para o córtex cerebral. A perda destas células resulta na degenerescência do nervo ótico e, consequentemente, na perda súbita da visão.

As causas genéticas que comprovam a presença desta doença são mutações em genes mitocondriais que codificam subunidades do complexo I da cadeia respiratória mitocondrial, nomeadamente *MT-ND1* (m.3460G>A), *MT-ND4* (m.11778G>A) e *MT-ND6* (m.14484T>C). A presença destas mutações não explica totalmente o fenótipo da doença, visto que a existência de indivíduos portadores de mutações do mtDNA homoplásmicas, sem manifestação da doença tem sido reportada.

Existem vários fatores incluindo fatores genéticos nucleares, que têm sido sugeridos como moduladores da manifestação da doença.

Um destes fatores genéticos poderá ser a presença de variantes em genes nucleares, que codificam subunidades e proteínas envolvidas na importação de proteínas para a mitocôndria e no processamento pós-tradução e importação. No sentido de avaliar esta possibilidade, foi realizada a pesquisa destas variantes em dados de sequenciação de exoma total e foi avaliada a previsão do impacto funcional das variantes relevantes. A análise resultou na identificação de 3 variantes promissoras: c.170delA e c.172_176delGGCAC no gene *TOMM20L* e c.280C>T no gene *MIPEP*, num doente afetado com LHON portador da mutação patogénica m.14484T>C. As variantes identificadas foram confirmadas por dois métodos adicionais, sequenciação de Sanger e PCR-RFLP. As implicações das variantes genéticas ao nível da proteína foram investigadas num estudo preliminar. Limitações ao nível da especificidade da expressão tecidular da proteína Tom20L não permitiram avaliar o impacto das variantes genéticas identificadas ao nível da expressão da proteína. A validação funcional e experimental adicional das variantes genéticas identificadas será necessária para clarificar a sua patogenicidade, mas os resultados sugerem que a alteração c.280C>T do gene *MIPEP* poderá ter um efeito potenciador da patogenicidade da variante m.14484T>C.

Palavras-chave: LHON; variantes genéticas nucleares; importação de proteínas mitocondriais; processamento de precursores; *MIPEP*; *TOMM20L*

1. INTRODUCTION

1.1. Mitochondria – structure and function

The mitochondria are essential organelles of eukaryotic cells involved in many processes required for the normal function of cells. The mitochondrial content varies among cell types, according to the energy demands. Mitochondria have double membrane systems composed by the mitochondrial outer membrane (MOM), which delimits the mitochondrial intermembrane space (IMS) and the mitochondrial inner membrane (MIM), which delimits the mitochondrial matrix. The MIM has invaginations called cristae where the several sets of five complexes of the mitochondrial respiratory chain (MRC) are located and catalyze the oxidative phosphorylation (OXPHOS) for ATP production[1,2]. The IMS is a hydrophilic compartment where molecules required for the mitochondrial metabolism are localized, including regulators of programmed cell death (apoptosis) and components for energy conversion[1,2]. Essential molecules are located in the matrix, specifically enzymes required for biosynthesis of metabolites, including tricarboxylic acid cycle (TCA), and ribosomes, required for translation of mitochondrial DNA (mtDNA), the mtDNA and matrix granules[1,2].

In addition to the energy production, this dynamic organelle has a strong involvement in apoptosis, metabolism of amino acids, lipids, ketone bodies, biosynthesis of heme and iron-sulfur clusters and it is also involved in the responses to cellular stress[3].

1.2. Nuclear-encoded proteins requiring mitochondrial import

Mitochondria contains more than 1,000 proteins that are encoded by two genomes, mtDNA and nuclear DNA (nDNA). Despite the fact that mitochondria contain their own DNA, it only codes for 13 peptides (in humans) that are part of the MRC complexes representing approximately 1 % of all proteins. The other proteins encoded by nDNA are synthesized in cytosolic ribosomes, in their precursor form. Therefore, the existence of protein import systems that recognize mitochondrial precursor proteins are mandatory for ensuring their efficient transport to the diverse compartments of mitochondria and consequently guaranteeing their functions. The proteins are mainly imported after translation[3–5]. However, several reports pointed to the presence of an additional co-translational protein import mechanism into mitochondria[6–10].

1.3. Precursor targeting for the diverse mitochondrial subcompartments

The precursor proteins have targeting signals that allows the import to the diverse subcompartments[3]. Approximately two-thirds of nuclear-encoded mitochondrial proteins have N-terminal targeting signals allowing their targeting into mitochondrial matrix, namely presequences, which are cleaved off after translocation into the matrix[3,11]. The presequences vary in length with approximately 15-50 residues long and form amphipathic α-helices with a positive net charge[3,11]. Some precursor proteins contain an additional hydrophobic sorting signal that is located after the matrix-targeting signal[3,11]. This signal arrests translocation through the MIM and the precursors are released laterally into the membrane[3,11]. Other precursor proteins have internal targeting signals within the mature protein sequence[3,12]. The internal signals direct the precursor proteins to all mitochondrial subcompartments[12]. The precursor proteins containing a C-terminal β-barrel signal consisting of a large polar amino acid (lysine or glutamine), an invariant glycine and two hydrophobic amino acids target these proteins to the MOM[5]. In addition, multi-spanning proteins are also targeted to this membrane. The proteins containing the characteristic Cx_3C and Cx_9C motifs are targeted to the IMS[12,13]. Hydrophobic metabolite carrier proteins of the MIM containing six or four transmembrane helices connected by short, matrix exposed loops have internal signals targeting them to the MIM[14–16].

1.4 Mitochondrial protein import machineries

Many types of machinery are involved in the transportation of proteins to the diverse mitochondrial subcompartments: outer membrane translocase (TOM complex), presequence translocase of the MIM (TIM23 complex), carrier translocase (TIM22 complex), mitochondrial intermembrane space import and assembly machinery (MIA complex), sorting and assembly (SAM) complex. In MIM, another machinery system is present, denominated mitochondrial contact site and cristae organizing system (MICOS), that is not involved in protein transportation but interacts with some subunits of SAM complex, leading to the formation of mitochondrial intermembrane space bridging complex (MIB)[17–19]. Studies focused in import using a diversity of precursor proteins and analysis of molecular interaction between the various translocases revealed an wide cooperation of protein machineries involved in mitochondrial protein sorting[13].

These mitochondrial import machineries (Figure 1) comprise many components (Table 1) and have been studied at more extent in fungal systems, like *Saccharomyces cerevisiae* and

Neurospora crassa[20]. The basic processes needed to the import of proteins into mitochondria seem to be conserved from fungi to mammals but differences in structure and complexity are rather evident in higher eukaryotes[20].

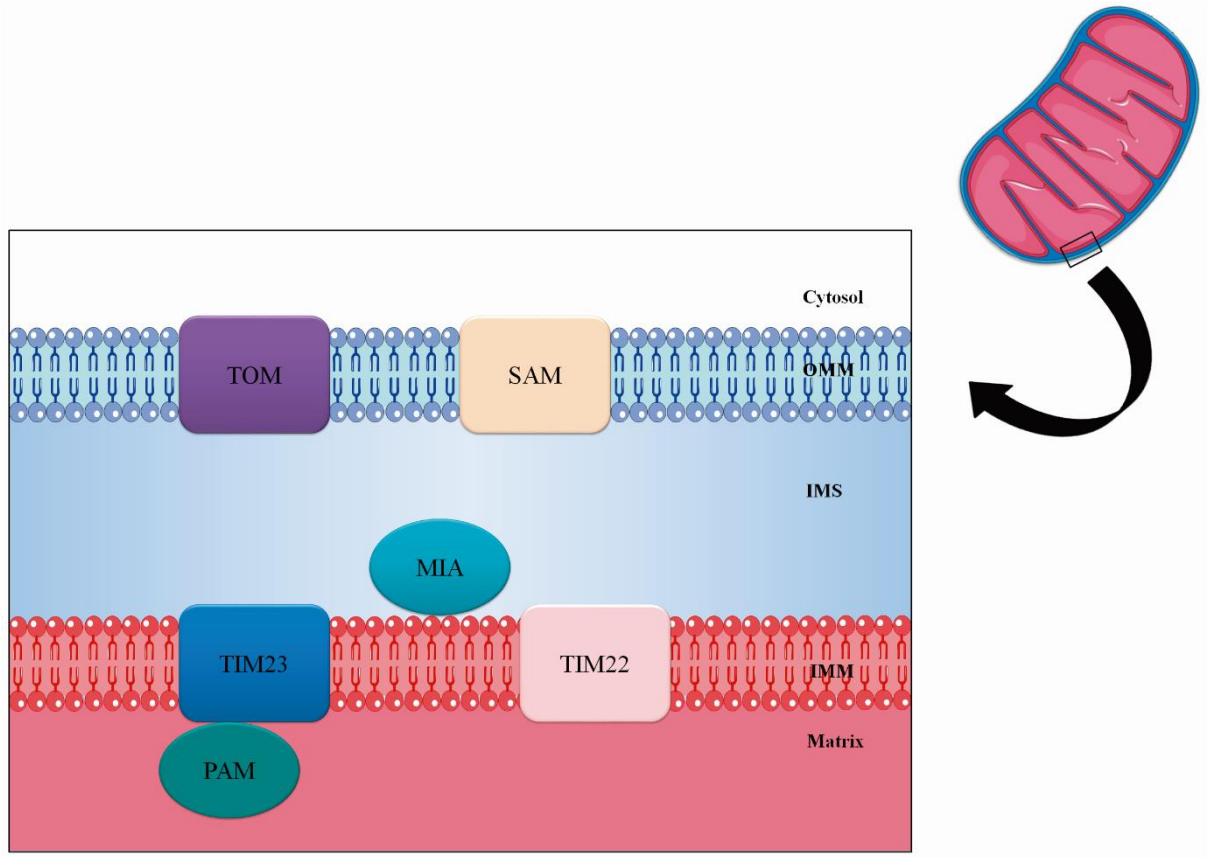

Figure 1. Human mitochondrial systems responsible for import of diverse precursor proteins into the diverse mitochondrial sub compartments. MOM, mitochondrial outer membrane; IMS, intermembrane space; MIM, mitochondrial inner membrane; TOM, outer membrane translocase; TIM23 complex, inner membrane presequence translocase; PAM, presequence translocase–associated motor; TIM22 complex, carrier translocase; SAM complex, sorting and assembly complex; MIA complex, mitochondrial intermembrane space import and assembly machinery.

Table 1: Human and yeast components of mitochondrial protein import systems

Gene symbol Human	Proteins/subunits human	Proteins/subunits yeast	Function	Local.	Ref.
TOM complex					
TOMM20	Tom20		Receptor for precursors with a presequence	MOM	21
TOMM20L	Tom20L				22
TOMM22	Tom22		Receptor, organizer of TOM	MOM	23
TOMM70	Tom70		Receptor for hydrophobic precursors	MOM	24
TOMM40	Tom40		Forms a translocation channel	MOM	25
TOMM40L					
TOMM5	Tom5		Transfer of precursors from receptors to channel, assembly of TOM	MOM	26
TOMM6	Tom6		Assembly (in yeast) and disassembly (in humans) of TOM	MOM	26
TOMM7	Tom7		Disassembly (in yeast), assembly (in humans) and dynamics of TOM	MOM	20
TOMM34	Tom34	-	Transport precursors across cytosol for delivery to TOM complex	Cytosol	27
HSPA8	Hsc70	Hsp70			
HSP90AA1	Hsp90				
TIM23 complex					
TIMM23	Tim23		Forms a translocation channel	MIM	28
TIMM23B					
TIMM17A	Tim17a	Tim17	Involved in lateral sorting of precursors	MIM	28
TIMM17B	Tim17b1 Tim17b2				
TIMM50	Tim50		Receptor, gating of Tim23 channel	MIM	29
TIMM21	Tim21		interacts with the IMS domain of Tom22	MIM	30
PAM					
HSPA9	Hsp70 member 9/ Mortalin	mtHsp70	ATP-dependent protein transport	MIM	31
TIMM44	Tim44		Membrane anchor for PAM subunits	MIM	32
PAM16	Magmas	Tim16/ Pam16	controls Pam18/DnaJC19/15 activity	MIM	33
DNAJC19	DnaJ Hsp40 C19	Tim14/ Pam18	stimulates ATPase activity of mtHsp70/mortalin	MIM	33
DNAJC15	DnaJ Hsp40 C15				
PAM17	-	Pam17	Integrity of Pam18-Pam16 module, binds to Tim23	MIM	34
GRPEL1	GrpE like 1	Mge1	Mitochondrial nucleotide-exchange factor for mtHsp70/Mortalin	MIM	35
GRPEL2	GrpE like 2				
TIM22 complex					
TIMM22	Tim22		Forms a double channel	MIM	36
TIMM29	-	Tim29	assembly and stability of TIM22 complex	MIM	37
-	-	Tim54	Binds Tim9-Tim10-Tim12 complex	MIM	38
-	-	Tim18	Involved in assembly of TIM22 complex	MIM	38
TIMM9	Tim9	Tim9	Transport of precursors through IMS	IMS	39
TIMM10	Tim10a	Tim10		IMS	
TIMM10B	Tim10b	Tim12		IMS	
MIA complex					
CHCHD4	Mia40		Mediates oxidation of precursors through disulfide bonds	IMS	40
GFER	ALR	Erv1	Oxidation of and cooperation with Mia40 in transfer of disulfide bonds	IMS	41,42
-	-	Hot13	Promotes oxidation of Mia40 by Erv1, binding of zinc ions	IMS	43
AIF1	AIF	-	Involved in the import of human Mia40	Nucleus Cytosol IMS	27

Table 1 (continuation)					
SAM complex					
SAMM50	Sam50		Forms a channel for insertion	MOM	[44]
MTX2	Metaxin 2	Sam35	Binding of β signal, partner of Sam50	MOM	[45]
MTX1	Metaxin 1	Sam37	Promotes release of precursors	MOM	[45]
TIMM9	Tim9	Tim9	Acceptance and transport of precursors through IMS	IMS	[39]
TIMM10	Tim10a	Tim10			
TIMM8A	Tim8a	Tim8			
TIMM8B	Tim8b				
TIMM13	Tim13	Tim13			

MOM, mitochondrial outer membrane; IMS, intermembrane space; MIM, mitochondrial inner membrane; Ref, reference; Local, Localization. Gene symbols searched in NCBI and MitoProteome databases.

Almost all precursor proteins are initially imported into the mitochondria by TOM complex; then, it cooperates with other protein complexes for sorting into the diverse mitochondrial subcompartments (Figure 1)[13,46]. Precursor proteins containing β-barrel signals and multi-membrane-spanning proteins directed to the MOM are passed on to SAM complex and IMS proteins require the MIA complex[12,13]. Insertion of precursor proteins into the MIM is mediated by TIM22 or TIM23, whereas the translocation of precursor proteins into the matrix is only mediated by TIM23 [3,12,46]. The precursor proteins are handed to TIM23 complex by TOM complex and then they can be imported into the matrix or inserted in the MIM lipid bilayer[12,47]. Hydrophobic metabolite carrier proteins of the MIM are imported by TIM22 complex[14–16].

1.4.1 TOM complex

As in yeast, the human TOM complex (Figure 2) is the main entry gate to mitochondria and it consists of three membrane-integral receptor subunits Tom20, Tom22 and Tom70, the channel-forming subunit Tom40 and three small Tom proteins Tom5, Tom6 and Tom7[12,13,21,25,27,48]. All homologous components were identified in humans[20,21,23,26,49] demonstrating that this complex is highly conserved and consequently highlighting its importance[26]. The Tom20 acts as the initial recognition site for precursors with presequences interacting with their hydrophobic surface and transfers the precursors to the central receptor Tom22[12,24,50–52]. Additionally, it may interact with chaperones that bind to precursors, preventing their aggregation and targeting them to TOM[12]. In human cells, the precursors with presequences delivered to Tom20 are transported by Hsc70 and another cytosolic factor, arylhydrocarbon receptor interacting protein (AIP)[27,53]. This delivery by cytosolic chaperones to Tom20 does not occur in yeast[48].

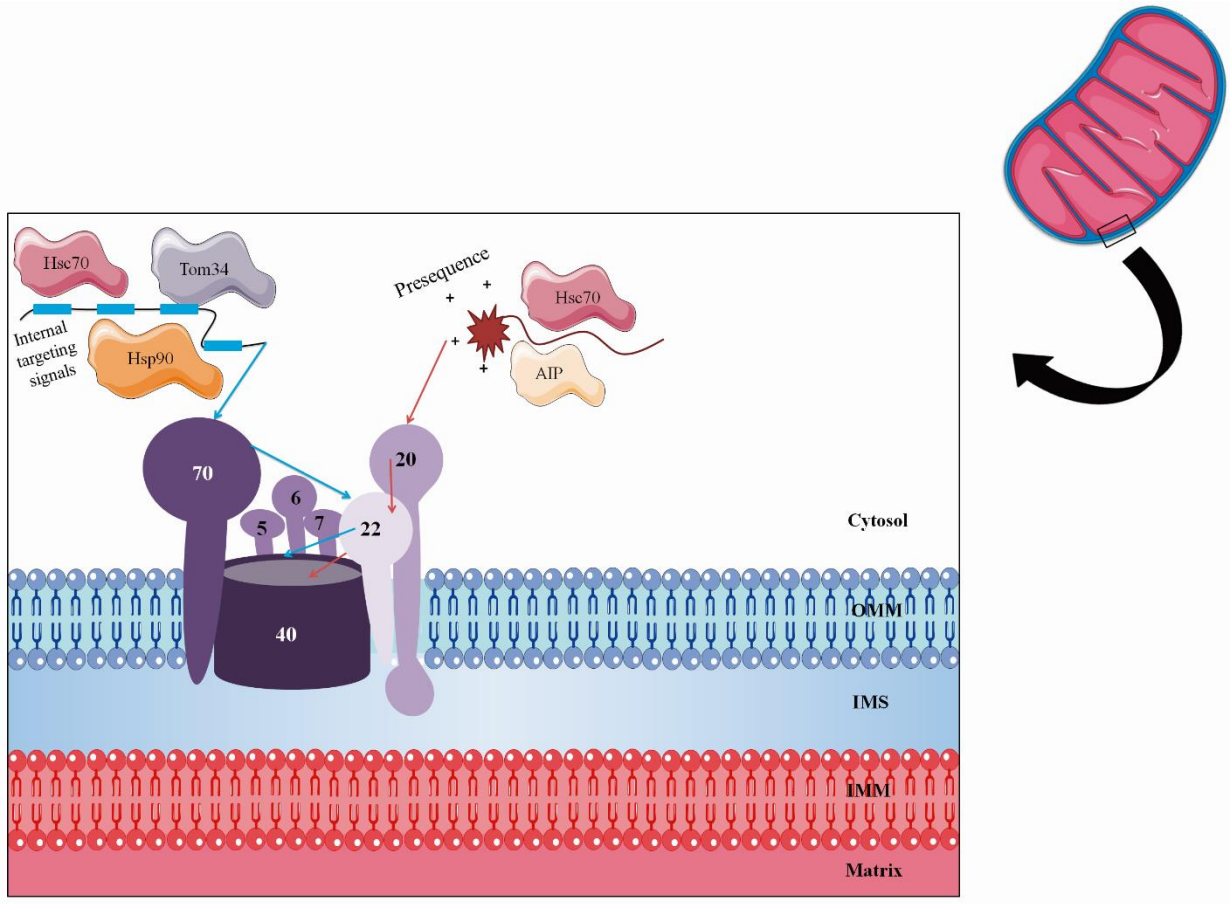

Figure 2. The majority of precursor proteins are initially imported through the TOM complex. MOM, mitochondrial outer membrane; IMS, intermembrane space; MIM, mitochondrial inner membrane; TOM, outer membrane translocase; Tom20, outer membrane translocase subunit 20; Tom22, outer membrane translocase subunit 22; Tom70, outer membrane translocase subunit 70; Tom40, outer membrane translocase subunit 40; Tom5, outer membrane translocase subunit 5; Tom6, outer membrane translocase subunit 6; Tom7, outer membrane translocase subunit 7; Hsc70, Heat shock cognate 71 kDa protein; AIP, arylhydrocarbon receptor interacting protein; Tom34, Mitochondrial import receptor subunit 34;Hsp90, Heat shock protein 90-alpha

The existence and characterization of a novel isoform of Tom20 in animals, including humans, rats and mice, was reported by Likic and colleagues (2005), that was designated type I Tom20 (also called Tom20-like/Tom20L) since the isoform structurally and functionally characterized at the time was designated as type II Tom20[22]. In mammalian cells, Tom20L is located in the mitochondria[22,54]. The presence of this isoform is also reported in other organisms, including some fish and invertebrates (*Drosophila melanogaster* and *Caenorhabditis elegans)*[22,54]. The study of the Tom20 isoforms in mice and rats reveled clear differences between Tom20 and Tom20L[22]. In mice, the isoforms presented a sequence identity of 34.9% with various conservative substitutions while in rats, they presented 60.3% of sequence identity and similarity[22]. The analysis of the 3D structure was also made using the rat Tom20 NMR structure. Furthermore, the interaction of mouse Tom20L with a presequence peptide was also achieved using the rat Tom20-presequence peptide complex as template [22]. The comparison of

these interactions was made for trying to understand if Tom20L was also capable of recognition of precursors and if the interaction was similar[22]. This analysis revealed obvious differences between the isoforms[22]. From the predicted presequence-interacting residues, the majority of different residues maintained their proprieties allowing the assumption that the binding between receptor-substrate is not disturbed[22]. Some residues changes led to a charge alteration, which could cause to a possible binding[22]. Even though there is this possibility, the use of a representative peptide of a precursor presequence does not accurately represent the real interaction, because in a native substrate, the region with charge modification would be located distantly from the N-terminal of the precursor[22].

Furthermore, the functional characterization of Tom20L was also performed. Analysis of the Tom20 isoforms expression in mouse revealed a differential expression of Tom20L with testis specificity unlike Tom20 that was ubiquitously expressed. A similar expression pattern was reveled in *D. melanogaster*[22]. The specificity of Tom20L was demonstrated before in this specimen being only present in male germline specifically in early spermiogenesis[54]. Other studies focused on the clarification of expression patterns of the totality of human genes demonstrated a clear testis preference for the expression of *TOMM20L*[55,56]. The second tissue with a higher expression of *TOMM20L* was the cerebral cortex and it was expressed in other tissues, but in a reduced extent[55,56]. Function loss experiments performed in *C. elegans* demonstrated that Tom20L cannot replace Tom20, but it is functional[22].

The Tom22 connects Tom20 to Tom40 and may assist Tom20 in the binding and unfolding of precursor proteins[48]. In addition, as in yeast, human Tom22 is also a crucial player in the integrity of the TOM complex[3,12,27]. The Tom70 recognizes hydrophobic precursors that contain internal targeting signals[12]. In humans, this subunit also interacts with three cytosolic chaperones (Hsc70,Hsp90 and Tom34) while in yeast it only interacts with Hsp70[12,13,27]. The Tom34 is absent in yeast, being exclusively present in mammalian cells, and its function remains indefinable[27]. While a report has proposed that this protein interacts with the mature portion of some precursor proteins, maintaining them in an unfolded state for import[57], another report suggested that the mitochondrial import was not disturbed by Tom34 depletion in mice[58], and consequently, further studies are needed for the assessment of the role played by Tom34[58]. After recognition, Tom70 transfers the precursor to Tom22[3,12]. Afterwards, the precursors are inserted into the Tom40 channel. The Tom40, a β-barrel protein, allows the passage of proteins across MOM[59] and a recently described conserved polar groove, uniquely present in this β-barrel protein, was suggested to interact with the presequence of precursors[11,60]. The small Tom proteins are involved in the assembly, dynamics and stability of TOM complex[3]. The Tom5 promotes the assembly of the TOM complex and participates in precursor protein transfer from

Tom22 to the Tom40 channel[20,26]. The Tom6 and Tom7 play antagonistic roles[20,26]. Moreover, the function of the individual proteins is not conserved between yeast and human mitochondria[20,27]. While yeast Tom6 stabilizes the large TOM complex and yeast Tom7 favors its disassembly, these roles seem to be inverted in humans[3,20,26]. The small Tom proteins seem to be essential to TOM as a group, since the loss of individual small TOM proteins causes merely reduced effects and their simultaneous depletion is lethal[12].

The expression of the human genes encoding the receptor subunits of TOM complex, namely Tom20 and Tom70, seem to be positively regulated via GA-binding protein (GABP) which is a transcription factor involved in the regulation of genes encoding a variety of molecules which have roles in multiple functions of mitochondria[61,62].

In yeast, the TOM complex is regulated by cytosolic kinases which phosphorylate the complex in response to different growth conditions and to the cell cycle[13]. This regulation influence the assembly of the TOM complex and therefore the availability of functional complexes[13,63].

1.4.2 TIM23 complex

The TIM23 complex (Figure 3) is composed by the essential subunits, Tim23, Tim17 and Tim50, and a non-essential subunit, Tim21[12,28]. All homologous components of the yeast complex have been identified in humans and presented high similarity demonstrating a structural and functional conservation[29,30,48,64]. The Tim23 and Tim17 form a channel for translocation of precursors, which is dependent on the membrane potential; in addition, Tim17 also regulates the channel and it is responsible for the differential sorting[5,12,39,65,66]. Moreover, the Tim23 cooperates with Tim50 in the recognition of the presequences[12,13,67]. In humans, Tim23 is represented by three variants: Tim23, Tim23b1 and Tim23b2, being the latter two proteins encoded by the same gene, *TIMM23B*[68]. The expression of *TIMM23* and *TIMM23B* is regulated by two different transcription factors, namely GABP and nuclear respiratory factor 1 (NRF1) having activating and repressing activity, respectively[68]. The human Tim17 is also represented by three variants: Tim17B1, Tim17B2 and Tim17A[30,69]. These variants are integrated in three distinct forms of human TIM23 complex, human translocase B1, human translocase B2, and human translocase A, respectively[28,69]. Even though Tim17 and Tim23 present three variants, the integration with different complexes is only reported in Tim17[68]. The Tim50 cooperates with Tim23 for recognition of precursors[48]. In contrast with yeast Tim50, the human Tim50 presents phosphatase activity and its role is still elusive[27,69]. The TIM23 complex interacts transiently with TOM complex through the subunits Tim50, Tim23 and Tim21, thus

assisting in the translocation of the precursor from the MOM to the MIM[13]. Apart from being a receptor of presequences in the IMS[3], the regulation of permeability of the translocation channel is also a task for Tim50[12]. The Tim21 also forms part of the mitochondrial translation regulation assembly intermediate of cytochrome c oxidase (MITRAC) being crucial for the assembly of subunits into the mature complex IV[13,64]. Additionally, it is needed for the transfer of precursor respiratory chain subunits containing presequences from the TIM23 complex[27,48,64]. The Tim21 might also regulate the association of the import motor with the subunits from the membrane belonging to the TIM23 complex[12].

Translocation mediated by the TIM23 complex is energetically driven by the membrane potential (Δψ) across the MIM and the hydrolysis of ATP[3,12,25,70–72]. The membrane region of the TIM23 complex can transfer only the targeting signal of precursor proteins and insert laterally in the lipid phase of MIM[3,12,25,72]. Then, the presequence translocase-associated motor (PAM) has to take over[12]. The PAM is responsible for completing the protein translocation across the MIM in a process that depends of energy obtained from ATP hydrolysis[12].

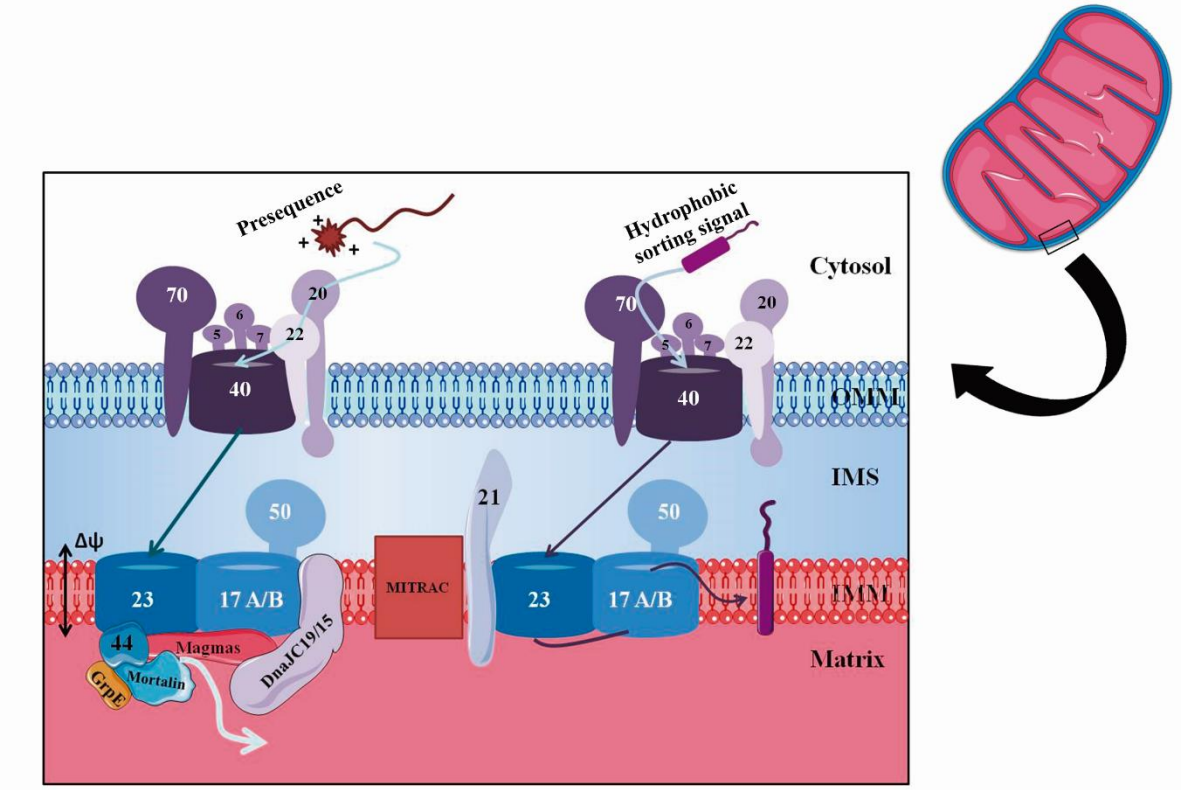

Figure 3. Cooperation between TOM and TIM23 complexes and import motor is necessary for translocation and insertion of precursors. MOM, mitochondrial outer membrane; IMS, intermembrane space; MIM, mitochondrial inner membrane; Δψ, electrochemical potential of the inner mitochondrial membrane; TOM, outer membrane translocase; Tom20, outer membrane translocase subunit 20; Tom22, outer membrane translocase subunit 22; Tom70, outer membrane translocase subunit 70; Tom40, outer membrane translocase subunit 40; Tom5, outer membrane translocase subunit 5; Tom6, outer membrane translocase subunit 6; Tom7, outer membrane translocase subunit 7; TIM23, presequence translocase; Tim17A/B, presequence translocase subunit 17A/B; Tim23, presequence translocase subunit 23; Tim50, presequence translocase subunit 50; Tim44, presequence translocase subunit 44; Tim21, presequence translocase subunit 21; PAM, presequence translocase-associated motor; GrpE, GrpE protein homolog 1, mitochondrial; DnaJC19/15, DnaJ homolog subfamily C member 19/15; MITRAC, mitochondrial translation regulation assembly intermediate of cytochrome c oxidase;Tim21, Mitochondrial import inner membrane translocase subunit 21

1.4.2.1 Import motor associated to TIM23 complex

In yeast, this motor consists of molecular chaperone heat shock protein 70 (mtHsp70), which represents the central subunit, Tim44, two subunits with DnaJ-like structures Tim16/Pam16 and Tim14/Pam18, Pam17, and nucleotide exchange factor Mge1[12]. In humans, all homologous components (Figure 3) were identified, except Pam17 with high functional evolutionary conservation[30,31,33,34,73]. Mortalin, the human equivalent of yeast mtHsp70 is responsible for the ATP-dependent transport of the mature part of the precursor proteins into the mitochondrial matrix[12]. The activity of Mortalin during precursor transport is controlled by co-chaperones[12], including GrpE (human Mge1), Magmas (human Pam16 orthologue)[74], DnaJC15 (also known as MCJ) and DnaJC19[69,75]. A recent study suggests that two GrpE orthologous are present in human cells, namely GrpEL1 and GrpEL2[35]. The Tim44, a hydrophilic matrix protein is partially attached to the inner membrane[12]. The Tim44 is in the proximity of precursor proteins and may bring PAM subunits to their surroundings and therefore promoting the translocation[12]. Beyond the formation of a docking platform, the Tim44 also binds to the precursor being the first PAM component to make contact with the precursor transiting through the Tim23 channel[3,12]. The Pam18 human orthologues, DnaJC15 and DnaJC19, contains a J domain that stimulates the ATPase activity of Mortalin[12]. As human Tim17, DnaJC15 and DnaJC19 assemble into different TIM23 complexes, with DnaJC15 associating with translocase containing Tim17A, whereas DnaJC19 associates with Tim17B1 or B2[27,69]. In terms of functionality, these orthologues are not similar, since only DnaJC15 is capable of complementing Pam18-depleted yeast cells[75,76]. The DnaJC15/19 binds to hTim17A/B1/B2, and in yeast Pam17 binds to the Tim23 being a player in the organization of PAM-TIM23 complex[3,27]. The Magmas are J-related proteins that form a stable complex with DnaJC15/19 and control its activity[12]. The GrpE promotes the release of ADP from Mortalin and thus stimulates a new round of ATP binding and precursor translocation[3].

Depending on the precursor to be imported, the TIM23 complex couples with PAM or remains in its free form (Figure 3). If the proteins carry a hydrophobic sorting/stop transfer signal behind the matrix-targeting signal, it is only needed the free form for the lateral release into the lipid phase of MIM[3,12,77]. If precursors are directed to the mitochondrial matrix, the association with PAM is crucial. When the precursor is translocated through the TOM channel, the Tim50 binds directly to the precursor and promotes its binding to the IMS domain of Tom22[3,78–80]. Then, the Tim21 binds transiently to Tom22 inducing release of the presequence, which then attaches to the Tim23 channel[3,78,81]. The $\Delta\psi$ activates the Tim23 channel and generates an electrophoretic force on the presequence[3,25,71]. The association of the TIM23 with complexes III

(bc1-complex) and IV (cytochrome c oxidase) supports the Δψ-dependent insertion of preproteins[3,78,82–84].

As mentioned before, if the precursor has a hydrophobic sorting signal behind the matrix-targeting signal, the translocation stops and lateral release occurs. Otherwise, precursors are imported completely into the matrix and for the occurrence of that process, the import motor needs to associate with TIM23[3]. For complete import, Magmas and DnaJC15/19 associate with the TIM23 complex while it is still bound to the precursor[3,48,74]. In yeast, the Pam17 also assembles at an early stage and it is released when the precursor emerges on the matrix side[3]. The assembly of Tim44-Mortalin activates the motor inducing the binding to the precursor in translocation and drives it to the matrix[3,85]. The mitochondrial processing peptidase (MPP) removes the presequence during or after translocation into the matrix and protein achieves its active, folded form[3].

There are cases in which the precursors that contain tightly folded domains are laterally sorted into the MIM, but the import motor for unfolding these domains is also required. Here, the precursor depends on lateral sorting and motor activities and thus requires the association of TIM23 and PAM for a correct import[3].

As in TIM23 complex, PAM has a role in the assembly of the cytochrome c oxidase, namely through the association with specific factors for promoting the biogenesis of the complex IV[13,86,87].

1.4.3 TIM22 complex

The TIM22 complex has been extensively studied in *Saccharomyces cerevisiae* but the same cannot be affirmed about the mammalian TIM22 complex. Taking into account the information obtained for the diverse translocases of the different organisms, one can assume that the TIM22 complex is significantly different between lower and higher eukaryotes[48].

In yeast, this complex consists of the following central subunits: Tim22, Tim54, Tim18 and Sdh3[36,38,88–90]. Additionally, there is an association with small Tim proteins, namely Tim9, Tim10 and Tim12[12,15,39,47,91–94]. In humans, only Tim22 and small Tim proteins (Figure 4) are present[15,39,48,95]. The gene encoding hTim22 presents ~40% similarity with the yeast gene. There no evidences for the existence of clear homologous of Tim54 or Tim18 and that Sdh3 homologous (SDHC) interacts with the TIM22 complex[15,39,96]. Concerning to the small Tim proteins, the yeast Tim10 presents two human homologous proteins, Tim10a and the Tim10b[15,39,95].

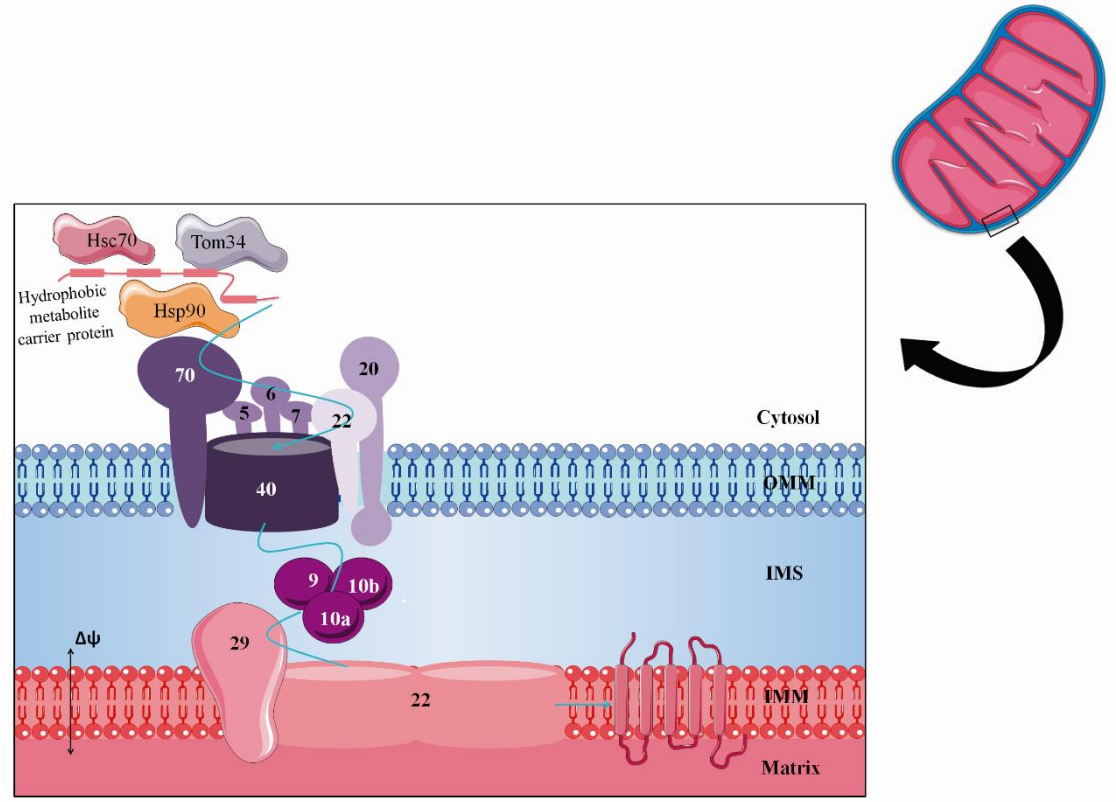

Figure 4. Import of hydrophobic metabolite carrier proteins of the inner membrane mediated by TOM and TIM22 complexes. MOM, mitochondrial outer membrane; IMS, intermembrane space; MIM, mitochondrial inner membrane; Δψ, electrochemical potential of the inner mitochondrial membrane; TOM, outer membrane translocase; Tom20, outer membrane translocase subunit 20; Tom22, outer membrane translocase subunit 22; Tom70, outer membrane translocase subunit 70; Tom40, outer membrane translocase subunit 40; Tom5, outer membrane translocase subunit 5; Tom6, outer membrane translocase subunit 6; Tom7, outer membrane translocase subunit 7; TIM22, carrier translocase; Tim22, carrier translocase subunit 22; Tim29, mitochondrial import inner membrane translocase subunit 29; Tim9, mitochondrial import inner membrane translocase subunit 9; Tim10a, mitochondrial import inner membrane translocase subunit 10a; Tim10b, mitochondrial import inner membrane translocase subunit 10b.

The Tim22 is the channel-forming subunit being the major component of the TIM22 complex and it is structurally related to Tim17 and Tim23 suggesting that both TIM complexes have evolved from a common ancestor[39,88]. The Tim54 appears to be required for the stability of Tim22 but it may not interact directly with the precursor in translocation[39]. The Tim18 is structurally related to the subunit IV of the succinate dehydrogenase (complex II of the MRC) and it might be involved in the assembly and stability of the TIM22 complex[3,39]. The small Tim proteins participate in the transport of metabolite carriers into the MIM, maintaining the precursor in a conformation that favors the import[92,95]. In humans, a complex formed by Tim9 and Tim10A associates with the precursor within the TOM complex through the binding to hydrophobic segments [15,39]. When the precursor is totally apart from TOM complex, it is delivered to another complex where Tim10B also associates with the complex [48,97]. Afterwards, the precursor is inserted into the translocase channel formed by Tim22 for the lateral insertion into the MIM lipid bilayer[3,39]. This insertion depends on the existence of the Δψ[3,39].

An additional human subunit has been identified, Tim29, which is an integral MIM protein and plays a role in the assembly of Tim22, leading to the stabilization of the TIM22 complex[15,37]. Furthermore, the Tim29 connects the TIM22 complex to the TOM complex[15,37].

1.4.4 MIA complex

The MIA complex (Figure 5) consists of a central component, Mia40[40] and Erv1[41,98]. A IMS protein, Hot13, also interacts with this complex[5]. In humans, Mia40 and Erv1 have been identified as CHCH4 and ALR, respectively[42,99]. The oxidoredutase Mia40 binds to the small IMS precursor proteins containing repetitive cysteine motifs, promoting their translocation through the MOM and the following assembly to the mature forms in the IMS [5,40,98]. The sulfhydryl oxidase Erv1 interacts with Mia40 allowing the biogenesis of the small IMS proteins mentioned before[5,41,98]. The Hot13 binds zinc ions used for oxidation of precursors maintaining Mia40 free of these ions and, therefore, allowing the re-oxidation of Mia40 through Erv1[3,5,43,98]. The main differences between yeast and human are the structure of Mia40/CHCH4 and the existence/absence of Hot13 and AIF[27]. The AIF is involved in the import of CHCH4. Therefore, it allows the correct function of the MIA complex[100]. In contrast with yeast form, the CHCH4 is not anchored to the MIM, being in a soluble form[42,48]. This change does not have a functional impact in CHCH4, maintaining the same abilities of the yeast form [42,48]. The human sulfhydryl oxidase ALR is similar to its yeast homologous[48]. The subunits from this complex act as a disulfide relay system[47] (Figure 5). The Mia40/CHCH4 forms transient intermolecular disulfide bonds with the precursor by using a redox-active CPC (cysteine–proline–cysteine), which is essential for its function and interacts with specific motifs of the precursor, when it emerges from the TOM complex[5,48,101,102]. Then, Mia40/CHCH4 promotes the formation of internal disulfide bonds within the precursor, leading to the removal of electrons and, consequently, the oxidation and release of the later into the IMS and reduction of cysteine residues in Mia40/CHCH4[3,5]. For the re-oxidation of cysteines of Mia40/CHCH4, it cooperates with Erv1/ALR that accepts electrons from Mia40/CHCH4 and transfers them via cytochrome c oxidase complex to the MRC and hence to O_2[3,5]. Alternatively, the Erv1 can also use O_2 directly, resulting in the production of hydrogen peroxide[3]. The active form of Erv1 is a homodimer, and an exchange of electrons between the two monomers has been identified as a central aspect of the reaction mechanism[5]. An alternative mechanism has been suggested for the protein oxidative folding. Here, Mia40, Erv1 and the precursor form a ternary complex, occurring, a direct cooperation between them[5,45,103]. In humans, the involvement of these mechanisms, independently or in cooperation, still needs confirmation[27].

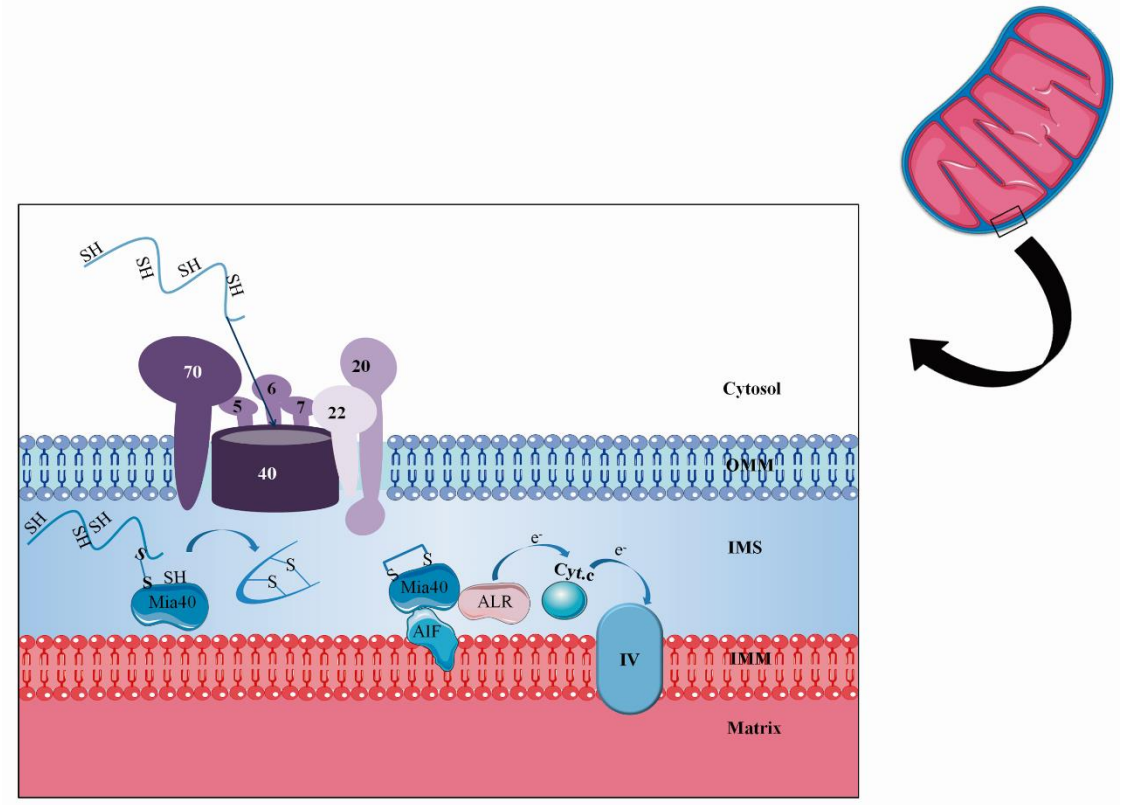

Figure 5. Import of precursors containing cysteine motifs is mediated by TOM and disulfide relay system. MOM, mitochondrial outer membrane; IMS, intermembrane space; MIM, mitochondrial inner membrane; e^-, electron; TOM, outer membrane translocase; Tom20, outer membrane translocase subunit 20; Tom22, outer membrane translocase subunit 22; Tom70, outer membrane translocase subunit 70; Tom40, outer membrane translocase subunit 40; Tom5, outer membrane translocase subunit 5; Tom6, outer membrane translocase subunit 6; Tom7, outer membrane translocase subunit 7; MIA, IMS import and assembly; Mia40, IMS import and assembly subunit 40; ALR, sulfhydryl oxidase; IV, cytochrome c oxidase; Cyt.c, cytochrome c; AIF, Apoptosis-inducing factor.

1.4.5 SAM complex

In yeast, the SAM complex consists of two essential proteins, Sam50 and Sam35[44,104]. The complex also includes a third subunit (Sam37), which is not essential to cell viability[3,45]. These proteins are conserved from bacteria to eukaryotes and the homologous protein of Sam50 has been identified in human mitochondria[59]. For Sam35 and 37, Metaxin 2 and Metaxin 1/3 have been suggested to be their functional substitutes, respectively, since they appear to be required for the biogenesis of β-barrel proteins in human cells[27]. Although a functional homology can be assumed, the direct sequence homology between metaxins and Sam35/Sam37 is still absent[18,27]. The Sam50, a β-barrel protein, represents the core of the SAM complex, forming a channel; its activity is induced by the internal targeting signal of the β-barrel precursor proteins[3,5,45]. The Sam35 (Metaxin 2, in humans) assists Sam50 in the specific response to the β-signal, recognizing this signal and inducing the opening of the channel[3,5,45]. The Sam37 (Metaxin 1/3, in humans) is involved in the release of precursor proteins from the SAM complex with the assistance of Sam50[3,5]. Kutik and colleagues (2008) proposed the following mechanism for the

translocation of β-barrel proteins: binding of the precursor β-signal to the hydrophilic, α-helice region of Sam35 located in the proximity of β-barrel domains of Sam50. This binding induces a conformational alteration, leading to the opening of the channel and the insertion of the β-strands. Then, the precursor is laterally released into the lipid phase of the MOM[5,45].

The interaction between Sam50, Metaxin 2, Metaxin 3 and MICOS has been reported[17,18,105,106]. The association of SAM complex subunits with MICOS leads to the formation of the MIB complex, which extend across the MIM and the MOM[19]. The MIB complex has important roles specifically in the maintenance of cristae structure and respiratory chain biogenesis[19,107].

Initially, the translocation of the precursors of β-barrel proteins is mediated by TOM complex and, therefore, enter into the IMS (Figure 6)[3]. In yeast, the small TIM proteins bind to the precursor and transport it to the surroundings of SAM complex[3]. In human mitochondria, the delivery to SAM complex by these small proteins was not reported[27]. In fact, it is still unknown if the precursors are delivered to the complex from the IMS[27]. The β-barrel precursors are inserted into the channel formed by Sam50 and, afterwards, released into the MOM[3].

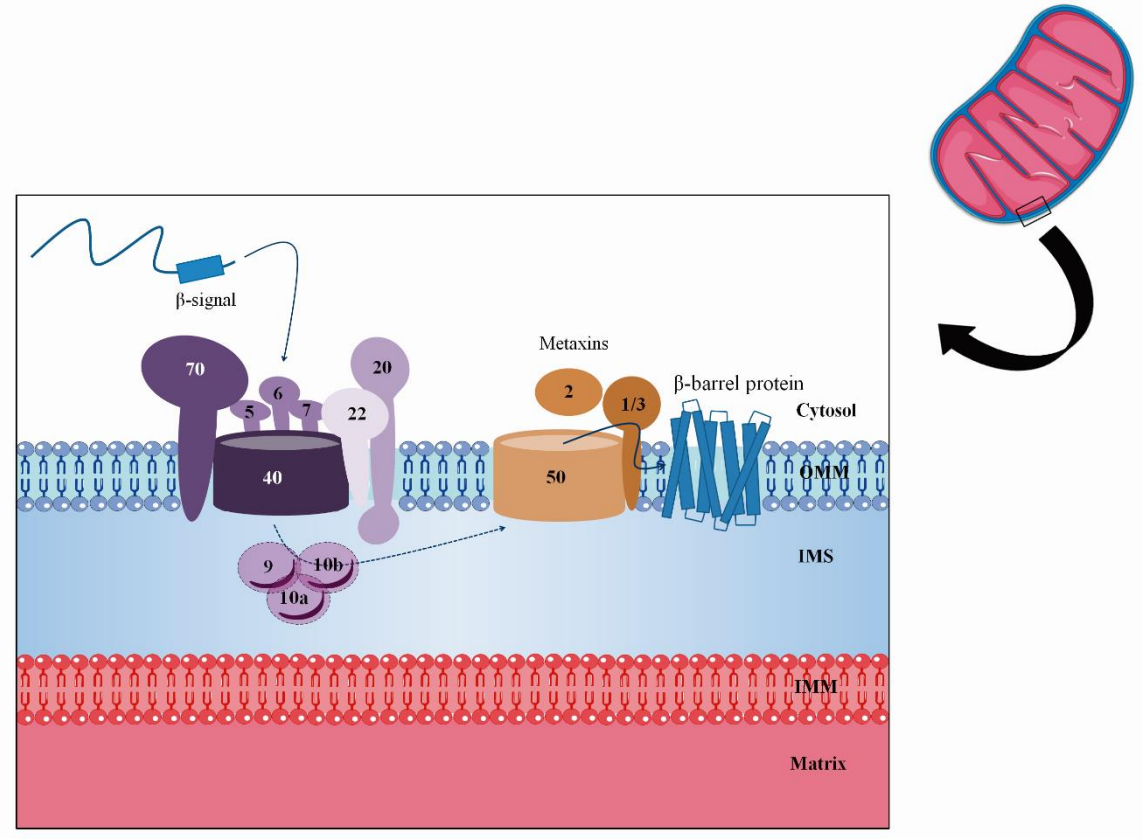

Figure 6. Insertion of β-barrel proteins into MOM is mediated by SAM complex. MOM, mitochondrial outer membrane; IMS, intermembrane space; MIM, mitochondrial inner membrane; TOM, outer membrane translocase; Tom20, outer membrane translocase subunit 20; Tom22, outer membrane translocase subunit 22; Tom70, outer membrane translocase subunit 70; Tom40, outer membrane translocase subunit 40; Tom5, outer membrane translocase subunit 5; Tom6, outer membrane translocase subunit 6; Tom7, outer membrane translocase subunit 7; SAM, sorting and assembly complex; 50, SAM subunit 50; 2, Metaxin 2; 1/3, Metaxin 1/3; Tim9, mitochondrial import inner membrane translocase subunit 9; Tim10a, mitochondrial import inner membrane translocase subunit 10a; Tim10b, mitochondrial import inner membrane translocase subunit 10b.

1.5 Post-translational modification of mitochondrial precursor proteins

Mitochondrial proteins are imported into the mitochondria in their precursor form that is the inactive form. For their maturation and consequent functional activation, the signals that allowed the targeting to the diverse subcompartments must be detached by proteolytic action and, in some cases, the precursors have to be folded. This processing is accomplished by mitochondrial peptidases. These enzymes are involved not only in proteolytic processing, but also in other biochemical processes essential for the mitochondrial function, including the maintenance of protein quality. This control consists of degrading damaged proteins and the regulation of mitochondrial gene expression and biogenesis by processing key components of the pathways[108,109]. In addition, these proteases regulate mitochondrial dynamics, mitophagy and apoptosis[108]. Here, the role of the mitochondrial peptidases involved in precursor processing is addressed.

Upon import into the matrix, most presequences are removed by proteolytic activity of MPP[3,13]. The MPP is a heterodimer localized in the mitochondrial matrix and it is composed by two homologous subunits: mitochondrial-processing peptidase subunit-α (α-MPP, PMPCA in humans) and mitochondrial-processing peptidase subunit-β (β-MPP, PMPCB in humans)[3,13,108]. Both subunits are highly conserved and essential[109] but only PMPCB has cleavage activity and it is a metallo, intrinsic protease[108]. The PMPCA recognizes and binds the presequences of precursors whereas PMPCB cleaves the mitochondrial import signal[108].

In proteins targeted to MIM, which have the hydrophobic sorting signal located after the matrix-targeting signal, the hydrophobic signal is usually responsible for the attachment of the mature protein to the membrane and, consequently, the cleavage does not occur[3].

For several IMS proteins, the hydrophobic sorting signal is cleaved off by an integral-MIM protease, the inner membrane peptidase (IMP), which contains the catalytic domain in the IMS, and thus, the mature protein is released into the IMS[3,109]. The IMP is an intrinsic, hetero-oligomeric, serinic protease that consists of two catalytic subunits, Imp1 and Imp2, and the auxiliary protein Som1[108]. In human mitochondria, IMMP1L and IMMP2L are the homologous of Imp1 and Imp2, respectively[108]. The IMP can cleave presequences, independent of MPP, but often after processing by MPP[109].

Some mitochondrial matrix and MIM proteins require additional cleavage events for total maturation and stabilization. These events only take place after the MPP processing. The mitochondrial intermediate peptidase MIP (Oct1, in yeast) and X-Pro aminopeptidase (XPNPEP3, Icp55 in yeast), which are intrinsic, and metallo proteases, remove an additional octapeptide sequence and one amino acid from the N-terminal, respectively[3,108,109]. These enzymes only

recognize and cleave their substrate after the removal of presequences[3,109]. The analysis of the N-terminal of mature proteins in yeast and plants using proteomic studies has shown an increasing number of substrates of these proteases, and also allowed the study of their cleavage sites[110–112].

The human MIP is encoded by *MIPEP* that is differently expressed in human tissues being highly expressed in heart, skeletal, muscle and pancreatic tissues[113] and, predominantly, in tissues with high oxygen consumption and energy demand[114]. The expression of this gene is potentially regulated by nuclear factor kappa-B (NF-kB) and activated by protein-1 (AP-1), which responds to redox cell state, since the promoter regions contain binding sites for these factors[114]. The promoter regions of this gene also contain specific sequences called Mt1, 3 and 4, which are present in nuclear genes involved in the oxidative metabolism and are also present in the region responsible for the mtDNA control[114]. The genomic region where the gene is located, has been associated with a form of muscular dystrophy, pancreatic carcinoma and deafness. More recently, a genome-wide association study (GWAS) in Han Chinese population identified a lung cancer susceptibility region where MIPEP is located[115]. Finally, Eldomery and colleagues (2016) demonstrated, for the first time, the involvement of MIP in human disease[116]. Using Whole-Exome Sequencing (WES), recessive variants of *MIPEP* were found in patients with left ventricular non-compaction, a disorder characterized by a failure in the development of a compact myocardium[116].

The human MIP is capable of performing efficiently as substitute of the yeast MIP (Oct1), revealing a conserved function of this protein between yeast and human[114].

The substrates of MIP and their roles are quite diverse. These substrates are involved in MRC, TCA cycle, the urea cycle, genome preservation and translation, protein folding and stress response[117]. In human mitochondria, the cytochrome c oxidase subunit 4 and the ornithine transcarbamylase are reported subtrates and more recently, the processing by this protease has been suggested to be required for total functional activation of sirtuin 3, which is responsible for the regulation of acetylation levels of metabolic enzymes[118,119]. Evidences also suggest that MIP may have a role in the regulation of the mechanism of action of Notch, which determinates the cell fate as a result of certain cellular signals[120]. Notch and gamma-secretase can also regulate mitochondrial proteins, including MIP that reveals a mutual regulation[120]. This regulation requires further studies to elucidate the molecular mechanisms involved and its impact in the cellular fate[120]. The second processing by MIP takes place when, after the MPP processing, the proteins have a phenylalanine, leucine or tyrosine residue in their N-terminal[112]. The destabilization of these proteins, which is caused by bulkiness of the mentioned amino acid residues and explained by the N-end rule of the protein degradation, makes this processing

necessary[112,117,121,122]. The N-end rule demonstrate that the amino acid residue of the amino terminal of proteins determine their in vivo half-life span[123]. In summary, the Oct1/MIP removes destabilizing residues from the N-terminal of precursors leading to the formation of mature proteins with a stable N-terminal and, therefore, acting as a quality control component of precursors undergoing double processing[112,117,124].

In addition to the protein processing, an interaction between MIP and frataxin has been suggested and, consequently, a role in iron usage[114].

1.6 Leber's Hereditary Optic Neuropathy (LHON)

1.6.1 Clinical presentation

Leber's hereditary optic neuropathy (LHON) is a rare maternally inherited disorder, although it is one of the most common mitochondrial diseases[125,126]. It is characterized by loss of retinal ganglion cells (RGCs) and degeneration of the optic nerve[127], leading to sudden and painless loss of central or total vision[126–128]. The first symptom is blurred or clouded vision in both eyes or, typically, in one eye and in the next few weeks or months, the second eye is also affected. Normally, LHON presents in males during adolescence or young adulthood but cases in which both genders' patients present a diverse range of ages of onset have been reported[126,128,129]. The disease is characterized by a rapid progression with small probability of partial visual recovery[127,128]. Typically, vision loss is the only clinical presentation. However, other anomalies can appear in LHON patients, namely cardiac and neurologic manifestations, and those cases are defined as LHON-*plus*[126,129].

1.6.2 Genetic context

1.6.2.1. mtDNA

About 90% of LHON cases with genetic cause identified are due to one of three primary mtDNA point mutations in genes affecting mitochondrial complex I subunit (NADH dehydrogenase) of the MRC, as follows: *MT-ND1* (m.3460G>A), *MT-ND4* (m.11778G>A), or *MT-ND6* (m.14484T>C)[130–132]. Additionally, other 16 rare primary mtDNA mutations affecting the same genes were identified as genetic causes of LHON. Other 19 mtDNA mutations present in genes encoding complex I subunits, ATP synthase 6, cytochrome c oxidase, cytochrome b and ATP synthase 8 have been suggested as candidate genetic variants being found in one individual or family (https://www.mitomap.org/foswiki/bin/view/MITOMAP/MutationsLHON#21) (Bacalhau, M. *et al*., unpublished data). Spontaneous visual recovery is more common in patients with the mutation m.14484T>C, with a partial recovery rate of 37%-58%, intermediate in the mutation m.3460G>A, with ~20% partial recovery rate; the mutation m.11778G>A has the lowest partial recovery rate of 4%[133]. Nevertheless, the majority of patients with characteristic clinical symptoms remain without a genetic cause identified.

1.6.2.2. nDNA

The mtDNA mutations identified as genetic causes of LHON cannot fully explain the pathogenicity of the disease. Since the mitochondrial function is highly dependent of nuclear encoded-proteins, nuclear gene factors have been suggested to be involved in this pathogenicity.

The higher prevalence of LHON in males suggested an influence of chromosome X genes in the disease. This influence was suggested in several genetic linkage studies revealing the involvement of modifier locus located in certain regions from this chromosome, in the manifestation of LHON phenotype[134–136]. Additionally, the inactivation of X chromosome could led the reduced prevalence of the disease in female since the LHON X-linked locus could be inactivated[134].

Additionally, polymorphisms in *EPHX1* and *p53* gene may be related with an earlier age onset of LHON patients with m.11778G>A mutation[137]. The *EPHX1* gene encodes a protein responsible for the hydrolysis of epoxides formed from the oxidative metabolism of endogenous or exogenous components, such as tobacco smoke components[137]. The reduced activity of the altered protein can compromise the processing of epoxides and their presence in the cell may cause the oxidative damage of neurons[137]. The p53 protein is involved in the apoptosis pathway in mitochondria and polymorphic gene variants affecting this protein potentiated cell apoptosis. Moreover, in a genome-wide linkage study of Thai LHON individuals, presenilins-associated rhomboid-like (*PARL*) gene was suggested to act as a potential nuclear modifier in LHON pathogenicity[138]. This association could not be proven in Chinese LHON patients[139]. The protein encoded by this gene is involved in apoptotic events and activate the proteolysis of proapoptotic proteins accumulated in the MOM[138].

A downregulation of *OPA1* was reported in LHON patients with m.11778G>A mutation suggesting a role in the disease pathogenicity[140]. The *OPA1* encoded protein is involved in the regulation of mitochondria shape, which is essential for the mitochondrial fusion[141].

The *YARS2* gene encoding the mitochondrial tyrosyl-tRNA synthetase was also suggested as a susceptibility gene in LHON individuals with the m.11778G>A mutation[142]. A c.572G>T variant in *YARS2* gene, causing an amino acid alteration in a conserved catalytic domain, was identified[142]. The presence of the mutated YARS2 protein compromised the mitochondrial translation in lymphoblastoid cell lines derived from patients [142].

1.6.3 Biochemical context

Taking into account that the primary mtDNA mutations occur in genes encoding subunits from the complex I of the MRC, one can assume that an impairment of the function in this complex could exist and, thus, a defective cell metabolism and failure in ATP production[126]. However, the biochemical phenotypes are variable according to the mutation present[143]. The presence of m.3460G>A primary mutation resulted in a substantial reduction of the complex I activity[127,144,145]. Several studies revealed the absence of a significant decrease of the activity of complex I when the m.11778G>A and m.14484T>C were present[127,145]. Other studies also reported the existence of complex I deficiencies in patients with the three primary mutations mentioned[146]. The phenotypic variability observed may be explained by its evaluation using different parameters, biological samples and analysis methods, since they may influence the results obtained[147].

1.6.4 "Incomplete penetrance" in LHON

Several cases show that individuals carrying a LHON mutation do not always present the disease. This is designated as "incomplete penetrance" and only ~50% of male and ~10% of female mutation carriers develop the symptoms[128,131]. Even when the pathogenic LHON mutations are homoplasmic (100% of mutant mtDNA) in all maternally-related individuals, only some individuals will express the disease phenotype[132]. Although the mtDNA mutation is a requirement to cause LHON, one can assume that it is not sufficient to cause the disease, leading to variation of disease "penetrance" in different families with the same mutation, and even within the same family[132,148,149]. Evidences suggest that mtDNA content and mitochondrial biogenesis are associated with the "incomplete penetrance" in LHON families[150]. The presence of modifier genes in nDNA and environmental causes may also contribute to the "incomplete penetrance" observed in LHON[149]. The vision loss in carriers can be related with smoking and alcohol consumption habits[131]. The female hormones, in particular estrogens, could also have a neuroprotective effect on RGCs function, and this feature could account, at least partly, for the relative protection of female LHON carriers against visual loss[151]. The presence of these hormones in females can contribute to the gender bias observed in LHON[151].

2. AIMS OF THE STUDY

The aim of this study was the identification of genetic variants present in genes encoding subunits and proteins involved in the protein import into human mitochondria and in the post import processing of precursor proteins, which may contribute to the "incomplete penetrance" verified in mtDNA mutations associated with LHON. The search of these variants in whole-exome sequencing (WES) data was performed and the functional impact prediction of the relevant variants was evaluated. The most promising variants were confirmed using two additional methods and the impact at the protein level was evaluated.

3. MATERIALS AND METHODS

3.1. Samples and controls

Total DNA from 29 samples (six samples from LHON patients with one of the primary mtDNA mutations, eight samples from patients without mtDNA mutation identified and fifteen samples from carriers of one mtDNA mutation) was previously extracted at the Laboratory of Mitochondrial Biomedicine and Theranostics (BioMiT lab) (previously known as Laboratory of Biochemical Genetics (LBG)).

Lymphocytes from control individuals and LHON patient were previously prepared at BioMiT lab. Skin fibroblasts from human control were also prepared.

Total lysates from mouse cortex and testis were kindly provided by Master Catarina Neves, Coimbra Institute for Clinical and Biomedical Research (iCBR), Faculty of Medicine, University of Coimbra, Coimbra, Portugal.

Total lysate from WiDR cells was kindly provided by researcher Salomé Pires PhD, Biophysics Institute, CNC.IBILI, Faculty of Medicine, University of Coimbra, Coimbra, Portugal.

3.2. Whole exome sequencing (WES)

The WES is a next generation sequencing (NGS) technique for analysis of the protein encoding regions and the exonic-intronic boundaries. The presence of alterations in these regions may influence the maturation of pre-mRNA. The sequencing, and following comparison with a human genome reference sequence, allows the identification of exonic and splice-site variants and also distinguish between common and rare, possible pathogenic variants.

The WES analysis was performed by an external service provided by ®STABVIDA, Lda. Briefly, the DNA integrity was assessed via electrophoresis in agarose gel (1%) in TBE buffer and the DNA was quantified by fluorometry. Then, the Illumina Nextera rapid capture exome kit v1.2 was used for the library construction and exome enrichment. The libraries were sequenced using a HiseqPE (100) cluster kit v4 in the platform Illumina Hiseq2500. After sequencing, the results obtained were analyzed using CLC Genomics workbench v9.0.1 with the following bioinformatics pipeline: alignment with a human genome reference GRCh38; variant calling taking into account its quality; variant filtering against 1000 genomes, phase 3 and ClinVar (v07-07-2016) databases; *in silico* predictive algorithms (amino acid changes and splice site effect).

3.3. Search of nuclear gene variants

The search for variants was focused in 46 genes encoding proteins involved in the nuclear-encoded mitochondrial proteins import systems and post-translation processing of precursor proteins (Table 2). The genomic location of the various genes was searched in NCBI and the variants located in these regions were selected.

Table 2. Human genes encoding the various subunits and proteins of mitochondrial proteins import systems and post-translation processing of precursor proteins

TOM	TIM23	PAM	TIM22	MIA	SAM	PROTEASES
TOMM20	*TIMM23*	*HSPA9*	*TIMM22*	*CHCHD4*	*SAMM50*	*PMPCB*
TOMM20L	*TIMM23B*	*TIMM44*	*TIMM29*	*GFER*	*MTX2*	*IMMP2L*
TOMM22	*TIMM17A*	*PAM16*	*TIMM9*	*AIF1*	*MTX1*	*PMPCA*
TOMM70	*TIMM17B*	*DNAJC19*	*TIMM10*		*TIMM9*	*IMMP1L*
TOMM40	*TIMM50*	*DNAJC15*	*TIMM10B*		*TIMM10*	*MIPEP*
TOMM40L	*TIMM21*	*GRPEL1*			*TIMM8A*	*XPNPEP3*
TOMM5		*GRPEL2*			*TIMM8B*	
TOMM6					*TIMM13*	
TOMM7						
TOMM34						
HSPA8						
HSP90AA1						

3.4. *In silico* analysis

The Mutalyser® nomenclature checker was used to determine if the variants correspond to an intronic or exonic region, in order to know which variants can be analysed for the prediction of functional changes. The variants found were analyzed using a set of bioinformatics tools that allows the prediction of variant pathogenicity, namely PolyPhen-2®, PROVEAN®, SIFT®, MutationAssessor®, Mutation Taster® and NNSplice®[152,153]. PolyPhen-2® prediction is based in the amino acid sequences and protein tridimensional structure. The variants are classified as "benign", "possible damaging" and "probably damaging"[154]. This tool uses two prediction models, HumVar and HumDiv. In this study, HumDiv was used because it is the most appropriate model for the study of rare alleles involved in complex phenotypes, which is the case of LHON. The PROVEAN® tool predicts the functional impact of substitutions, insertions, deletions and multiple substitutions considering the conservation, substitution frequency and the variant neighborhood classifying the variants as "neutral" or "deleterious"[155,156]. The SIFT® prediction is

based on the sequence homology and the variants are classified as “tolerated” or “deleterious”[157]. The Mutation Assessor® predicts the functional impact of a genetic alteration taking into account patterns of evolutionary conservation derived from families and subfamilies of homologous sequences and classifies the variants as “neutral”, “low”, “medium” or “high impact”[158,159]. The Mutation Taster® allows a rapid evaluation of the disease-causing potential of DNA sequence alterations considering evolutionary conservation, splice-site changes, loss of protein features and changes that might affect the amount of mRNA and classifies the variant as “polymorphism” or “disease-causing”[160]. The NNSplice® predicts the donor and acceptor splicing sites (5’ and 3’, respectively) in the exon-intron boundaries of protein encoding-genes with a score that reflects the probability of occurring splicing upon the sequence variation occurrence[161].

For the variants detected in coding regions, the first analysis for functional impact prediction was made using PROVEAN® and SIFT® (J.Craig Venter®). In addition, these tools also give information about the amino acid sequence changes, the protein ID, codon and strand of DNA (sense or antisense), the length of the protein, the position and the amino acid change and also the annotation from dbSNP database.

The variants identified as non-synonymous were analyzed by other tools. The PolyPhen-2® prediction was performed, using the protein Uniprot ID, the position of the altered amino acid, the reference amino acid (A1) and the altered amino acid (A2). Then, in order to proceed with the analysis using Mutation Assessor®, the version of the human genome, chromosome and position of the variant, reference nucleotide and altered nucleotide were required. The prediction can also be made using the protein Uniprot ID, the position of the amino acid sequence, the reference and the altered amino acid.

At last, it is necessary to enter the gene name, Ensembl transcript ID, the altered position and the new base or the positions of the bases before and after the alteration and the inserted bases, for the Mutation Taster® prediction.

According to the prediction, the variants classified as having a functional impact at PROVEAN® (deleterious – score<-2.5), SIFT® (damaging – score<0.05), PolyPhen-2® (possibly/probably damaging), MutationAssessor® (medium/high impact) and MutationTaster® (Disease-causing) were retained for further analysis.

For variants detected in intronic regions, in the proximity of exon-intron boundaries, the splicing effect analysis was performed using NNSplice. From the variants with possible splicing effect, those presenting gain or loss of a donor or acceptor site and/or score alteration were also retained for a more detailed study.

Afterwards, the assigned variants with a predicted functional impact by all or in the majority of the bioinformatics tools were selected. Then, the minor allele frequency (MAF) was searched in Exome Aggregation Consortium (ExAc) browser, and together with the functional impact prediction, contributed for the selection of the promising variants, namely those with a low MAF (<0.05).

3.5. Identification of variants with functional impact using the ExomeLoupe® software

In addition to the manual identification of the gene variants with potential functional impact, the identification was also performed for the sample containing the promising variants using the genetic variants prioritization software, ExomeLoupe® developed by the Genoinseq (http://genoinseq.com/pt/home/servicossequenciacaoservicos/servicos/). This software performs the *in silico* analysis of exome sequencing data using the bioinformatics tools for the prediction of variant functional impact, and presents numerous informations related to all the variants identified, namely the name of the gene where the variants are located, chromosomic location, reference and the altered allele and variant genotype (homo or heterozygous). The software also presents the transcript where the variant has the uppermost impact, the predicted amino acid change and the degree of functional impact (low, medium, high). Additionally, it presents the reference cluster ID (rs), allele frequencies in the ESP and 1000 Genomes project, clinical significance and diseases associated with the variants.

For the identification of variants from genes encoding the mitochondrial protein import systems and precursor protein processing, the following filters were applied: a list with the name of the genes under evaluation, variants with a medium or high impact and genotype of the variants (hetero or homozygosity). The most promising variants were additionally validated using PCR-RFLP and Sanger sequencing.

3.6. Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP)

The polymerase-chain reaction (PCR) is a molecular technique developed by Kary Mullis in the 1980s[162]. It allows the amplification of a specific DNA sequence from various sources since simple prokaryotes to higher organisms, as humans. The PCR allows the amplification of the desired DNA in an exponential manner, starting with a small amount of DNA, which is a valuable

advantage since it enables the yield of an elevated number of DNA copies and therefore, allowing the performance of a myriad of additional techniques, such as sequencing, Restriction Fragment Length Polymorphism (RFLP) and many others[162–164]. The combination of PCR and RFLP is widely used to search for DNA variants. Besides being a sensitive method, PCR-RFLP is also simple for the detection of variants[165]. The fragment with the alteration is amplified via PCR, enzimatically digested using a suitable endonuclease and, then, separated by agarose gel electrophoresis.

3.6.1. PCR

The amplification via PCR consists in three steps: denaturation, annealing and extension, which are repeated in each PCR cycle[166]. An additional step of denaturation and extension is used in the beginning and at the ending of PCR, respectively[166]. Each step occurs in a different temperature, which leads to a specific outcome. First, the initial denaturation ensures the total separation of the double-stranded template for proper amplification. Then, the cycling starts with the denaturation allowing the separation of the double strand of DNA template via the breakdown of hydrogen bonds between the complementary nucleotides of each strand. Then, the annealing step occurs with the binding of each primer to each single stranded DNA template. The primers must be designed considering several characteristics using, for example, the software Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/)[167–169]. After the primers annealing, the DNA polymerase is able to start the extension of the new DNA strand only in the presence of a double stranded template. The extension occurs by the addition of deoxyribonucleotides triphosphate (dNTPs) to the growing strand.

To amplify the fragments of interest, the PCR was optimized to maximize the amplification of the desired products. Since the PCR specificity and yield are deeply dependent on the annealing temperature, the optimization of this parameter was critical[170–172]. When the annealing temperature is too low, the appearance of unspecific products is highly probable[171,172]. On the other hand, when the temperature is too high, the reaction yield can be reduced or even inexistent[171,172]. The annealing temperature is closely related with the primers used, namely the length and composition of the primer sequence and its concentration[167]. In order to determine the optimal temperature, a gradient of temperatures was used in the C1000™ Thermal Cycler (Bio-Rad). Even in the presence of the optimal temperature, the unspecific products may still be present and/or the reaction yield insufficient. To address this situation, the PCR cycling conditions, as the initial denaturation, were optimized. Here, the time and temperature were extended and increased, respectively[173]. These variations allowed the total denaturation of the

template, which enhanced the binding of primers to the correct sites within the template and, therefore, it reduced the unspecific binding and formation of primer dimmers, increasing the reaction specificity. The fragment from *TOMM20L* (one of the genes in study) presents an elevated percentage of GC content (≥60%); thus, an additive, dimethylsulfoxide (DMSO), was added to the reaction[174]. Increasing percentages of DMSO (0-10%) were added to the reaction mix to determine the optimal percentage to use. The DMSO acts in the base pairing of the template leading to its disruption [172,175].

The fragments were amplified by PCR in a final volume of 20µL containing 20-50ng of genomic DNA, 1x PCR reaction buffer (with 1.5mM $MgCl_2$) (GE Healthcare), 50µM of forward and reverse primers (Invitrogen) (Table 3), 2mM of deoxyribonucleotide triphosphates (dNTPs) (GE Healthcare), 1 U of Taq Polymerase (GE Healthcare) and 10% DMSO (Sigma) that was only used for the for *TOMM20L* fragment amplification. The cycling conditions are described in Table 4.

Table 3. Primer sequences (designed using the Primer3 software) for amplification of fragments from *MIPEP* and *TOMM20L*

Gene	Primer sequence	
	Forward	Reverse
MIPEP	5'-TGT GGA GTG ATA TGG GGA GTA-3'	5'-CCC AAT GGA TCT CAG TTT CAG A-3'
TOMM20L	5'-GAC AGT GAG TGG GAC CGA G-3'	5'-GTG AGA GGA GAG TGA CTG GG-3'

Table 4. Cycling conditions for amplification of *MIPEP* and *TOMM20L* fragments

Steps	Temperature		Time	Cycles
Initial denaturation	95ºC		5 min	1
Denaturation	95ºC		1 min	35
Annealing	60ºC (*MIPEP*)	63.5ºC (*TOMM20L*)	30 sec	
Extension	72°C		30 sec	
Final extension	72ºC		10 min	1

3.6.2. Agarose gel electrophoresis

For the evaluation of the PCR amplification, namely the presence of the desired DNA fragments in a considerable amount and the presence of unspecific products, the samples were run in a 2% agarose gel by electrophoresis. This type of electrophoresis is an effective and efficient method for the identification of the PCR-amplified DNA fragment[176]. In the presence of an electric field, the DNA fragments migrate to the positive electrode and they are separated

according to their size, since the mass and charge are in the same proportion[177]. The gel may be prepared in various percentages of agarose depending on the expected product. This percentage reflects the size of the pores within the agarose matrix. The increase of agarose concentration reduces the size of the pores, and consequently, it promotes a better separation of fragments of DNA. To detect the DNA fragments, ethidium bromide (1mg/mL) was used. This chemical agent intercalates within the base pairs from the DNA double strand and it fluoresces when it is exposed to UV light[178]. The detection and analysis of the gel were performed using ChemiDoc™ XRS+ System (Bio-Rad).

3.6.3. RFLP

For the enzymatic digestion, a proper enzyme must be chosen. The creation or removal of a restriction site as a result of the variant occurrence in a different band pattern after digestion, allowing the identification in the agarose gel electrophoresis. In addition, the fragments obtained from the restriction analysis must have different sizes for a good separation and identification. The choice of the endonuclease was made using software available online, namely the PCR-RFLP simulator (http://insilico.ehu.es/restriction/main/) and NEBcutter (http://nc2.neb.com/NEBcutter2/). Whereas the simulator compares the normal and altered sequence presenting the restriction enzymes that digest the sequences and lead to a different restriction pattern, the NEB cutter only indicates the restriction enzymes and the patterns for each of the sequences. So, in this case, the user has to identify the best restriction enzyme. The resulting band pattern corresponds to the two alleles and therefore, the zygosity of the variant has to be taken into account. When the variant is present in heterozygosity, the modification of the band pattern is only caused by one of the alleles, so the bands' pattern observed is the sum of both mutant and wild-type patterns.

The resulting fragments from *MIPEP* and *TOMM20L* were digested using AvaII and BrsBI, respectively (New England Biolabs, USA). The digestions were performed in a final volume of 20µL containing 15µL of PCR products, 1× buffer and 1U or 5U of BrsBI and AvaII, respectively. After overnight incubation at 37°C, with agitation, the samples were submitted to electrophoresis in a 3% agarose gel, stained with ethidium bromide. The RFLP pattern was then evaluated under UV light. The sequences recognized by the enzymes and the expected bands' patterns are described in Table 5.

Table 5. Restriction enzymes and expected band patterns.

***MIPEP* (475 bp)**				
Restriction enzyme	**Recognition sequence**	**Band pattern**		
AvaII	G'GWC_C	**GG (Wild-type)**	**GA (heterozygous)**	**AA (mutated homozygous)**
		221 bp 254 bp	475 bp 221 bp 254 bp	475 bp
***TOMM20L* (429 bp)**				
Restriction enzyme	**Recognition sequence**	**Band pattern**		
BrsBI	CCG'CTC	**AA (Wild-type)**	**A/delA (heterozygous)**	**delA/delA (mutated homozygous)**
		429 bp	429 bp 240 bp 188 bp	240 bp 188 bp

3.7. Sanger sequencing

The Sanger sequencing, also known as chain termination method is a DNA sequence determination method that was developed by Frederick Sanger and his colleagues in 1977[179]. Nowadays, the majority of the sequencing steps are performed using automated sequencers[180].

Initially, the amplification of the interest fragments were performed using PCR with the previously optimized conditions. Subsequently, amplification was confirmed using a 2 % agarose gel electrophoresis. Then, the PCR product was treated with the addition of ExoSAP-IT®, an enzymatic cocktail that inactivate the unincorporated dNTPs and primers though dephosphorylation and degradation, respectively[181]. This treatment was performed with the incubation of PCR product with ExoSAP-IT® at 37°C for 30 minutes and then at 80°C for 15 minutes to inactivate the enzymes. This incubation was carried out using the C1000™ ThermalCycler (Bio-rad).

Afterward, the PCR product is amplified using a sequencing PCR, which is similar to the conventional one but with some important differences. Firstly, the primers are used separately with the forward and reverse primer in independent reaction. Secondly, in addition to the dNTPs, dideoxynucleotides (ddNTPs) are also added to the reaction[182]. The ddNTPs do not have the free hydroxyl group at the 3'carbon and consequently, when are added to the DNA growing strand, the reaction is terminated. The ddNTPs are added to the reaction in a certain proportion that allows the formation of fragments with all lengths, between the length of primer and the complete DNA fragment[182]. Each ddNTP (ddATP, ddTTP, ddCTP and ddGTP) is bound to a different fluorophore that allows the identification of the ddNTP incorporated in the DNA strand[180,182]. The cycle sequencing was performed using the BigDye™ Terminator Cycle Sequencing v3.1 kit (Applied Biosystems), in a final volume of 10µL containing 2µL of PCR

product, 1x BigDye sequencing buffer, 0.32µM of forward or reverse primer and 1× BigDye™ Terminator Cycle Sequencing v3.1 premix. The used cycling conditions are presented in Table 6. The PCR product was purified using a Sephadex column, which allowed the separation of the fragments from the fluorescent dye-terminators by molecular exclusion, meaning that the smaller components are trapped in the porous beads of the column whereas the fragments can pass them and, consequently, they are isolated from the other reaction components. The columns were prepared by the addition of a resin of Illustra Sephadex™ G-50 Fine Grade to the Illustra™ Autoseq™ G-50 columns (GE healthcare) and then the columns were centrifuged at 2,000×g for 1 min. Subsequently, the samples were added to the columns and then, they were centrifuged again.

Table 6. Cycling conditions used in sequencing PCR

Steps (25 cycles)	Temperature	Time	Cycles
Denaturation	96°C	10s	25
Annealing	50°C	5s	
Extension	60°C	4min	

Afterward, the fragments were sequenced using the automated sequencer 3130 Genetic Analyzer (Applied Biosystems). The sequencer performs a high-resolution capillary electrophoresis where the fragments are separated by size allowing the distinction between fragments with a single nucleotide difference[180]. When the fluorescently labeled fragments pass the detection region, the fluorophore is excited by a laser and emits fluorescence at one of the four different wavelengths, which allows the identification of the ddNTP and therefore, the nucleotide at the end of the fragment[180]. The combination of the separation by length of the fragments and the order of detection of the ddNTPs allows the identification of the DNA sequence, since the ddNTPs are detected by the order of appearance in the DNA sequence. The purified sequencing PCR products were diluted in a 96-well plate by adding 1-3µL of product to 17-19µL of H_2O miliQ and then sequenced.

3.8. Western Blot

Western Blot is molecular method first described in 1979 and it is widely used for detection and analysis of proteins using antibodies[177]. The proteins are separated through an electrophoresis with denaturing conditions (SDS-PAGE) and then transferred to a membrane. This membrane is incubated with a primary antibody, which binds to the antigen (protein) and

after it is incubated with a secondary antibody containing an enzyme attached. This enzyme cleaves a chemiluminescent molecule leading to emission of light that is detected.

Lymphocytes (total lysate) from control individuals and patient were previously prepared. Skin fibroblasts (mitochondrial-enriched fractions) and WiDr cells (total lysate) from human, as well as cortex and testis (total lysate) from mouse were used as controls to verify the tissue specificity of Tom20L protein expression. The total lysates were centrifuged for separation of the total cellular protein from lipids that can difficult the identification of the desired bands. The samples were quantified using the Bradford method, which uses a reagent containing Comassie Brilliant Blue G-250 Blue that shifts the maximum absorbance from 465 to 595nm when protein is present[183–185]. This shift leads to the appearance of color, which intensity is proportional to the quantity of protein. The determination of the protein concentration was accomplished using a standard curve made with bovine serum albumin (BSA) standards with increasing final concentrations between 0 and 2mg/mL. The assay was performed in triplicate with final volume of 200µL for each standard and sample containing 78µL of H_2O miliQ, 1µL of BSA standard or sample, 1µL of sample buffer and 120µL of Bradford reagent. An incubation at room temperature was made and then, the absorbance was measured at 595nm using the SpectraMax® Plus 384 Microplate spectrophotometer and the total protein concentration was calculated. Volumes were adjusted according to the total protein concentration of each sample, loading buffer (Tris 1M, pH 6.8, 20% SDS (w/v), 40% glycerol (w/v), 0.02% bromophenol blue (w/v), 15% β-mercaptoethanol (v/v)) was added to the sample in a ratio of 1:1 (¼ loading 4× and ¾ sample) and afterwards, the samples were denaturated.

Next, the size-based electrophoresis was performed. First, a gel (Mini-Protean® TGX™ Precast gel, Bio-Rad) was assemble with the electrophoresis cell (Mini-PROTEAN® electrophoresis cell, Bio-Rad) and the assembly was filled with running buffer (Tris-glycine 1×, 0.1% SDS (w/v), pH=~8.25). The molecular marker Precision Plus Protein™ Dual Xtra Standards and 30-40 µg of total protein for each sample were loaded into the gel. The electrophoresis was performed at 30V at the beginning, and then at 120V for approximately 1h-1h30min. The gel was then removed and placed in transfer buffer (Tris-glycine 1x, 20% methanol (v/v)). Next, a PVDF membrane (Hybond P 0.5µM, Amersham) was activated in methanol, washed and equilibrate in transfer buffer. The transfer sandwich was assembled and placed in the transference module (Mini PROTEAN® tetra cell system, Bio-Rad) and the latter was filled with transfer buffer. The transference was performed at 4°C, 100V for, approximately, 1h30min-1h45min.

After the transfer, the membrane was blocked to prevent unspecific bond of antibody, for 1h at room temperature with a saline solution buffered with Tris (50mM Tris and 150mM

NaCl) containing 0.1% Tween 20 (v/v) (TBS-T 1×) and 5% skimmed milk. The blocking was followed by incubation overnight at 4°C, with rabbit polyclonal anti-human Tom20L at 1:300, or rabbit polyclonal anti-human MIPEP (Abcam) at 1:500, or rabbit monoclonal anti-human citrate synthase (CS, ab129095, Abcam) at 1:5,000, or mouse monoclonal anti-human actin (Santa Cruz Biotechnology) at 1:200, or mouse monoclonal anti-COX IV (ab14744, Abcam) at 1:5,000. Subsequently, the membrane was washed in TBS-T and incubated at room temperature for 1h-1h30min with a solution containing TBS-T 1x, 2.5% skimmed milk and the appropriate HRP-conjugated secondary antibody (Bio-Rad) at 1:5,000. After the incubations, the membrane was washed in TBS-T to remove the unbound antibody. Afterward, the membrane was incubated with a chemiluminescence substrate (Clarity Western ECL Substrate, Bio-rad) and the detection was carried out using the ChemiDoc™ XRS+ System (Bio-Rad). The membrane and, consequently, the bands were observed and quantified using the Quantity One® 1-D software (Bio-Rad).

4. RESULTS

4.1. Identification of genetic variants and *in silico* analysis for prediction of pathogenicity

The search of variants located in genes encoding the mitochondrial protein import machineries and proteases responsible for precursor protein processing after import, resulted in the identification of 87 relevant variants, on average, 54 variants in heterozygosity and 33 homozygous. On average, 64 variants are intronic, six variants are located in UTR sequences, specifically four in 3'UTR and two in 5'UTR, 11 variants are synonymous and six variants are nonsynonymous (Figure 7).

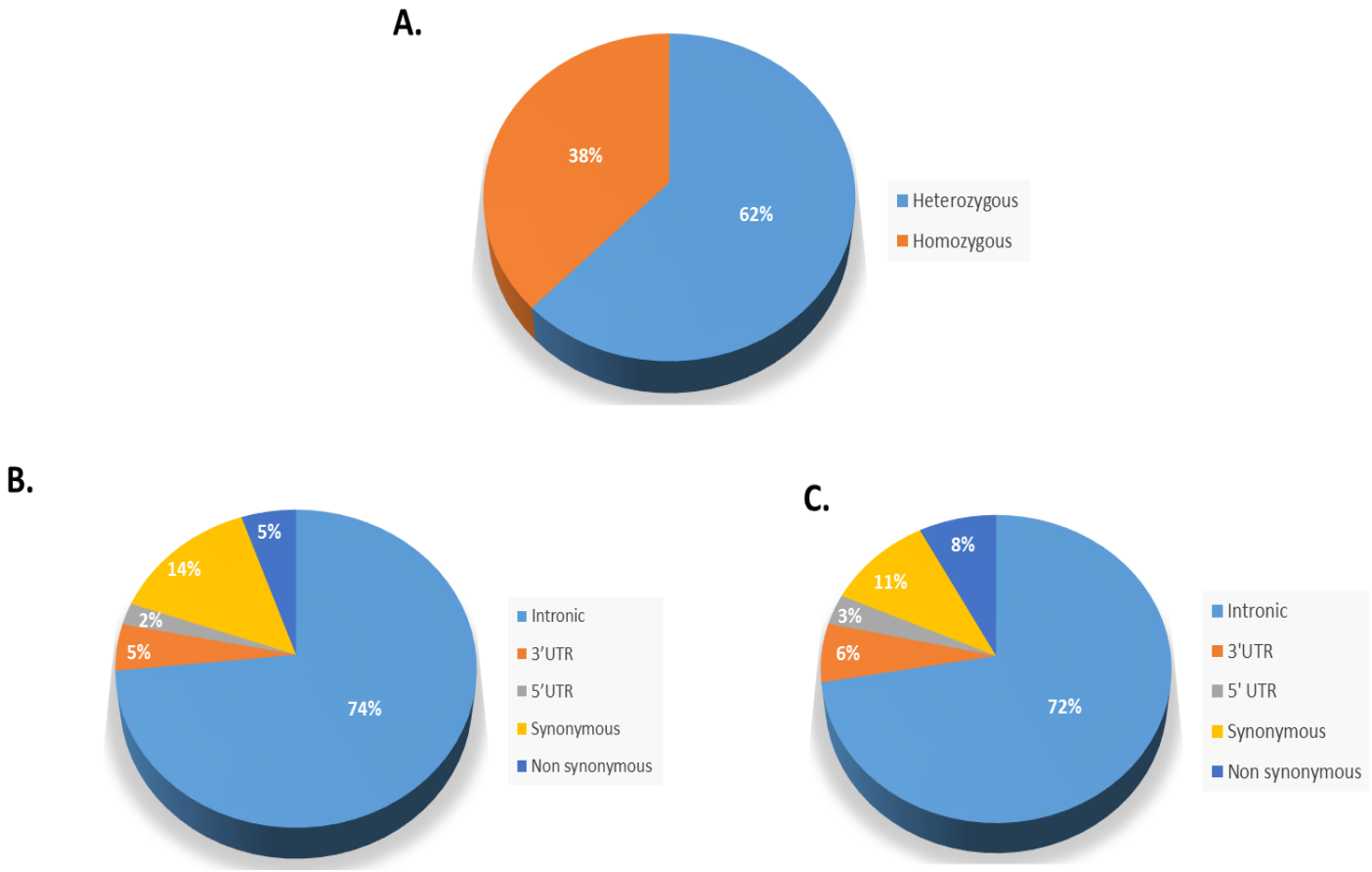

Figure 7. Genetic variants located in genes encoding proteins involved in the mitochondrial protein import and processing.
A. Zygosity of the total variants identified. **B.** Type of variants in heterozygosity. **C.** Type of variants found in homozygosity.

The analysis of the variants identified, using bioinformatics tools, allowed the identification of seven variants with a possible functional impact (Table 7). The variants that demonstrated a higher probability of being pathogenic were localized in *HSPA9*, *MIPEP* and *TOMM20L* encoding PAM subunit Mortalin, Mitochondrial intermediate peptidase (MIP) and TOM complex subunit Tom20L, respectively. Moreover, these variants are rare since they present a low MAF. Although these variants are heterozygous, they are present in the same patient (*MIPEP* and *TOMM20L* in sample 22) and are predicted to have a pathogenic effect. In addition, the MIP and Tom20L proteins encoded by the altered gene may have a negative functional influence in the respective proteins encoded by the normal copy of the gene. Taking these observations into account, these variants were chosen for further studies. The sample 22 belongs to a male individual, with clinical and molecular (m.14484T>C) diagnosis of LHON with age onset at 35 years old. The biochemical analysis using lymphocytes demonstrated a decrease of activity in the complex IV (56.4%) with higher values for complex II (128.3%), complex III (70.1%) and segment II+III (96.0%). Complexes I and V were not evaluated.

4.2. Identification of variants with functional impact using the ExomeLoupe®

The analysis of the sample belonging to the LHON patient with the promising variants revealed an additional important variant localized in the *TOMM20L* gene, namely a multiple deletion of five nucleotides (c.172_176delGGCAC), in heterozygosity. The variant present a low MAF (0.0003/32). This variant is located two nucleotides downstream of the c.170delA variant. The ExomeLoupe® analysis also identified the c.280C>T variant in *MIPEP* gene, previously identified by the manual analysis (Table 7).

4.3. Comparison of Tom20L sequences in the presence of the different variants

The variant localized in *TOMM20L* gene could have multiple effects in the protein sequence. Since the variants were present in heterozygosity, they could be localized in the same allele or in different alleles. These possible locations could distinctly influence the sequence of the protein leading to different outcomes as follows (Figure 8).

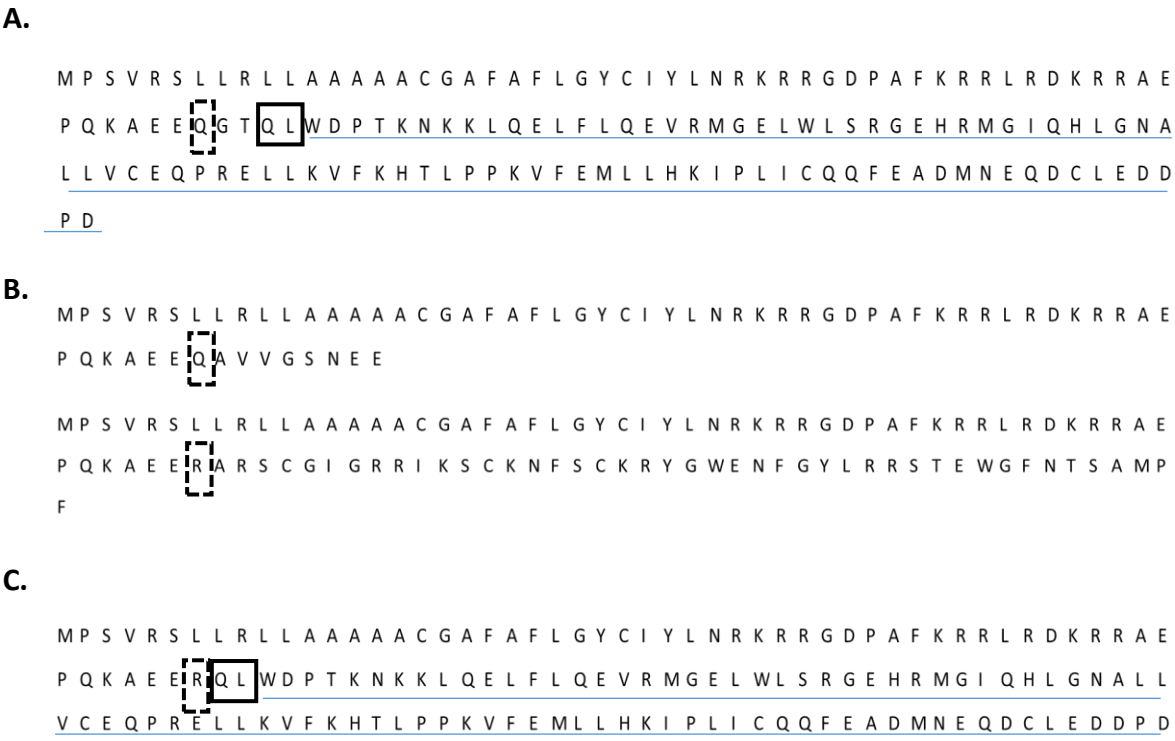

Figure 8. Possible effects of *TOMM20L* gene variants in the protein sequence. A – Wild-type protein sequence; B – protein sequence resulting from the presence of the variants in different alleles; C – protein sequence resulting from the presence of the variants in the same allele.

If the variants were present in different alleles (Figure 8. B), both alleles would encode a truncated protein with a different amino acid sequence after the positon 57. The allele containing the c.172_176delGGCAC would encode a smaller protein than the allele containing the c.170delA. If the variants were present in the same allele (Figure 8C), the c.170delA variant would cause a change of the glutamine (Q) residue to an arginine (R) residue, at the position 57. The c.172_176delGGCAC would led to the elimination of the glutamine (G) and threonine (T) residues and then, the rest of protein sequence would be unchanged (underlined sequence).

Table 7. Variants with a predictive functional impact identified in genes encoding proteins belonging to mitochondrial import systems and processing of precursors.

Variant (GRCh38)	Sample	Zygosity	Gene	Protein sequence alteration	PROVEAN (cut-off=-2.5)	SIFT (cut-off=0.05)	MUTATION ASSESSOR	POLYPHEN-2	MUTATION TASTER	dbSNP	NNSplice	MAF (ExAc)
1,201957275,T,G **Tim17A: c.27-6T>G**	1, 5, 7, 9 11, 16-18, 21, 23 29	Heterozygous	*TIMM17A*	–	–	–	–	–	Polymorphism	rs61821487 ClinVar: NA	**Acceptor site loss 1770 -1810 (S=0.88)**	0.07963
5,138 561 689,A,G **Mortalin: c.1073T>C**	2, 27, 28	Heterozygous	*HSPA9*	p.L358P	**Deleterious S=-6.80**	**Damaging S=0**	**High impact S=4.27**	**Probably Damaging S=1**	**Disease causing**	-	-	**0.00004118**
6,31615535,G,A **AIF:** c.40G>A	5, 8, 9 24, 25	Heterozygous	*AIF1*	p.G14R	**Deleterious S=-5.05**	Tolerated S=0.056	Low Impact S=0.056	**Probably Damaging S=0.945**	Polymorphism	rs2736182 ClinVar: NA	–	0.0554
10,45982847,G,A NOVEL **Tim23: c.261G>A**	25	Heterozygous	*TIMM23* (NOTE: transcript absent in version 38)	Synonymous	Neutral S=0	Tolerated S=0	-	-	**Disease causing**	-	Acceptor site 10373-10413 Score diminished from 0.52 to 0.42	-
13,23886416,G,A **MIP: c.280C>T**	22	Heterozygous	*MIPEP*	p.R94C	**Deleterious S=-3.49**	**Damaging S=0.002**	**Medium Impact S=2.175**	**Possibly Damaging S=0.932**	**Disease causing**	rs147048880 ClinVar: NA	-	0.0004367
14,58396331,delA **Tom20L: c.170delA**	22	Heterozygous	*TOMM20L*	p.Q57Rfs*46	**Frameshift**	-	-	-	**Disease causing**	rs752079244 ClinVar: NA	-	**0.0002762**
19,7928074,T,C **Tim44: c.1128+3A>G**	1, 10-12, 14, 16, 25, 29	Heterozygous	*TIMM44*	–	–	–	–	–	Polymorphism	rs34425383 ClinVar: Benign	**Donor site loss 1345-1359 S=0.41**	0.0712

S – score; NA – not available; MAF – minor allele frequency; L – leucine; P – proline; G – glycine; R – arginine; C – cysteine; Q – glutamine;

Table 8. Variants present in *MIPEP* and *TOMM20L* identified in sample 22 using the ExomeLoupe® software.

Chr	Start	End	R	A	dbSNP (rs ID)	Impact	Impact Severity	Gene	aa change	Polyphen-2 Pred.	Polyphen-2 S	SIFT Pred.	SIFT S	Genotype
13	24460554	24460555	G	A	rs147048880	Missense	MEDIUM	*MIPEP*	R/C	Possibly Damaging	0.462	deleterious	0	G/A
14	58863047	58863049	CA	C	rs752079244	Frameshift	HIGH	*TOMM20L*	Q/X	-	None	-	none	CA/C
14	58863049	58863055	GGGCAC	G	rs760772539	Frameshift	HIGH	*TOMM20L*	GT/X	-	none	-	none	GGGCAC/G

Chr, chromosome; R, reference nucleotide; A, altered nucleotide; aa, amino acid residue; Pred, Prediction; S, Score; R, arginine; C, cysteine; G, glycine; T, Threonine

4.4. Validation of the presence of the *MIPEP* and *TOMM20L* variants

The validation of the 170delA and c.280G>A variants from *TOMM20L* and *MIPEP* gene was performed by two additional methods, namely PCR-RFLP and Sanger sequencing. The observation of the bands' patterns resulted from the digestion of *TOMM20L* and *MIPEP* fragments using BsrI (figure 9A) and AvaII (figure 9B), respectively, confirmed the presence of the variants in the patient (P). The absence of the variants were verified in 200 controls, being the controls 1, 2 and 3, representative of the results.

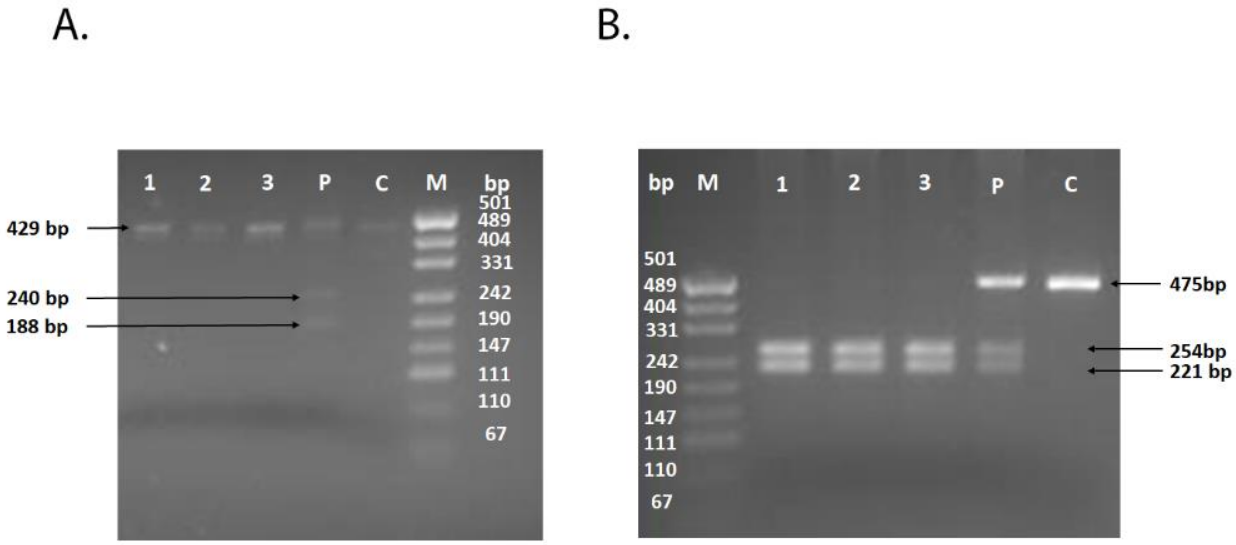

Figure 9. Bands' patterns from RFLP. A - Restriction digestion of *TOMM20L* fragment containing the c.280G>A variant using BsrI. B - Restriction digestion of *MIPEP* fragment containing the c.280G>A variant using AvaII. M, pUC19 DNA/MspI marker, 23; 1, 2, 3 – Controls from healthy individuals; P – Patient sample; C – Non- digested control

The presence of the variants in the LHON patient was also demonstrated in the electropherograms, which resulted from the Sanger sequencing (Figure 10). The heterozygous character of the variants was also clear since the wild-type and the deleted or altered nucleotide are both present. The comparison of the *TOMM20L* forward sequences from the control (figure10B.b) and patient (figure10B.a) led to the identification of the c.170delA variant, while the reverse sequences (figure10B.c, d) led to the identification of the c.172_176delGGCAC.

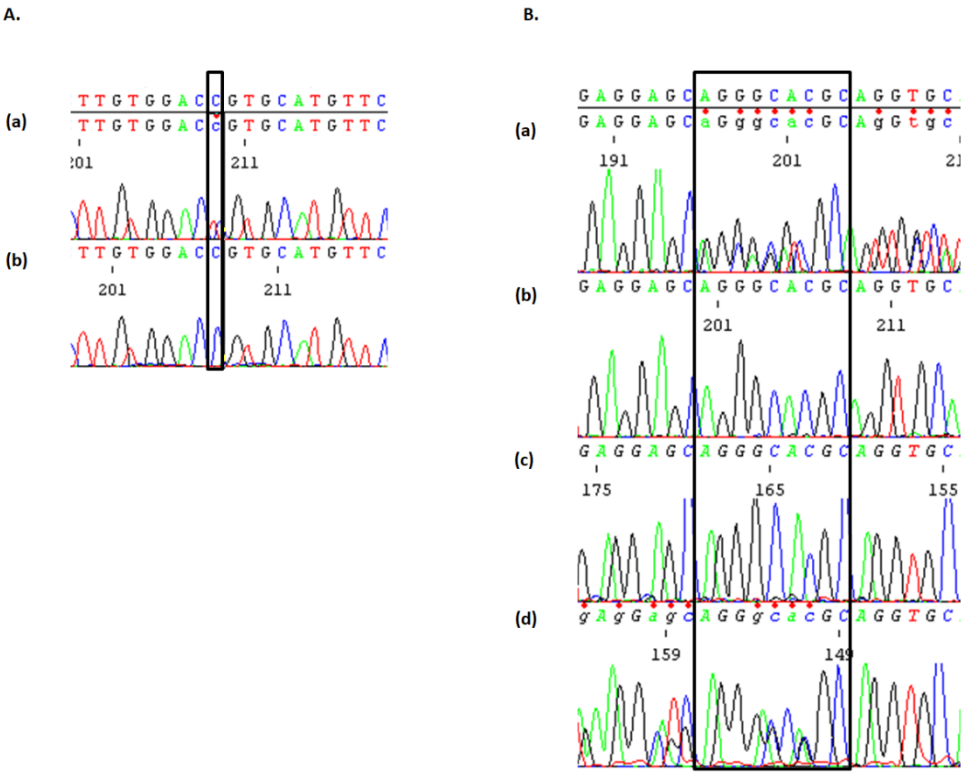

Figure 10. Electropherograms comparing the fragments sequences in a control individual and the LHON patient. A – Sequence from the *MIPEP* fragment containing the c.280C>T variant, (a) patient, (b) control individual; **B** – Sequence from the *TOMM20L* fragment containing the c.170delA and c.172_176delGGCAC, (a) patient forward sequence, (b) control individual forward sequence, (c) control individual reverse sequence, (d) patient reverse sequence.

4.5. Detection and analysis of the proteins affected by the variants

The analysis of the impact of the amino acid modifications caused by the deletions and single nucleotide variant in the presence and quantity of the MIP and Tom20L proteins was archived by using immunodetection (Figure 11). The tissue specificity of the expression of Tom20L protein was also assessed using mouse tissues (Figure 11B). The positive control for anti-Tom20L used was the cell lysate of WiDr cell line. The Tom20L protein was absent in control and patient's lymphocytes, as well as in control fibroblasts demonstrating that the protein is not expressed in these cells or the quantity of protein is vestigial being undetected by Western Blot. Additionally, the Tom20L protein was absent in mouse cerebral cortex and present in mouse testis. The level of MIP protein was highly reduced (80%) in the patient's lymphocytes in comparison to the control, while the level of cytochrome c oxidase subunit 4 (COX4) was normal.

The statistical significance was not assessed, since the blot was only performed once.

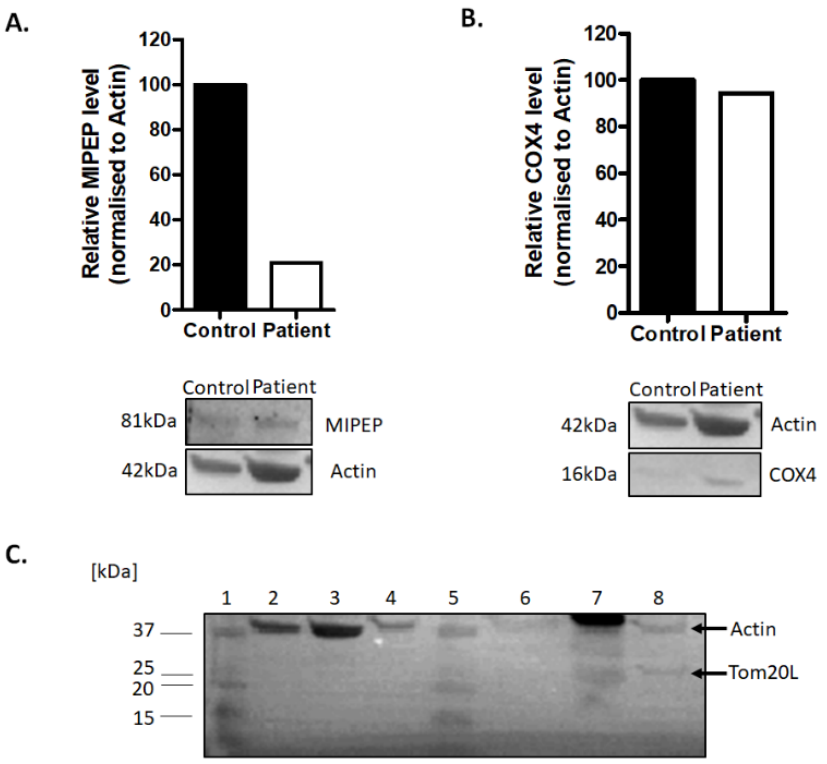

Figure 11. Assessment of the presence and/or quantity of MIP, COX4 and Tom20L proteins from human and mouse cells. Resulting blot (n=1) after immunodetection of **(A)** MIP protein and **(B)** COX4 protein from control (C) and patient (P) lymphocytes and **(C)** Tom20L protein (predicted size: 18kDa) from various human and mouse cells. The WiDr cell lysate (8) was used as positive control for anti-Tom20L. 1, 5 - Precision Plus Protein™ Dual Xtra Standards; 2 – human control lymphocytes; 3 – Patient lymphocytes; 4 – human control fibroblasts; 6 – mouse cortex; 7 – mouse testis; 8 – WiDr cells. MIP protein levels were normalized to actin.

5. DISCUSSION

Variants in genes encoding the mitochondrial protein import machineries and post import processing have been reported to have a role in mitochondrial disorders pathology[27,48,108]. *MIPEP* compound heterozygous and homozygous variants have been reported in individuals with combined oxidative phosphorylation deficiency-31 being the first report of disease caused by variants in this gene[116]. Variants in other genes from mitochondrial protein import, namely *TIMM50* (TIM23 complex), *TIMM8A* (SAM complex) and *AGK* (TIM22 complex) have been reported in mitochondrial epileptic encephalopathy, 3-methylglutaconic aciduria and variable complex V deficiency, Mohr-Tranebjaerg disease and Sengers syndrome, respectively[186–188].

The next-generation sequencing techniques, which includes the WES, had revolutionized the sequencing and analysis of the human genome, allowing the identification of genetic variants in whole genome. Consequently, it has allowed the finding of more genetic causes associated with disease[189,190]. In particular, WES is the preferred approach for the identification of rare and *de novo* variants with a possible functional impact, since the majority of the causative variants are present in coding regions and exon-intron boundaries[191,192]. This approach already allowed the identification of genetic factors with a role in LHON pathogenicity[142]. These sequencing techniques produce an enormous quantity of data that must be processed using bioinformatics tools, so the relevant information must be further analyzed.

In this study, the sequencing and the initial data filtering was performed by an external service provided by ®STABVIDA. This filtering resulted in a list of the total variants localized in whole exome, where the positions of the variants within the chromosomic regions were presented, but not the gene, leading to the need of the identification of the variants of interest using the chromosomic regions. This type of identification is time-consuming, which makes the process of identification of relevant genetic variants difficult. The use of a software as the ExomeLoupe®, facilitates this process in a great extent, since it produces a huge amount of information about the variants, such as the name of the genes containing the variants identified in the analyzed sample, by applying a filter containing the genes' names. Additionally, the software presents the type of the variants and the *in silico* analyze results, allowing the rapid identification of the promising variants. In contrast, in the manual identification, this analysis has to be made for each variant, separately. The software also presents the rs ID associated to each variant allowing the access to important information, as clinical significance and MAF. In addition, the use of the ExoumeLoupe® allowed the identification of other relevant variants, namely the multiple deletion present in *TOMM20L* gene that was not identified by manual search of the variants of interest. This difference is caused by the application of the genomic localization filter that excluded the multiple nucleotide deletions and insertions. For this type of

variants, the chromosomic region was presented with the position of the first nucleotide altered and the position of last nucleotide separated by two dots. When the interval of the first and last position of the gene in the chromosomic region was used as a filter, the latter variants were excluded. The search for this type of variants was also performed to ensure that other relevant variants were not missed. The *in silico* analysis performed manually or using the software allowed the exclusion of variants without functional impact reducing the number of variants, which would require further studies for the evaluation of their pathogenic effect[189].

Considering the obtained results using the manual and software identification of relevant variants, three promising variants were further studied, namely c.172_176delGGCAC and c.170delA variants from *TOMM20L* gene and c.280C>T from *MIPEP* gene. The deletions are present in heterozygosity and they can be present in the same allele or in different alleles. The possible way to know in which allele the variants are present would be the study of the patient's parents or the offspring (if present). If each parent/offspring had one of the variants, the alterations would be in different alleles, whereas variants would be in the same allele if one of the parents/offspring had both variants, since the progeny only inherit one allele from each parent. In the first case, the functional impact would be higher, since each variant in separate leads to a truncated protein and all the synthesized protein amount would be truncated, which likely lead to lack of functional protein. In the second case, the protein would have an amino acid residue exchange and the elimination of the following two amino acids, glycine and threonine. Then, the remaining sequence of the protein would be maintained, possibly leading to a slightly different protein. The functional impact is unknown. The protein structure is often unaffected by this type of alteration (in frame deletions)[193,194]. Indeed, a study reported several individuals with in frame deletions in the dystrophin gene without or with reduced phenotypical manifestation[195]. However, a single amino acid residue deletion can cause disease, as in cystic fibrosis[196]. Another consideration should be taken into account, related to the presence of *de novo* variants in the patient. The family studies would allow the clarification of this point, but the samples of family members were not available for analysis.

Even though the WES has been allowing the identification of disease-causing variants, the identified alterations must be confirmed by a second method because of the WES limitations, such as the low coverage of GC-regions, a reduced quality of read extremities, difficulties in detection of variants in repetitive regions and sequencing errors that could lead to a false variant identification[190,197–200]. Since this information is used for the diagnosis of genetic disorders, the presence of the variant should be ensured. Here, the PCR-RFLP and Sanger sequencing were performed since both methods can be used for the validation of the presence of variants[201,202]. Both methods confirmed the presence of the promising variants identified

using the WES analysis. Concerning the confirmation of deletions affecting the *TOMM20L* gene, a consideration related to the restriction enzyme AvaII must be taken into consideration. The presence of c.170delA variant or the combination of c.170delA and c.172_176delGGCAC variants would lead to the same restriction digestion result. Thus, the band pattern would be similar, with the exception of the second fragment that would be five nucleotides shorter, resulting from the deletion of these nucleotides. This small difference would not be perceived in the agarose gel, therefore the presence of the multiple deletion would remain unnoticed. This evidence reveals the lack of specificity of restriction enzymes. It means that different variants can create fragments that result in the same target sequence for restriction enzymes, originating in the same bands' patterns and that not be distinguished. The combination of PCR-RFLP and Sanger sequencing would be a good way to overcome the latter limitation since it would allow the identification of the distinct variants.

The impact of variants in the 3D structure of the proteins could be evaluated using the PyMol® Molecular Graphics Sytem, Version 2.0 Schroödinger, LLC. This software allows the visualization of the 3D structure of proteins when the structure experimentally obtained/structural model from the pretended protein or a protein presenting a high sequence identity are available in protein databases such Protein Data Bank (PDB) and Model Portal[203–205]. The alterations of interest within the protein structure could be analyzed and the altered and wild type proteins could be compared. This comparison would allow the observation of structural and conformational modifications and their possible implications on the protein function[206]. Since the 3D structure of Tom20L and MIP, or similar proteins, are not available this comparison could not be performed.

After the genetic study and confirmation of the promising variants, the molecular study of the altered proteins was performed for the evaluation of the impact of the variants on the expression and quantity of the Tom20L and MIP.

It was possible to observe that the Tom20L protein is absent in the control lymphocytes/fibroblasts and the patient lymphocytes. This absence can be explained by a tissue differential expression. Studies for the mapping of the total human proteome using RNA sequencing demonstrated a testis expression specificity of Tom20L with a mild expression in the cerebral cortex[55,56]. This expression pattern was also reported in mouse and *D. melanogaster*[22]. Even though the specificity of expression could not be evaluated using patient's cells, cortex and testis from mouse were used. Undeniably, the difficult access to samples from patients is one of the limitations of studies with human tissues. Another limitation is the high frequency of variant detection in one individual or one family with LHON. Furthermore, the study of the family is often difficult and the possibility of finding another patient with the same variations is small.

The specificity was assessed using Western Blot allowing the complementation of the evidences from transcript sequencing[22,55,56]. Indeed, the Tom20L protein was expressed in mouse testis corroborating the specificity of expression reported in expression databases, specifically in the Expression Atlas (https://www.ebi.ac.uk/gxa/about.html) and Human Protein Atlas project (https://www.proteinatlas.org/). Even though the expression of Tom20L protein was absent in the cerebral cortex, the actin band (control protein) was considerably weak, suggesting that more extract should be added to the gel for the detection of the protein. If the protein would be absent, even with an increase of the protein added, the Tom20L mRNA detected from transcript quantification may not be translated and the protein would be absent. Since the Tom20L was only expressed in testis, specifically in primary spermatocytes and early spermatids, it could have a particularly important role in this tissue[54]. Although the role remain uncertain, some suggestions have been made. The Tom20L could reinforce the recognition of precursors made by Tom20 and, consequently reinforce the protein import to accomplish the extensive mitochondrial biogenesis or it could be incorporated into TOM complex to favor the import of specific proteins during different stages of spermatogenesis[22,54].

Concerning the Tom20L protein, functional validation of the variants was not possible due to the specificity of tissue expression of this protein. As previously discussed, the patient cells available for the study (lymphocytes) do not express the Tom20L protein and its presence is necessary for the validation.

The MIP protein from the patient's lymphocytes was clearly reduced in comparison to the control lymphocytes and consequently, the processing of substrates may be diminished leading to an increase of unprocessed and unstable proteins. This protease cleave the octapeptide from proteins belonging to the MRC system, specifically the COX4[118]. As well as the other MRC complexes, the complex IV is composed by various subunits, which associate to each other forming various subassemblies leading to the formation of the complete holoenzyme[207]. The processing of mitochondrial proteins encoded by nDNA is essential for their assembly and function[109,118]. The initial complex subassembly seems to include the COX4 being essential for the association of the other subunits and, consequently for the complete assemble of the holoenzyme[208]. When the octapeptide is not cleaved, only a small quantity of the mature protein may be available. Accordingly, the number of assembled complex IV units may also be reduced, leading to depletion of activity. Indeed, the evaluation of the patient MRC activity for diagnosis purposes was previously made, revealing a 50% decrease of complex IV activity. The immunodetection of this protein was also performed, in order to determine levels from the COX4. Although the levels of COX4 in the patient's lymphocytes were normal relatively to the control, the protein could be not functional. For the evaluation of the assembly of complex IV

and consequent clarification of which type of COX4 (intermediate or mature) is present in the patient's lymphocytes, the Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE) may be performed. Furthermore, this technique would allow the *in vitro* functional validation, which is one of the strongest type of study for validation of variant pathogenicity[209,210]. The functional studies are also important to avoid false pathogenicity prediction by the computational algorithms used on the *in silico* analysis tools[211,212].The BN-PAGE would evaluate the assembly of the complex IV and consequently, reveal the presence of defects in the MRC, since the functions of the complexes are interdependent [213]. Evidences suggest that the assembled complex IV helps to maintain the activity of complex I[214]. Since defects in complex I are hallmarks of LHON disease, the reduction of assembled complex IV may aggravate the defect caused by the mtDNA mutations.

6. CONCLUSIONS AND FUTURE PERSPECTIVES

This study allowed the identification of additional gene sequence variants in a LHON patient, namely alterations in nuclear genes encoding the Tom20L and MIP proteins, which are involved in the mitochondrial protein import and processing of precursor proteins imported into the mitochondrial matrix, respectively, in a LHON patient.

The influence of the gene sequence variants present in Tom20L could not be elucidated due to specificity of the protein tissue expression, but considering the possible roles of this protein, one can assume that these variants would have a negative impact on the spermatogenesis.

The genetic variant present in the *MIPEP* gene led to a reduction of levels of MIP protein in patient's lymphocytes. This decrease may compromise the processing of COX4 and consequently, the assembly of complex IV, which is compatible with the activity reduction observed in the patient cells. Additionally, the reduction of assembled complex IV may worsen the defect caused by the m.14448T>C present in the LHON patient, since the complex IV may be involved in the stabilization and activity of the complex I. This possible effect must be clarified though further functional studies, namely BN-PAGE using the patient's lymphocytes to evaluate if the complex VI is properly assembled or if it is compromised by the possible reduction of the COX4 processing.

Although this study is still preliminary, it reinforces the role of the nuclear gene factors in mitochondrial disorders specially, LHON. The nuclear gene variants reported in LHON affect proteins involved in many mitochondrial functions, specifically oxidative metabolism, apoptosis, mitochondrial translation and fusion[137,139,140,142]. The identification of promising variants in genes encoding the mitochondrial protein import machineries and post import processing in a LHON individual was first accomplished in this study, representing a significant contribution for the field and opening new perspectives in LHON research.

7. REFERENCES

1. Kühlbrandt, W. Structure and function of mitochondrial membrane protein complexes. *BMC Biol.* **13,** 89 (2015). doi: 10.1186/s12915-015-0201-x

2. McCarron, J. G. *et al.* From Structure to Function: Mitochondrial Morphology, Motion and Shaping in Vascular Smooth Muscle. *J. Vasc. Res.* **50,** 357–371 (2013). doi: 10.1159/000353883

3. Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T. and Pfanner, N. Importing Mitochondrial Proteins: Machineries and Mechanisms. *Cell* **138,** 628–644 (2009). doi: 10.1016/j.cell.2009.08.005

4. Schatz, G. The protein import system of mitochondria. *Journal of Biological Chemistry* (1996). doi:10.1074/jbc.271.50.31763

5. Dudek, J., Rehling, P. and van der Laan, M. Mitochondrial protein import: Common principles and physiological networks. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833,** 274–285 (2013). doi: 10.1016/j.bbamcr.2012.05.028

6. Kellems, R. E., Allison, V. F. and Butow, R. A. Cytoplasmic type 80s ribosomes associated with yeast mitochondria: IV. Attachment of ribosomes to the outer: Membrane of isolated mitochondria. *J. Cell Biol.* (1975). doi:10.1083/jcb.65.1.1

7. Verner, K. Co-translational protein import into mitochondria: an alternative view. *Trends Biochem. Sci.* (1993). doi:10.1016/0968-0004(93)90090-A

8. Margeot, A. *et al.* Why are many mRNAs translated to the vicinity of mitochondria: A role in protein complex assembly? in *Gene* (2005). doi:10.1016/j.gene.2005.04.022

9. MacKenzie, J. A. and Payne, R. M. Ribosomes Specifically Bind to Mammalian Mitochondria via Protease-sensitive Proteins on the Outer Membrane. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M307167200

10. Yogev, O., Karniely, S. and Pines, O. Translation-coupled translocation of yeast fumarase into mitochondria in vivo. *J. Biol. Chem.* (2007). doi:10.1074/jbc.M704201200

11. Melin, J. *et al.* Presequence Recognition by the Tom40 Channel Contributes to Precursor Translocation into the Mitochondrial Matrix. *Mol. Cell. Biol.* **34,** 3473–3485 (2014). doi: 10.1128/MCB.00433-14

12. Neupert, W. and Herrmann, J. M. Translocation of Proteins into Mitochondria. *Annu. Rev. Biochem.* **76,** 723–749 (2007). doi: 10.1146/annurev.biochem.76.052705.163409

13. Wenz, L.-S., Opaliński, Ł., Wiedemann, N. and Becker, T. Cooperation of protein machineries in mitochondrial protein sorting. *Biochim. Biophys. Acta - Mol. Cell Res.* **1853,** 1119–1129 (2015). doi: 10.1016/j.bbamcr.2015.01.012

14. Nelson, D. R., Felix, C. M. and Swanson, J. M. Highly conserved charge-pair networks in the mitochondrial carrier family. *J. Mol. Biol.* (1998). doi:10.1006/jmbi.1997.1594

15. Callegari, S. *et al.* TIM29 is a subunit of the human carrier translocase required for protein transport. *FEBS Lett.* **590,** 4147–4158 (2016). doi:10.1002/1873-3468.12450

16. Straub, S. P., Stiller, S. B., Wiedemann, N. and Pfanner, N. Dynamic organization of the

mitochondrial protein import machinery. *Biol. Chem.* **397,** 1097–1114 (2016). doi: 10.1515/hsz-2016-0145

17. Xie, J., Marusich, M. F., Souda, P., Whitelegge, J. and Capaldi, R. A. The mitochondrial inner membrane protein Mitofilin exists as a complex with SAM50, metaxins 1 and 2, coiled-coil-helix coiled-coil-helix domain-containing protein 3 and 6 and DnaJC11. *FEBS Lett.* (2007). doi:10.1016/j.febslet.2007.06.052

18. Huynen, M. A., Mühlmeister, M., Gotthardt, K., Guerrero-Castillo, S. and Brandt, U. Evolution and structural organization of the mitochondrial contact site (MICOS) complex and the mitochondrial intermembrane space bridging (MIB) complex. *Biochim. Biophys. Acta - Mol. Cell Res.* (2016). doi:10.1016/j.bbamcr.2015.10.009

19. Kozjak-Pavlovic, V. The MICOS complex of human mitochondria. *Cell and Tissue Research* (2017). doi:10.1007/s00441-016-2433-7

20. Johnston, A. J. *et al.* Insertion and Assembly of Human Tom7 into the Preprotein Translocase Complex of the Outer Mitochondrial Membrane. *J. Biol. Chem.* **277,** 42197–42204 (2002). doi: 10.1074/jbc.M205613200

21. Yano, M. *et al.* Functional Analysis of Human Mitochondrial Receptor Tom20 for Protein Import into Mitochondria. *J. Biol. Chem.* **273,** 26844–26851 (1998). doi: 10.1074/jbc.273.41.26844

22. Likić, V. A. *et al.* Patterns that Define the Four Domains Conserved in Known and Novel Isoforms of the Protein Import Receptor Tom20. *J. Mol. Biol.* **347,** 81–93 (2005). doi: 10.1016/j.jmb.2004.12.057

23. Yano, M., Hoogenraad, N., Terada, K. and Mori, M. Identification and Functional Analysis of Human Tom22 for Protein Import into Mitochondria. *Mol. Cell. Biol.* **20,** 7205–7213 (2000). doi: 10.1128/MCB.20.19.7205-7213.2000

24. Brix, J., Dietmeier, K. and Pfanner, N. Differential Recognition of Preproteins by the Purified Cytosolic Domains of the Mitochondrial Import Receptors Tom20, Tom22, and Tom70. *J. Biol. Chem.* **272,** 20730–20735 (1997). doi: 10.1074/jbc.272.33.20730

25. Schulz, C., Schendzielorz, A. and Rehling, P. Unlocking the presequence import pathway. *Trends Cell Biol.* **25,** 265–275 (2015). doi: 10.1016/j.tcb.2014.12.001

26. Kato, H. and Mihara, K. Identification of Tom5 and Tom6 in the preprotein translocase complex of human mitochondrial outer membrane. *Biochem. Biophys. Res. Commun.* **369,** 958–963 (2008). doi: 10.1016/j.bbrc.2008.02.150

27. Kang, Y., Fielden, L. F. and Stojanovski, D. Mitochondrial protein transport in health and disease. *Semin. Cell Dev. Biol.* (2017). doi:10.1016/j.semcdb.2017.07.028

28. Demishtein-Zohary, K. and Azem, A. The TIM23 mitochondrial protein import complex: function and dysfunction. *Cell Tissue Res.* **367,** 33–41 (2017). doi: 10.1007/s00441-016-2486-7

29. Guo, Y. *et al.* Tim50, a Component of the Mitochondrial Translocator, Regulates Mitochondrial Integrity and Cell Death. *J. Biol. Chem.* **279,** 24813–24825 (2004). doi: 10.1074/jbc.M402049200

30. Bauer, M. F. *et al.* Genetic and structural characterization of the human mitochondrial

inner membrane translocase 1 1Edited by J. Karn. *J. Mol. Biol.* **289,** 69–82 (1999). doi: 10.1006/jmbi.1999.2751

31. Dores-Silva, P. R., Barbosa, L. R. S., Ramos, C. H. I. and Borges, J. C. Human Mitochondrial Hsp70 (Mortalin): Shedding Light on ATPase Activity, Interaction with Adenosine Nucleotides, Solution Structure and Domain Organization. *PLoS One* **10,** e0117170 (2015). doi: 10.1371/journal.pone.0117170

32. Ting, S. Y., Yan, N. L., Schilke, B. A. and Craig, E. A. Dual interaction of scaffold protein Tim44 of mitochondrial import motor with channel-forming translocase subunit Tim23. *Elife* (2017). doi:10.7554/eLife.23609

33. Elsner, S., Simian, D., Iosefson, O., Marom, M. and Azem, A. The Mitochondrial Protein Translocation Motor: Structural Conservation between the Human and Yeast Tim14/Pam18-Tim16/Pam16 co-Chaperones. *Int. J. Mol. Sci.* **10,** 2041–2053 (2009). doi: 10.3390/ijms10052041

34. Mokranjac, D. and Neupert, W. The many faces of the mitochondrial TIM23 complex. *Biochim. Biophys. Acta - Bioenerg.* **1797,** 1045–1054 (2010). doi: 10.1016/j.bbabio.2010.01.026

35. Srivastava, S. *et al.* Regulation of mitochondrial protein import by the nucleotide exchange factors GrpEL1 and GrpEL2 in human cells. *J. Biol. Chem.* jbc.M117.788463 (2017). doi:10.1074/jbc.M117.788463

36. Kovermann, P. *et al.* Tim22, the essential core of the mitochondrial protein insertion complex, forms a voltage-activated and signal-gated channel. *Mol. Cell* **9,** 363–373 (2002). doi: 10.1016/S1097-2765(02)00446-X

37. Kang, Y. *et al.* Tim29 is a novel subunit of the human TIM22 translocase and is involved in complex assembly and stability. *Elife* **5,** (2016). doi: 10.7554/eLife.17463

38. Kerscher, O., Sepuri, N. B. and Jensen, R. E. Tim18p is a new component of the Tim54p-Tim22p translocon in the mitochondrial inner membrane. *Mol. Biol. Cell* (2000). doi:10.1091/mbc.11.1.103

39. Mühlenbein, N., Hofmann, S., Rothbauer, U. and Bauer, M. F. Organization and Function of the Small Tim Complexes Acting along the Import Pathway of Metabolite Carriers into Mammalian Mitochondria. *J. Biol. Chem.* **279,** 13540–13546 (2004). doi: 10.1074/jbc.M312485200

40. Chacinska, A. *et al.* Essential role of Mia40 in import and assembly of mitochondrial intermembrane space proteins. *EMBO J.* (2004). doi:10.1038/sj.emboj.7600389

41. Rissler, M. *et al.* The essential mitochondrial protein Erv1 cooperates with Mia40 in biogenesis of intermembrane space proteins. *J. Mol. Biol.* (2005). doi:10.1016/j.jmb.2005.08.051

42. Sztolsztener, M. E., Brewinska, A., Guiard, B. and Chacinska, A. Disulfide Bond Formation: Sulfhydryl Oxidase ALR Controls Mitochondrial Biogenesis of Human MIA40. *Traffic* **14,** 309–320 (2013). doi: 10.1111/tra.12030

43. Mesecke, N. *et al.* The zinc-binding protein Hot13 promotes oxidation of the mitochondrial import receptor Mia40. *EMBO Rep.* (2008). doi:10.1038/embor.2008.173

44. Kozjak, V. *et al.* An Essential Role of Sam50 in the Protein Sorting and Assembly Machinery of the Mitochondrial Outer Membrane. *J. Biol. Chem.* **278,** 48520–48523 (2003). doi: 10.1074/jbc.C300442200

45. Kutik, S. *et al.* Dissecting Membrane Insertion of Mitochondrial β-Barrel Proteins. *Cell* **132,** 1011–1024 (2008). doi: 10.1016/j.cell.2008.01.028

46. Bohnert, M., Pfanner, N. and van der Laan, M. A dynamic machinery for import of mitochondrial precursor proteins. *FEBS Lett.* **581,** 2802–2810 (2007). doi: 10.1016/j.febslet.2007.03.004

47. Okamoto, H., Miyagawa, A., Shiota, T., Tamura, Y. and Endo, T. Intramolecular disulfide bond of TIM22 protein maintains integrity of the TIM22 complex in the mitochondrial inner membrane. *J. Biol. Chem.* **289,** 4827–4838 (2014). doi: 10.1016/j.febslet.2007.03.004

48. Sokol, A. M., Sztolsztener, M. E., Wasilewski, M., Heinz, E. and Chacinska, A. Mitochondrial protein translocases for survival and wellbeing. *FEBS Lett.* **588,** 2484–2495 (2014). doi: 10.1016/j.febslet.2014.05.028

49. Edmonson, A. M., Mayfield, D. K., Vervoort, V., DuPont, B. R. and Argyropoulos, G. Characterization of a human import component of the mitochondrial outer membrane, TOMM70A. *Cell Commun. Adhes.* **9,** 15–27 (2002). doi: 10.1080/10615380290011265

50. Schleiff, E., Shore, G. C. and Goping, I. S. Interactions of the human mitochondrial protein import receptor, hTom20, with precursor proteins in vitro reveal pleiotropic specificities and different receptor domain requirements. *J. Biol. Chem.* (1997). doi:10.1074/jbc.272.28.17784

51. Abe, Y. *et al.* Structural Basis of Presequence Recognition by the Mitochondrial Protein Import Receptor Tom20. *Cell* **100,** 551–560 (2000). doi: 10.1016/S0092-8674(00)80691-1

52. Saitoh, T. *et al.* Tom20 recognizes mitochondrial presequences through dynamic equilibrium among multiple bound states. *EMBO J.* **26,** 4777–4787 (2007). doi: 10.1038/sj.emboj.7601888

53. Yano, M., Terada, K. and Mori, M. AIP is a mitochondrial import mediator that binds to both import receptor Tom20 and preproteins. *J. Cell Biol.* **163,** 45–56 (2003). doi: 10.1083/jcb.200305051

54. Hwa, J. J. *et al.* Germ-line specific variants of components of the mitochondrial outer membrane import machinery in Drosophila. *FEBS Lett.* **572,** 141–146 (2004). doi:10.1016/j.febslet.2004.07.025

55. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015). doi:10.1016/j.febslet.2007.03.004

56. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science (80-.).* **347,** 1260419–1260419 (2015). doi: 10.1126/science.1260419

57. Mukhopadhyay, A., Avramova, L. V and Weiner, H. Tom34 unlike Tom20 does not interact with the leader sequences of mitochondrial precursor proteins. *Arch. Biochem. Biophys.* (2002). doi:10.1006/abbi.2002.2777

58. Terada, K. *et al.* Expression of Tom34 splicing isoforms in mouse testis and knock-out of Tom34 in mice. *J. Biochem.* **133,** 625–631 (2003). doi: 10.1093/jb/mvg080

59. Humphries, A. D. *et al.* Dissection of the Mitochondrial Import and Assembly Pathway for Human Tom40. *J. Biol. Chem.* **280,** 11535–11543 (2005). doi: 10.1074/jbc.M413816200

60. Gessmann, D. *et al.* Structural elements of the mitochondrial preprotein-conducting channel Tom40 dissolved by bioinformatics and mass spectrometry. *Biochim. Biophys. Acta - Bioenerg.* **1807,** 1647–1657 (2011). doi: 10.1016/j.bbabio.2011.08.006

61. Blesa, J. R., Prieto-Ruiz, J. A., Hernández, J. M. and Hernández-Yago, J. NRF-2 transcription factor is required for human TOMM20 gene expression. *Gene* (2007). doi:10.1016/j.gene.2006.12.024

62. Blesa, J. R. and Hernández-Yago, J. Distinct functional contributions of 2 GABP-NRF-2 recognition sites within the context of the human TOMM70 promoter. *Biochem. Cell Biol.* (2006). doi:10.1139/o06-064

63. Rao, S. *et al.* Biogenesis of the preprotein translocase of the outer mitochondrial membrane: protein kinase A phosphorylates the precursor of Tom40 and impairs its import. *Mol. Biol. Cell* **23,** 1618–1627 (2012). doi: 10.1091/mbc.e11-11-0933

64. Mick, D. U. *et al.* MITRAC links mitochondrial protein translocation to respiratory-chain assembly and translational regulation. *Cell* **151,** 1528–1541 (2012). doi: 10.1016/j.cell.2012.11.053

65. Bauer, M. F., Sirrenberg, C., Neupert, W. and Brunner, M. Role of Tim23 as voltage sensor and presequence receptor in protein import into mitochondria. *Cell* (1996). doi:10.1016/S0092-8674(00)81320-3

66. Alder, N. N., Sutherland, J., Buhring, A. I., Jensen, R. E. and Johnson, A. E. Quaternary structure of the mitochondrial TIM23 complex reveals dynamic association between Tim23p and other subunits. *Mol. Biol. Cell* (2008). doi:10.1091/mbc.E07-07-0669

67. Waegemann, K., Popov-Čeleketić, D., Neupert, W., Azem, A. and Mokranjac, D. Cooperation of TOM and TIM23 complexes during translocation of proteins into mitochondria. *J. Mol. Biol.* **427,** 1075–1084 (2015). doi: 10.1016/j.jmb.2014.07.015

68. Prieto-Ruiz, J. A. *et al.* Expression of the human TIMM23 and TIMM23B genes is regulated by the GABP transcription factor. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1861,** 80–94 (2018). doi: 10.1016/j.bbagrm.2018.01.006

69. Sinha, D., Srivastava, S., Krishna, L. and D'Silva, P. Unraveling the Intricate Organization of Mammalian Mitochondrial Presequence Translocases: Existence of Multiple Translocases for Maintenance of Mitochondrial Function. *Mol. Cell. Biol.* **34,** 1757–1775 (2014). doi: 10.1128/MCB.01527-13

70. Schleyer, M., Schmidt, B. and Neupert, W. Requirement of a Membrane Potential for the Posttranslational Transfer of Proteins into Mitochondsria. *Eur. J. Biochem.* (1982). doi:10.1111/j.1432-1033.1982.tb06657.x

71. Martin, J., Mahlke, K. and Pfanner, N. Role of an energized inner membrane in mitochondrial protein import: ΔΨ drives the movement of presequences. *J. Biol. Chem.* **266,** 18051–18057 (1991).

72. van der Laan, M. *et al.* Motor-free mitochondrial presequence translocase drives membrane integration of preproteins. *Nat. Cell Biol.* (2007). doi:10.1038/ncb1635

73. Borges, J. C., Fischer, H., Craievich, A. F., Hansen, L. D. and Ramos, C. H. I. Free human mitochondrial GrpE is a symmetric dimer in solution. *J. Biol. Chem.* **278,** 35337–35344 (2003). doi: 10.1074/jbc.M305083200

74. Sinha, D., Joshi, N., Chittoor, B., Samji, P. and D'Silva, P. Role of Magmas in protein transport and human mitochondria biogenesis. *Hum. Mol. Genet.* (2010). doi:10.1093/hmg/ddq002

75. Schusdziarra, C., Blamowska, M., Azem, A. and Hell, K. Methylation-controlled J-protein MCJ acts in the import of proteins into human mitochondria. *Hum. Mol. Genet.* (2013). doi:10.1093/hmg/dds541

76. Sinha, D., Srivastava, S. and D'Silva, P. Functional diversity of human mitochondrial j-proteins is independent of their association with the inner membrane presequence translocase. *J. Biol. Chem.* (2016). doi:10.1074/jbc.M116.738146

77. Dudek, J., Rehling, P. and van der Laan, M. Mitochondrial protein import: Common principles and physiological networks. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833,** 274–285 (2013). 10.1016/j.bbamcr.2012.05.028

78. Chacinska, A. *et al.* Mitochondrial presequence translocase: Switching between TOM tethering and motor recruitment involves Tim21 and Tim17. *Cell* (2005). doi:10.1016/j.cell.2005.01.011

79. Mokranjac, D. *et al.* Role of Tim50 in the Transfer of Precursor Proteins from the Outer to the Inner Membrane of Mitochondria. *Mol. Biol. Cell* (2009). doi:10.1091/mbc.E08-09-0934

80. Tamura, Y. *et al.* Tim23 - Tim50 pair coordinates functions of translocators and motor proteins in mitochondrial protein import. *J. Cell Biol.* (2009). doi:10.1083/jcb.200808068

81. Mokranjac, D., Bourenkov, G., Hell, K., Neupert, W. and Groll, M. Structure and function of Tim14 and Tim16, the J and J-like components of the mitochondrial protein import motor. *EMBO J.* (2006). doi:10.1038/sj.emboj.7601334

82. van der Laan, M. *et al.* A Role for Tim21 in Membrane-Potential-Dependent Preprotein Sorting in Mitochondria. *Curr. Biol.* (2006). doi:10.1016/j.cub.2006.10.025

83. Wiedemann, N., Van Der Laan, M., Hutu, D. P., Rehling, P. and Pfanner, N. Sorting switch of mitochondrial presequence translocase involves coupling of motor module to respiratory chain. *J. Cell Biol.* (2007). doi:10.1083/jcb.200709087

84. Dienhart, M. K. and Stuart, R. A. The Yeast Aac2 Protein Exists in Physical Association with the Cytochrome bc1-COX Supercomplex and the TIM23 Machinery. *Mol. Biol. Cell* (2008). doi:10.1091/mbc.E08-04-0402

85. Okamoto, K. *et al.* The protein import motor of mitochondria: A targeted molecular ratchet driving unfolding and translocation. *EMBO J.* (2002). doi:10.1093/emboj/cdf358

86. Fontanesi, F., Soto, I. C., Horn, D. and Barrientos, A. Mss51 and Ssc1 facilitate translational regulation of cytochrome c oxidase biogenesis. *Mol. Cell. Biol.* (2010). doi:10.1128/MCB.00983-09

87. Bottinger, L. *et al.* A complex of Cox4 and mitochondrial Hsp70 plays an important role in the assembly of the cytochrome c oxidase. *Mol. Biol. Cell* (2013). doi:10.1091/mbc.E13-02-0106

88. Sirrenberg, C., Bauer, M. F., Guiard, B., Neupert, W. and Brunner, M. Import of carrier proteins into the mitochondrial inner membrane mediated by Tim22. *Nature* (1996). doi:10.1038/384582a0

89. Kerscher, O., Holder, J., Srinivasan, M., Leung, R. S. and Jensen, R. E. The Tim54p-Tim22p complex mediates insertion of proteins into the mitochondrial inner membrane. *J. Cell Biol.* (1997). doi:10.1083/jcb.139.7.1663

90. Gebert, N. *et al.* Dual function of Sdh3 in the respiratory Chain and TIM22 protein translocase of the mitochondrial inner membrane. *Mol. Cell* (2011). doi:10.1016/j.molcel.2011.09.025

91. Sirrenberg, C. *et al.* Carrier protein import into mitochondria mediated by the intermembrane proteins Tim10/Mrs11 and Tim12/Mrs5. *Nature* (1998). doi:10.1038/36136

92. Koehler, C. M. *et al.* Import of mitochondrial carriers mediated by essential proteins of the intermembrane space. *Science (80-.).* (1998). doi:10.1126/science.279.5349.369

93. Koehler, C. M. *et al.* Tim9p, an essential partner subunit of Tim10p for the import of mitochondrial carrier proteins. *EMBO J.* (1998). doi:10.1093/emboj/17.22.6477

94. Adam, A. *et al.* Tim9, a new component of the TIM22·54 translocase in mitochondria. *EMBO J.* (1999). doi:10.1093/emboj/18.2.313

95. Mühlenbein, N., Hofmann, S., Rothbauer, U. and Bauer, M. F. Organization and Function of the Small Tim Complexes Acting along the Import Pathway of Metabolite Carriers into Mammalian Mitochondria. *J. Biol. Chem.* **279,** 13540–13546 (2004). doi:10.1074/jbc.M312485200

96. Bauer, M. F. *et al.* The mitochondrial TIM22 preprotein translocase is highly conserved throughout the eukaryotic kingdom. *FEBS Lett.* (1999). doi:10.1016/S0014-5793(99)01665-8

97. Gebert, N. *et al.* Assembly of the three small Tim proteins precedes docking to the mitochondrial carrier translocase. *EMBO Rep.* **9,** 548–554 (2008). doi: 10.1038/embor.2008.49

98. Gabriel, K. *et al.* Novel Mitochondrial Intermembrane Space Proteins as Substrates of the MIA Import Pathway. *J. Mol. Biol.* **365,** 612–620 (2007). doi: 10.1016/j.jmb.2006.10.038

99. Banci, L. *et al.* MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat. Struct. Mol. Biol.* **16,** 198–206 (2009). doi: 10.1038/nsmb.1553

100. Hangen, E. *et al.* Interaction between AIF and CHCHD4 Regulates Respiratory Chain Biogenesis. *Mol. Cell* **58,** 1001–1014 (2015). doi: 10.1016/j.molcel.2015.04.020

101. Herrmann, J. M. and Riemer, J. Mitochondrial disulfide relay: Redox-regulated protein import into the intermembrane space. *J. Biol. Chem.* **287,** 4426–4433 (2012). doi: 10.1074/jbc.R111.270678

102. Backes, S. and Herrmann, J. M. Protein Translocation into the Intermembrane Space and Matrix of Mitochondria: Mechanisms and Driving Forces. *Front. Mol. Biosci.* **4,** (2017). doi: 10.3389/fmolb.2017.00083

103. Stojanovski, D. *et al.* Mitochondrial protein import: precursor oxidation in a ternary complex with disulfide carrier and sulfhydryl oxidase. *J. Cell Biol.* (2008). doi:10.1083/jcb.200804095

104. Ishikawa, D., Yamamoto, H., Tamura, Y., Moritoh, K. and Endo, T. Two novel proteins in the mitochondrial outer membrane mediate β-barrel protein assembly. *J. Cell Biol.* (2004). doi:10.1083/jcb.200405138

105. Gentle, I. E. *et al.* Conserved Motifs Reveal Details of Ancestry and Structure in the Small TIM Chaperones of the Mitochondrial Intermembrane Space. *Mol. Biol. Evol.* **24,** 1149–1160 (2007). doi: 10.1093/molbev/msm031

106. Ding, C. *et al.* Mitofilin and CHCHD6 physically interact with Sam50 to sustain cristae structure. *Sci. Rep.* (2015). doi:10.1038/srep16064

107. Ott, C. *et al.* Sam50 Functions in Mitochondrial Intermembrane Space Bridging and Biogenesis of Respiratory Complexes. *Mol. Cell. Biol.* **32,** 1173–1188 (2012). doi: 10.1128/MCB.06388-11

108. Quirós, P. M., Langer, T. and López-Otín, C. New roles for mitochondrial proteases in health, ageing and disease. *Nat. Rev. Mol. Cell Biol.* **16,** 345–359 (2015). doi: 10.1038/nrm3984

109. Mossmann, D., Meisinger, C. and Vögtle, F. N. Processing of mitochondrial presequences. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819,** 1098–1106 (2012). doi: 10.1016/j.bbagrm.2011.11.007

110. Vögtle, F. N. *et al.* Global Analysis of the Mitochondrial N-Proteome Identifies a Processing Peptidase Critical for Protein Stability. *Cell* **139,** 428–439 (2009). doi: 10.1016/j.cell.2009.07.045

111. Burkhart, J. M., Taskin, A. A., Zahedi, R. P. and Vögtle, F. N. Quantitative Profiling for Substrates of the Mitochondrial Presequence Processing Protease Reveals a Set of Nonsubstrate Proteins Increased upon Proteotoxic Stress. *J. Proteome Res.* **14,** 4550–4563 (2015). doi: 10.1021/acs.jproteome.5b00327

112. Poveda-Huertes, D., Mulica, P. and Vögtle, F. N. The versatility of the mitochondrial presequence processing machinery: cleavage, quality control and turnover. *Cell Tissue Res.* **367,** 73–81 (2017). doi: 10.1007/s00441-016-2492-9

113. Chew, A. *et al.* Cloning, Expression, and Chromosomal Assignment of the Human Mitochondrial Intermediate Peptidase Gene (MIPEP). *Genomics* **40,** 493–496 (1997). doi: 10.1006/geno.1996.4586

114. Chew, A., Sirugo, G., Alsobrook, J. P. and Isaya, G. Functional and Genomic Analysis of the Human Mitochondrial Intermediate Peptidase, a Putative Protein Partner of Frataxin. *Genomics* **65,** 104–112 (2000). doi: 10.1006/geno.2000.6162

115. Hu, Z. *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat. Genet.* (2011). doi:10.1038/ng.875

116. Eldomery, M. K. *et al.* MIPEP recessive variants cause a syndrome of left ventricular non-compaction, hypotonia, and infantile death. *Genome Med.* **8,** 106 (2016). doi: 10.1186/s13073-016-0360-6

117. Vögtle, F. N. *et al.* Mitochondrial protein turnover: role of the precursor intermediate peptidase Oct1 in protein stabilization. *Mol. Biol. Cell* **22,** 2135–2143 (2011). doi: 10.1091/mbc.e11-02-0169

118. Hervouet, E. *et al.* Inhibition of cytochrome c oxidase subunit 4 precursor processing by the hypoxia mimic cobalt chloride. *Biochem. Biophys. Res. Commun.* **344,** 1086–1093 (2006). doi:10.1016/j.bbrc.2006.04.014

119. Kobayashi, M. *et al.* Mitochondrial intermediate peptidase is a novel regulator of sirtuin-3 activation by caloric restriction. *FEBS Lett.* **591,** 4067–4073 (2017). doi: 10.1002/1873-3468.12914

120. Lee, S.-F. *et al.* γ-Secretase-regulated Proteolysis of the Notch Receptor by Mitochondrial Intermediate Peptidase. *J. Biol. Chem.* **286,** 27447–27453 (2011). doi: 10.1074/jbc.M111.243154

121. Tasaki, T., Sriram, S. M., Park, K. S. and Kwon, Y. T. The N-End Rule Pathway. *Annu. Rev. Biochem.* (2012). doi:10.1146/annurev-biochem-051710-093308

122. Tasaki, T. and Kwon, Y. T. The mammalian N-end rule pathway: new insights into its components and physiological roles. *Trends in Biochemical Sciences* (2007). doi:10.1016/j.tibs.2007.08.010

123. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **20,** 1298–1345 (2011). doi:10.1002/pro.666

124. Lionaki, E. and Tavernarakis, N. Oxidative stress and mitochondrial protein quality control in aging. *Journal of Proteomics* (2013). doi:10.1016/j.jprot.2013.03.022

125. Carelli, V., Ross-Cisneros, F. N. and Sadun, A. A. Mitochondrial dysfunction as a cause of optic neuropathies. *Progress in Retinal and Eye Research* (2004). doi:10.1016/j.preteyeres.2003.10.003

126. Piotrowska, A., Korwin, M., Bartnik, E. and Tońska, K. Leber hereditary optic neuropathy — Historical report in comparison with the current knowledge. *Gene* **555,** 41–49 (2015). doi: 10.1016/j.gene.2014.09.048

127. Kirches, E. LHON: Mitochondrial Mutations and More. *Curr. Genomics* **12,** 44–54 (2011). doi: 10.2174/138920211794520150

128. Jankauskaitė, E., Bartnik, E. and Kodroń, A. Investigating Leber's hereditary optic neuropathy: Cell models and future perspectives. *Mitochondrion* **32,** 19–26 (2017). doi: 10.1016/j.mito.2016.11.006

129. Grazina, M. M. *et al.* Atypical presentation of Leber's hereditary optic neuropathy associated to mtDNA 11778G>A point mutation-A case report. *Eur. J. Paediatr. Neurol.* **11,** 115–118 (2007). doi: 10.1016/j.mito.2016.11.006

130. Mackey, D. A. *et al.* Primary pathogenic mtDNA mutations in multigeneration pedigrees with Leber hereditary optic neuropathy. *Am. J. Hum. Genet.* **59,** 481–5 (1996).

131. Pilz, Y. L., Bass, S. J. and Sherman, J. A Review of Mitochondrial Optic Neuropathies: From Inherited to Acquired Forms. *Journal of Optometry* (2016). doi:10.1016/j.optom.2016.09.003

132. Giordano, C. *et al.* Efficient mitochondrial biogenesis drives incomplete penetrance in Leber's hereditary optic neuropathy. *Brain* **137,** 335–353 (2014).

133. McClelland, C., Meyerson, C. and Van Stavern, G. Leber hereditary optic neuropathy: current perspectives. *Clin. Ophthalmol.* **39,** 1165 (2015).

134. Bu, X. D. and Rotter, J. I. X chromosome-linked and mitochondrial gene control of Leber hereditary optic neuropathy: evidence from segregation analysis for dependence on X chromosome inactivation. *Proc. Natl. Acad. Sci. U. S. A.* (1991). doi:10.1073/pnas.88.18.8198

135. Shankar, S. P. *et al.* Evidence for a novel X-linked modifier locus for leber hereditary optic neuropathy. *Ophthalmic Genet.* (2008). doi:10.1080/13816810701867607

136. Ji, Y. *et al.* Evaluation of the X-linked modifier loci for Leber hereditary optic neuropathy with the G11778A mutation in Chinese. *Mol. Vis.* (2010).

137. Ishikawa, K., Funayama, T., Ohde, H., Inagaki, Y. and Mashima, Y. Genetic variants of TP53 and EPHX1 in Leber's hereditary optic neuropathy and their relationship to age at onset. *Jpn. J. Ophthalmol.* **49,** 121–126 (2005). doi: 10.1007/s10384-004-0166-8

138. Istikharah, R. *et al.* Identification of the variants in PARL, the nuclear modifier gene, responsible for the expression of LHON patients in Thailand. *Exp. Eye Res.* **116,** 55–67 (2013). doi: 10.1016/j.exer.2013.08.007

139. Zhang, A.-M., Jia, X., Zhang, Q. and Yao, Y.-G. No association between the SNPs (rs3749446 and rs1402000) in the PARL gene and LHON in Chinese patients with m.11778G>A. *Hum. Genet.* **128,** 465–468 (2010). doi: 10.1007/s00439-010-0875-7

140. Abu-Amero, K. K., Jaber, M., Hellani, A. and Bosley, T. M. Genome-wide expression profile of LHON patients with the 11778 mutation. *Br. J. Ophthalmol.* **94,** 256–259 (2010). doi: 10.1136/bjo.2009.165571

141. Lee, S. *et al.* Mitochondrial Fission and Fusion Mediators, hFis1 and OPA1, Modulate Cellular Senescence. *J. Biol. Chem.* **282,** 22977–22983 (2007). doi: 10.1074/jbc.M700679200

142. Jiang, P. *et al.* The exome sequencing identified the mutation in YARS2 encoding the mitochondrial tyrosyl-tRNA synthetase as a nuclear modifier for the phenotypic manifestation of Leber's hereditary optic neuropathy-associated mitochondrial DNA mutation. *Hum. Mol. Genet.* **25,** 584–596 (2016). doi: 10.1093/hmg/ddv498

143. Chalmers, R. M. and Schapira, A. H. V. Clinical, biochemical and molecular genetic features of Leber's hereditary optic neuropathy. *Biochim. Biophys. Acta - Bioenerg.* (1999). doi:10.1016/S0005-2728(98)00163-7

144. Carelli, V. *et al.* Leber's hereditary optic neuropathy: Biochemical effect of 11778/ND4 and 3460/ND1 mutations and correlation with the mitochondrial genotype. *Neurology* **48,** 1623–1632 (1997). doi: 10.1212/WNL.48.6.1623

145. Brown, M. D., Trounce, I. A., Jun, A. S., Allen, J. C. and Wallace, D. C. Functional Analysis

of Lymphoblast and Cybrid Mitochondria Containing the 3460, 11778, or 14484 Leber's Hereditary Optic Neuropathy Mitochondrial DNA Mutation. *J. Biol. Chem.* **275,** 39831–39836 (2000). doi: 10.1074/jbc.M006476200

146. Chevrollier, A. *et al.* Hereditary optic neuropathies share a common mitochondrial coupling defect. *Ann. Neurol.* (2008). doi:10.1002/ana.21385

147. Cruz-Bermúdez, A. *et al.* Functional Characterization of Three Concomitant MtDNA LHON Mutations Shows No Synergistic Effect on Mitochondrial Activity. *PLoS One* **11,** e0146816 (2016). doi: 10.1371/journal.pone.0146816

148. Huoponen, K. Leber hereditary optic neuropathy: Clinical and molecular genetic findings. *Neurogenetics* (2001). doi:10.1007/s100480100115

149. Carelli, V., Giordano, C. and D'Amati, G. Pathogenic expression of homoplasmic mtDNA mutations needs a complex nuclear-mitochondrial interaction. *Trends in Genetics* **19,** 257–262 (2003). doi: 10.1016/S0168-9525(03)00072-6

150. Giordano, C. *et al.* Efficient mitochondrial biogenesis drives incomplete penetrance in Leber's hereditary optic neuropathy. *Brain* **137,** 335–353 (2014). doi: 10.1093/brain/awt343

151. Yu-Wai-Man, P., Votruba, M., Moore, A. T. and Chinnery, P. F. Treatment strategies for inherited optic neuropathies: past, present and future. *Eye* **28,** 521–537 (2014). doi: 10.1038/eye.2014.37

152. Bacalhau, M. *et al.* In silico analysis for predicting pathogenicity of five unclassified mitochondrial DNA mutations associated with mitochondrial cytopathies' phenotypes. *Eur. J. Med. Genet.* **60,** 172–177 (2017). doi: 10.1016/j.ejmg.2016.12.009

153. Wildeman, M., Van Ophuizen, E., Den Dunnen, J. T. and Taschner, P. E. M. Improving sequence variant descriptions in mutation databases and literature using the mutalyzer sequence variation nomenclature checker. *Hum. Mutat.* (2008). doi:10.1002/humu.20654

154. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7,** 248–249 (2010). doi: 10.1038/nmeth0410-248

155. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. and Chan, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One* **7,** (2012). doi: 10.1371/journal.pone.0046688

156. Choi, Y. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. *Proc. ACM Conf. Bioinformatics, Comput. Biol. Biomed. - BCB '12* 414–417 (2012). doi:10.1145/2382936.2382989

157. Kumar, P., Henikoff, S. and Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,** 1073–1081 (2009). doi: 10.1038/nprot.2009.86

158. Reva, B., Antipin, Y. and Sander, C. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.* **39,** (2011). doi: 10.1093/nar/gkr407

159. Reva, B., Antipin, Y. and Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.* **8,** (2007). doi: 10.1186/gb-2007-8-11-

r232

160. Schwarz, J. M., Cooper, D. N., Schuelke, M. and Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11,** 361–362 (2014). doi: 10.1038/nmeth.2890

161. Reese, M. G., Eeckman, F. H., Kulp, D. and Haussler, D. Improved Splice Site Detection in Genie. *J. Comput. Biol.* **4,** 311–323 (1997). doi: 10.1089/cmb.1997.4.311

162. Mullis, K. B. The Unusual Origin of the Polymerase Chain Reaction. *Sci. Am.* **262,** 56–65 (1990). doi: 10.1038/scientificamerican0490-56

163. Mullis, K. B. and Faloona, F. A. Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction. *Methods Enzymol.* **155,** 335–350 (1987). doi: 10.1016/0076-6879(87)55023-6

164. Joshi, M. and Deshpande, J. D. Polymerase Chain Reaction: Methods, Principles And Application. *Int. J. Biomed. Res.* **2,** (2011). doi: 10.7439/ijbr.v2i1.83

165. Berg, H. in *Gel Electrophoresis - Principles and Basics* (InTech, 2012). doi:10.5772/37724

166. Najafov, A., Hoxhaj, G., Najafov, A. and Hoxhaj, G. in *PCR Guru* 7–18 (2017). doi:10.1016/B978-0-12-804231-1.00002-X

167. Dieffenbach, C. W., Lowe, T. M. and Dveksler, G. S. General concepts for PCR primer design. *Genome Res.* S30–S37 (2010). doi:10.1101/gr.3.3.S30

168. Koressaar, T. and Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23,** 1289–1291 (2007). doi: 10.1093/bioinformatics/btm091

169. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40,** e115–e115 (2012). doi: 10.1093/nar/gks596

170. Rychlik, W., Spencer, W. J. and Rhoads, R. E. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* **19,** 698–698 (1991). doi: 10.1093/nar/19.3.698-a

171. Roux, K. H. Optimization and Troubleshooting in PCR. *Cold Spring Harb. Protoc.* **2009,** pdb.ip66-pdb.ip66 (2009). doi: 10.1101/pdb.ip66

172. Najafov, A. and Hoxhaj, G. Optimization and Troubleshooting. *PCR Guru* 31–43 (Elsevier, 2017). doi:10.1016/B978-0-12-804231-1.00004-3

173. Grunenwald, H. Optimization of Polymerase Chain Reactions. *PCR Protocols* **226,** 89–100 (Humana Press, 2003). doi: 10.1385/1-59259-384-4:89

174. Lotte Hansen, L. and Justesen, J. PCR Amplification of Highly GC-Rich Regions. *Cold Spring Harb. Protoc.* **2006,** pdb.prot4093 (2006). doi: 10.1101/pdb.prot4093

175. Musso, M., Bocciardi, R., Parodi, S., Ravazzolo, R. and Ceccherini, I. Betaine, Dimethyl Sulfoxide, and 7-Deaza-dGTP, a Powerful Mixture for Amplification of GC-Rich DNA Sequences. *J. Mol. Diagnostics* **8,** 544–550 (2006). doi: 10.2353/jmoldx.2006.060058

176. Lee, P. Y., Costumbrado, J., Hsu, C.Y. and Kim, Y. H. Agarose Gel Electrophoresis for the Separation of DNA Fragments. *J. Vis. Exp.* 3923 (2012). doi:10.3791/3923

177. Ylmaz, M., Ozic, C. and Gok, L. in *Gel Electrophoresis - Principles and Basics* (InTech, 2012). doi:10.5772/38654

178. Sharp, P. A., Sugden, B. and Sambrook, J. Detection of Two Restriction Endonuclease Activities in Haemophilus parainfluenzae Using Analytical Agarose-Ethidium Bromide Electrophoresis. *Biochemistry* **12,** 3055–3063 (1973). doi: 10.1021/bi00740a018

179. Sanger, F., Nicklen, S. and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* (1977). doi:10.1073/pnas.74.12.5463

180. Shendure, J. A. *et al.* Overview of DNA Sequencing Strategies. *Curr. Protoc. Mol. Biol.* **96,** 7.1.1-7.1.23 (2011). doi: 10.1002/0471142727.mb0701s96

181. Bell, J. R. A Simple Way to Treat PCR Products Prior to Sequencing Using ExoSAP-IT ®. *Biotechniques* **44,** 834 (2008). doi: 10.2144/000112890

182. Metzker, M. L. Emerging technologies in DNA sequencing. *Genome Research* (2005). doi:10.1101/gr.3770505

183. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72,** 248–254 (1976). doi: 10.1016/0003-2697(76)90527-3

184. Kruger, N. J. The Bradford Method for Protein Quantitation. *Protein Protocols Handbook, The* 15–22 (Humana Press, 2002). doi:10.1385/1-59259-169-8:15

185. Noble, J. E. and Bailey, M. J. A. Chapter 8 Quantitation of Protein. *Methods in Enzymology* 73–95 (2009). doi:10.1016/S0076-6879(09)63008-1

186. Shahrour, M. A. *et al.* Mitochondrial epileptic encephalopathy, 3-methylglutaconic aciduria and variable complex V deficiency associated with TIMM50 mutations. *Clin. Genet.* **91,** 690–696 (2017). doi: 10.1111/cge.12855

187. Roesch, K., Curran, S. P., Tranebjaerg, L. and Koehler, C. M. Human deafness dystonia syndrome is caused by a defect in assembly of the DDP1/TIMM8a-TIMM13 complex. *Hum. Mol. Genet.* **11,** 477–486 (2002). doi: 10.1093/hmg/11.5.477

188. Kang, Y. *et al.* Sengers Syndrome-Associated Mitochondrial Acylglycerol Kinase Is a Subunit of the Human TIM22 Protein Import Complex. *Mol. Cell* **67,** 457–470.e5 (2017). doi: 10.1016/j.molcel.2017.06.014

189. Do, R., Kathiresan, S. and Abecasis, G. R. Exome sequencing and complex disease: Practical aspects of rare variant association studies. *Hum. Mol. Genet.* (2012). doi:10.1093/hmg/dds387

190. Zhang, X. Exome sequencing greatly expedites the progressive research of Mendelian diseases. *Front. Med.* **8,** 42–57 (2014). doi: 10.1007/s11684-014-0303-9

191. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 19096–101 (2009). doi: 10.1073/pnas.0910672106

192. Rabbani, B., Tekin, M. and Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *J. Hum. Genet.* **59,** 5–15 (2014). 10.1038/jhg.2013.114

193. Kim, R. and Guo, J. Systematic analysis of short internal indels and their impact on protein folding. *BMC Struct. Biol.* **10,** 24 (2010). doi: 10.1186/1472-6807-10-24

194. Lin, M. *et al.* Effects of short indels on protein structure and function in human genomes. *Sci. Rep.* **7,** 9313 (2017). doi: 10.1038/s41598-017-09287-x

195. Melis, M. *et al.* Elevation of serum creatine kinase as the only manifestation of an intragenic deletion of the dystrophin gene in three unrelated families. *Eur. J. Paediatr. Neurol.* **2,** 255–261 (1998). doi: 10.1016/S1090-3798(98)80039-1

196. Elborn, J. S. Cystic fibrosis. *Lancet* **388,** 2519–2531 (2016). doi: 10.1016/S0140-6736(16)00576-6

197. Ku, C. S. *et al.* Exome versus transcriptome sequencing in identifying coding region variants. *Expert Review of Molecular Diagnostics* (2012). doi:10.1586/erm.12.10

198. Fernandez-Marmiesse, A., Gouveia, S. and Couce, M. L. NGS Technologies as a Turning Point in Rare Disease Research , Diagnosis and Treatment. *Curr. Med. Chem.* (2018). doi:10.2174/0929867324666170718101946

199. Salgado, D., Bellgard, M. I., Desvignes, J. P. and Béroud, C. How to Identify Pathogenic Mutations among All Those Variations: Variant Annotation and Filtration in the Genome Sequencing Era. *Human Mutation* (2016). doi:10.1002/humu.23110

200. Sandmann, S. *et al.* Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. *Sci. Rep.* **7,** 43169 (2017). doi: 10.1038/srep43169

201. Mu, W., Lu, H. M., Chen, J., Li, S. and Elliott, A. M. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *J. Mol. Diagnostics* **18,** 923–932 (2016). doi: 10.1016/j.jmoldx.2016.07.006

202. Hamilton, A. *et al.* Concordance between whole-exome sequencing and clinical Sanger sequencing: implications for patient care. *Mol. Genet. Genomic Med.* **4,** 504–512 (2016). doi: 10.1002/mgg3.223

203. Berman, H., Henrick, K. and Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural Biology* (2003). doi:10.1038/nsb1203-980

204. Haas, J. *et al.* The protein model portal - A comprehensive resource for protein structure and model information. *Database* (2013). doi:10.1093/database/bat031

205. Glusman, G. *et al.* Mapping genetic variations to three-dimensional protein structures to enhance variant interpretation: a proposed framework. *Genome Med.* **9,** 113 (2017). doi: 10.1186/s13073-017-0509-y

206. Seifi, M. and Walter, M. A. Accurate prediction of functional, structural, and stability changes in PITX2 mutations using in silico bioinformatics algorithms. *PLoS One* **13,** e0195971 (2018). doi: 10.1371/journal.pone.0195971

207. Signes, A. and Fernandez-Vizarra, E. Assembly of mammalian oxidative phosphorylation complexes I–V and supercomplexes. *Essays Biochem.* **62,** 255–270 (2018). doi: 10.1042/EBC20170098

208. Li, Y., Park, J. S., Deng, J. H. and Bai, Y. Cytochrome c oxidase subunit IV is essential for assembly and respiratory function of the enzyme complex. *J. Bioenerg. Biomembr.*

(2006). doi:10.1007/s10863-006-9052-z

209. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17,** 405–423 (2015). doi: 10.1038/gim.2015.30

210. Manolio, T. A. *et al.* Bedside Back to Bench: Building Bridges between Basic and Clinical Genomic Research. *Cell* **169,** 6–12 (2017). doi: 10.1016/j.cell.2017.03.005

211. Sun, S. *et al.* An extended set of yeast-based functional assays accurately identifies human disease mutations. *Genome Res.* **26,** 670–680 (2016). doi: 10.1101/gr.192526.115

212. Starita, L. M. *et al.* Variant Interpretation: Functional Assays to the Rescue. *Am. J. Hum. Genet.* **101,** 315–325 (2017). doi: 10.1016/j.ajhg.2017.07.014

213. Calvaruso, M. A., Smeitink, J. and Nijtmans, L. Electrophoresis techniques to investigate defects in oxidative phosphorylation. *Methods* (2008). doi:10.1016/j.ymeth.2008.09.023

214. Li, Y. *et al.* An assembled complex IV maintains the stability and activity of complex I in mammalian mitochondria. *J. Biol. Chem.* (2007). doi:10.1074/jbc.M701056200