

Université de Montréal

Étude du protéome alternatif d'origine mitochondriale chez l'humain

Par
Laura Kienzle

Département de sciences biologiques, Faculté des Arts et des Sciences

Mémoire présenté
en vue de l'obtention du grade de maîtrise (M. Sc.) en sciences biologiques

Avril 2021

© Laura Kienzle, 2021

Ce mémoire intitulé

Étude du protéome alternatif d'origine mitochondriale chez l'humain

Présenté par
Laura Kienzle

A été évalué par un jury composé des personnes suivantes

Bernard Angers
Président-rapporteur

Sophie Breton
Directrice de recherche

Annie Angers
Codirectrice de recherche

Benoît Vanderperre
Membre du jury

Résumé

Les mitochondries, organelles d'origine bactérienne, sont trouvées dans les cellules de presque tous les organismes eucaryotes. Elles exercent des rôles centraux dans les fonctions cellulaires tels que la production d'énergie, la signalisation cellulaire et l'apoptose et ont aussi un impact sur le vieillissement ainsi que certains cancers et maladies neurodégénératives. Chez l'humain et les mammifères en général, le génome mitochondrial est une molécule d'ADN double brin circulaire composée de 37 gènes. Seulement 13 de ces gènes codent des protéines mitochondriales et les 24 autres produisent 22 ARNt (ARN de transfert) et 2 ARNr (ARN ribosomal) qui sont nécessaires à la traduction des 13 protéines mitochondriales. L'ADNmt (ADN mitochondrial) étant très compact, ceci suggère qu'il y a peu de possibilités pour des nouveautés évolutives. Cependant, de récentes recherches ont permis de révéler la présence de près d'une dizaine de petits ORF (cadres de lecture ouverts) fonctionnels à l'intérieur des gènes mitochondriaux 12S ARNr et 16S ARNr. Ceci remet en question la complexité du génome mitochondrial et montre que son potentiel codant a été sous-estimé. Une analyse approfondie du génome mitochondrial humain a révélé la présence de 227 séquences potentiellement traduites en protéines mitochondriales à travers l'ensemble du génome. Dans cette étude, nous avons sélectionné 9 de ces 227 séquences afin de déterminer si effectivement, elles produisent un peptide identifiable. Pour ce faire, des expériences d'immunobuvardage, d'immunofluorescence et d'immunoprécipitation ont été réalisées sur des cellules HeLa et des cellules HEK293T. Ces expériences ont permis d'identifier une protéine mitochondriale alternative nommée MTALTND4 dont la séquence codante est trouvée à l'intérieur du gène *nd4*, dans un cadre de lecture alternatif. MTALTND4 est traduite dans la mitochondrie et peut être exportée dans le cytoplasme ainsi qu'à l'extérieur de la cellule puisqu'elle a été retrouvée dans le plasma humain. Bien que la fonction de cette protéine n'ait pas encore été confirmée, des

résultats préliminaires indiquent qu'elle a un impact sur la respiration cellulaire. MTALTND4 diminue la respiration mitochondriale et nos résultats suggèrent que son action serait induite par l'hypoxie. La découverte de ce nouveau gène mitochondrial humain confirme que le potentiel codant du génome mitochondrial est beaucoup plus vaste que ce que nous croyions. Il existe fort probablement encore plusieurs autres protéines mitochondriales dont les effets pourraient se révéler d'une grande importance. En effet, plusieurs des protéines dérivées du génome mitochondrial découvertes à ce jour ont des impacts majeurs au niveau du métabolisme et pourraient agir en tant que molécules thérapeutiques importantes. Nos résultats amènent à repenser l'évolution et les pressions de sélection exercées sur le génome mitochondrial et ouvrent la porte à de nombreuses recherches futures qui permettront de re-caractériser le génome mitochondrial et d'avoir une compréhension encore plus approfondie du rôle des mitochondries dans les fonctions cellulaires.

Mots-clé: Mitochondries, Protéome Alternatif, Humain, ORF, altORF, smORF, Humanine, MOTS-c, SHLPs, Gau, MTALTND4

Abstract

Mitochondria, organelles of bacterial origin, are found in almost every eukaryotic organism and play a central role in cellular functions such as energy production, cellular signaling and apoptosis and are also known to have an impact on aging, certain cancers and neurodegenerative diseases. In humans and mammals in general, the mitochondrial DNA is a small double-stranded circular molecule coding for only 37 genes. Only 13 of them code for mitochondrial proteins and the other 24 genes produce 22 tRNAs (transfer RNA) and 2 rRNAs (ribosomal RNA) necessary for the translation of the 13 protein coding genes. The extremely compact nature of mtDNA (mitochondrial DNA) suggests that there is little room for evolutionary novelties. However, recent research revealed the presence of about ten small functional ORFs inside the mitochondrial genes 12S rRNA and 16S rRNA. This calls into question the complexity of the mitochondrial genome and shows that its coding potential has been greatly underestimated. A thorough examination of the human mitochondrial genome revealed the presence of 227 sequences potentially translated into mitochondrial alternative proteins across the entire genome. In this study, we selected 9 of the 227 sequences to determine if they indeed produce identifiable peptides. This was done by immunoblotting, immunofluorescence and immunoprecipitation experiments on HeLa and HEK293T cells. These experiments allowed us to identify one alternative protein named MTALTND4 whose coding sequence is found inside the *nd4* gene, in an alternative sequence. MTALTND4 is translated inside the mitochondria and can be exported in the cytoplasm as well as outside the cell since it has been found in human plasma. Although the function of this protein has not yet been confirmed preliminary results indicate its impact on cellular respiration. MTALTND4 decreases mitochondrial respiration and our results suggest that its action could be induced by hypoxia. The discovery of this new human mitochondrial gene confirms that the coding

potential of the mitochondrial genome is much larger than we thought. There are most likely still many other mitochondrial proteins whose effects could prove to be of great importance. Indeed, several of the mitochondrial derived proteins discovered to date have major impacts on metabolism and could act as important therapeutic molecules. Our results lead to rethink the evolution and the selection pressures exerted on the mitochondrial genome and open the door to many future researches which will allow to re-characterize the mitochondrial genome and to have an even deeper understanding of the role of mitochondria in cellular functions.

Keywords: Mitochondria, Alternative Proteome, Human, ORF, altORF, smORF, Humanin, MOTS-c, SHLPs, Gau, MTALTND4

Table des matières

Résumé.....	2
Abstract.....	5
Table des matières	7
Liste des tableaux	10
Liste des figures.....	11
Liste des sigles et abréviations.....	16
Remerciements.....	19
Introduction	20
1.1 Les mitochondries	20
1.2 Génome mitochondrial	21
1.3 Initiation de la traduction	23
1.3.1 Initiation de la traduction chez les procaryotes.....	24
1.3.2 Initiation de la traduction chez les eucaryotes: génome nucléaire.....	26
1.3.3 Initiation de la traduction chez les eucaryotes: génome mitochondrial	28
1.4 Annotation des génomes.....	29
1.4.1 ORF.....	29
1.4.2 smORF et altORF.....	30

1.5 Des gènes codants supplémentaires dans les génomes mitochondriaux	33
1.6. Hypothèse et objectifs.....	36
The alternative human proteome of mitochondrial origin.....	37
ABSTRACT	38
INTRODUCTION	39
MATERIALS & METHODS.....	40
Identification of mitochondrial smORFs and altORFs and choice of candidates for antibody production	40
Analyses of Mtaltd4 sequences.....	43
Rabbit anti-MTALTND4 antibody generation	44
Cell culture.....	44
Western blotting.....	45
Deglycosylation	46
Immunofluorescence.....	47
Immunoprecipitation and sequencing	47
RESULTS & DISCUSSION.....	49
3. Discussion	57
3.1 Identification d'une protéine alternative par immunobuvardage	57
3.2 Détermination de l'origine de MTALTND4.....	61

3.3 Localisation de MTALTND4	62
3.4 Poids moléculaire de MTALTND4.....	63
3.5 Immunoprécipitation et séquençage par spectrométrie de masse	66
3.6 Fonction de MTALTND4	67
Conclusion et perspectives	70
Références	73
Annexes	87
Supplementary Material.....	87
Supplementary Figures	87
Supplementary Tables.....	90

Liste des tableaux

Table S1. Chosen candidates for antibody production.

Table S2. Identification of mitochondrial smORFs and altORFs - in silico + kozak approach.

Table S3. Identification of mitochondrial smORFs and altORFs - OpenProt approach and sequences selected for antibody production.

Table S4. Identification of mitochondrial smORFs and altORFs - MS approach and sequences selected for antibody production.

Table S5. List of proteins found by mass spectrometry following immunoprecipitation of MTALTND4.

Liste des figures

Figure 1. Carte du génome mitochondrial humain représentant les 37 gènes et leur organisation. Les 13 gènes codant pour des protéines sont représentés en vert, mauve, bleu ou rouge selon le complexe protéique qu'ils forment, les ARNr sont en jaune, les ARNt sont en gris foncé et les régions non-codantes en blanc. Tirée et modifiée de Amorim *et al.* (2019)..... 23

Figure 2. Liaison de la séquence SD (Shine-Dalgarno) de l'ARNm et de la partie complémentaire en 3' de l'ARNr 16S de la petite sous-unité ribosomique procaryote. La complémentarité des séquences permet de positionner le codon d'initiation de l'ARNm au site P du ribosome. Tirée de Yang *et al.* (2016)..... 25

Figure 3. Multiples sites de liaison du ribosome avec un ARNm procaryote. L'ARNm code pour plusieurs protéines, c'est une molécule polycistronique. Tirée et modifiée de Alberts *et al.* (2011)..... 25

Figure 4. Coiffe en 5' et queue polyA en 3' d'un ARNm nucléaire eucaryote. La molécule est monocistronique. Tirée et modifiée de Alberts *et al.* (2011)..... 27

Figure 5. Circularisation de l'ARNm par l'interaction entre la coiffe en 5', les facteurs eIF4E et eIF4G et la protéine PABP liée à la queue polyA. La petite sous-unité ribosomale est recrutée à l'extrémité 5' de l'ARNm et scanne la molécule vers l'extrémité 3'. Une fois arrivée au codon AUG, la

grosse sous-unité ribosomale est recrutée et la traduction commence. Tirée et modifiée de Kasinath *et al.* (2009).. 28

Figure 6. Structure d'un ARNm eucaryote typique en trois régions avec une seule séquence codante annotée (refCDS). Les régions 5'UTR et 3'UTR ne sont pas traduites. Tirée et modifiée de Mouilleron *et al.* (2015)..... 30

Figure 7. Organisation des altORF (en bleu) par rapport aux refCDS (en gris foncé) d'un ARNm. Les altORF sont dans le même cadre de lecture (à gauche) ou dans un autre cadre de lecture (à droite) que le refCDS. Les codons start des refCDS sont représentés par les triangles pleins et les codons start alternatifs par les triangles vides. Les altORF peuvent être dans les régions non-traduites des ARNm ou complètement à l'intérieur du refCDS ou encore partiellement à l'intérieur. Tirée et modifiée de Orr *et al.* (2019)..... 31

Figure 8. (a) Représentation typique d'un ARNm eucaryote mature et ses altORF possibles. (b) Représentation du gène ATXN1 et de son altORF AltATXN1. AltATXN1 est dans un autre cadre de lecture que le refORF et il chevauche son côté N-terminal. Une fois les deux protéines traduites, la protéine alternative AltATXN1 interagit avec le domaine N-terminal de la protéine ATXN1. Tirée de Landry *et al.* (2015)..... 33

Figure 9. Emplacement des peptides Humanine, MOTS-c, SHLP1-6 et Gau dans l'ADNmt humain. Tirée de Minasyan *et al.* (2017)..... 34

Figure 10. Carte du génome mitochondrial humain montrant les nouveaux cadres de lecture caractérisés (en bleu) et tous les cadres de lecture ouverts potentiels sur les brins sens (flèches orangées) et antisens (flèches vertes). Tirée de Angers *et al.* (2019)..... 35

Figure 11. Figure 1. Map of the human mitochondrial genome. All 37 typical genes are represented on the outside circle of the map. New characterized alternative mitochondrial peptides such as Humanin, MOTS-c, SHLP1-6 and Gau are represented by blue arrows. 227 unannotated ORFs of >60 nucleotides are represented by black, white, pink and red arrows. ORFs in frames 1, 2 and 3 are represented by black arrows, ORFs in frames -1, -2 and -3 are represented by white arrows. The 9 ORFs chosen for this study are represented by pink arrows and the red arrow that codes for MTALTND4 alternative protein..... 50

Figure 12. Figure 2. Identification of a novel altORF in the *nd4* gene of the human mitochondrial genome. (A) Location of *mtaltnd4* ORF in the *nd4* gene and its translation in mitochondria considering AGA and AGG codons as coding. The antigenic sequence used for antibody production is in bold and underlined. (B) Peptide sequence alignment in 5 primate species: human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), gorilla (*Gorilla gorilla*) and orangutan (*Pongo pygmaeus*). (C) Immunoblotting of MTALTND4 on HeLa and HEK293T cells. The apparent molecular weight is 25kDa. (D) Immunoblotting of MTALTND4, mitochondrial-encoded cytochrome C oxidase 1 and nuclear-encoded actin and ATP5 on normal HeLa cells and HeLa- ρ 0 cells. (E) Immunoblotting of MTALTND4, mitochondrial-encoded cytochrome C oxidase 1 and nuclear-encoded actin and ATP5 on normal HeLa cells (-) and HeLa cells treated with chloramphenicol antibiotic (+). (F) Immunoblotting of

MTALTND4 and ATP5 on human plasma. MTALTND4 appears at 25 kDa and ATP5 is absent. **(G)** Immunoblotting of MTALTND4 on non-treated HeLa cells (-) and HeLa cells treated with deglycosylation enzymes (+). **(H)** Coomassie Blue staining of non-treated Fetuin (-) and Fetuin treated with deglycosylation enzymes (+). **(I)** Immunoblotting of MTALTND4 on non-treated HeLa cells (-) and HeLa cells treated with Lambda phosphatase (+)..... 53

Figure 13. Figure 3. Localization of MTALTND4 by immunofluorescence. **(A)** Nuclear marker DAPI (blue), nuclear-encoded mitochondrial marker ATP5 (red) and MTALTND4 (green) by immunofluorescence in HeLa cells. Scale bar, 20µm. **(B)** Nuclear marker DAPI (blue), nuclear-encoded mitochondrial marker ATP5 (red) MTALTND4 (green) by immunofluorescence in HeLa cells treated with chloramphenicol after 48 hours. Scale bar, 20µm..... 55

Figure 14. Immunobuvardage du peptide synthétique. Immunobuvardage de MTALTND4 sur le peptide synthétique en présence d’agents dénaturants (Peptide D) et en absence d’agents dénaturants (Peptide ND) 65

Figure 15. Respirométrie à haute résolution de cellules HeLa. La concentration d’oxygène est représentée en bleu et le flux d’oxygène est représenté en rouge. La figure du haut est le contrôle et montre que l’ajout d’eau dans la chambre cellulaire ne cause pas de diminution du flux d’oxygène, tandis que dans la figure du bas, l’ajout du peptide synthétique cause une nette diminution du flux d’oxygène..... 68

Figure 16. Immunobuvardage de MTALTND4 sur des cellules HeLa en hypoxie. Une bande principale est observée à 25 kDa et deux autres bandes pointées par les flèches en rouge à environ 11,5 kDa et 14 kDa peuvent aussi être observées..... 69

Figure 17. Figure S1. Interrogation of mitochondria-derived mass spectrometry datasets using the novel PepQuery tool

Predicted protein sequences are digested *in silico* with trypsin. Then each peptide is queried using the novel PepQuery tool from which any Peptide Spectrum Match (PSM) that hits better with the queried peptide than with any protein from the reference database is retrieved..... 87

Figure 18. Figure S2. Comparison of MTALTND4 peptide encoded in mitochondrial DNA (mtDNA) and hypothetically in nuclear DNA that have been transferred from mtDNA (NUMT) through evolution. Only sequences with >50% identities and complete without stop codons are shown..... 88

Figure 19. Figure S3. Multiple peptide sequence alignment in 15 mammal species: human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), naked mole rat (*Heterocephalus glaber*), dog (*Canis lupus familiaris*), cow (*Bos taurus*), zebrafish (*Danio rerio*), lion (*Panthera leo*), bear (*Ursus arctos*), horse (*Equus caballus*), dolphin (*Tursiops truncatus*)..... 89

Liste des sigles et abréviations

Les termes en anglais sont en caractères italiques, les termes en français sont en caractères droits.

ADP: Adénosine Diphosphate

ADNmt: Acide Désoxyribonucléique mitochondrial

altORF: *alternative Open Reding Frame* / cadre de lecture ouvert alternatif

ARN: Acide Ribonucléique

ARNm: ARN messenger

ARNr: ARN ribosomal

ARNt: ARN de transfert

ATP: Adénosine Triphosphate

BLAST: *Basic Local Alignment Search Tool*

BLASTN: *from a nucleotide sequence to a nucleotide sequence*

TBLASTn: *from a protein sequence to a translated nucleotide sequence*

CBC: *Cap Binding Protein*

CTE: Chaîne de Transport des Électrons

DMEM: *Dulbecco's Modified Eagle Medium*

DTT: Dithiothreitol

eIF3: *eukaryotic translation Initiation Factor 3*

eIF4E: *eukaryotic translation Initiation Factor 4E*

eIF4G: *eukaryotic translation Initiation Factor 4G*

ELISA : *Enzyme-Linked Immunosorbent Assay*

FDR: *False Discovery Rate*

GPx-1: Glutathione peroxidase

HN: *Humanin* / Humanine

HRP: *Horseradish Peroxidase*

mIF2: *mitochondrial Initiation Factor 2*

mIF3: *mitochondrial Initiation Factor 3*

MOTS-c: *Mitochondrial Open reading frame of the 12S rRNA-c*

MS: *Mass Spectrometry*

MTALTND4: *Mitochondrial Alternative ND4 protein*

mtaltORF: *mitochondrial alternative Open Reading Frame*

MUHC: *McGill University Health Center* / Centre Universitaire de Santé McGill

NCBI: *National Center for Biotechnology Information*

NEB: *New England Biolabs*

NGS: *Normal Goat Serum*

NUMTs: *Nuclear Mitochondrial DNA sequence*

ORF: *Open Reading Frame* / Cadre de lecture ouvert

OXPHOS: *Oxidative Phosphorylation*/Phosphorylation oxidative

PABP1: *Polyadenylate-Binding Protein 1*

PAGE: *Polyacrylamide Gel Electrophoresis*

PBS: *Phosphate-Buffered Saline*

PCA: Polycycloalkanes

PMSF: Phenylmethylsulfonyl Fluoride

PSM: *Peptide Spectrum Match*

PTM: *Post-Translational Modifications*

PVDF: Polyvinylidene Difluoride

refCDS: *reference Coding Sequence* / séquence codante de référence

refORF: *reference Open Reading Frame* / cadre de lecture ouvert de référence

SD: Shine-Dalgarno

SDS: Sodium Dodecyl Sulfate

SHLPs: *Small Humanin-Like Peptides*

smORF: *small Open Reading Frame*

TBS: *Tris-buffered saline*

TBS-T: *Tris-buffered saline with Tween 20*

TM: *Transmembrane*

UTR: *Untranslated Region*

Remerciements

Je tiens premièrement à remercier ma directrice de recherche Sophie Breton pour son soutien, son dynamisme et ses encouragements constants tout au long de ma maîtrise. Merci Sophie de m'avoir continuellement guidée, aidée et motivée à travers ce projet ambitieux et à travers les nombreux événements qui ont interrompu mon travail. Je tiens aussi à remercier tous mes collègues qui m'ont intégrée à leur équipe et qui ont su créer une ambiance des plus agréable et motivante. Mélanie, Stefano, Charlotte, Georges, Andréanne, Thierry et Rosa, merci pour tous les bons moments passés ensemble que ce soit au laboratoire ou à l'extérieur. Je tiens particulièrement à remercier Mélanie Tassé qui a été et qui restera toujours une amie extraordinaire. Merci Mélanie de m'avoir encouragée et aidée à travers ma maîtrise et d'avoir été mon compagnon de laboratoire en temps de pandémie et de confinement. Tu as su embellir mes journées et les rendre drôles et agréables. Je tiens aussi à remercier ma co-directrice de recherche Annie Angers. Merci Annie pour tout le temps que tu as consacré à m'enseigner les techniques de western blot, d'immunofluorescence et d'immunoprécipitation. J'ai toujours apprécié ces moments pour apprendre et échanger des idées avec toi.

Je tiens aussi à remercier tous les membres de ma famille et ami(e)s qui m'ont toujours soutenue et encouragée dans mon projet. Je remercie particulièrement ma mère et David pour leur support exceptionnel et pour avoir manifesté un grand intérêt pour ma recherche. Merci à tous d'avoir écouté mes histoires de mitochondries. Pour finir, je tiens à remercier mon copain Guillaume pour tout le support moral qu'il m'a procuré. Merci Guillaume pour ton aide, ta générosité et ton soutien tout au long de ces deux années.

Introduction

1.1 Les mitochondries

Les mitochondries, organelles d'origine bactérienne, sont des composantes essentielles de presque toutes les cellules eucaryotes. À l'exception d'un seul microorganisme eucaryote unicellulaire qui est dépourvu de mitochondries (*Monocercomonoides* sp.), tous les autres eucaryotes possèdent ces organelles époustouflantes (Karnkowska et al., 2016). Exerçant plusieurs fonctions dans la cellule, le rôle le plus connu des mitochondries est leur capacité à produire de l'énergie. L'énergie chimique, sous forme de molécules d'ATP dans la cellule, est produite dans la mitochondrie par un processus nommé la phosphorylation oxydative (OXPHOS) durant lequel des électrons sont transportés à travers une chaîne de complexes protéiques jusqu'à l'oxygène. Le transport des électrons à travers la CTE (chaîne de transport des électrons) génère un gradient de protons qui force éventuellement leur passage à travers l'ATP synthase. Ce passage engendre la phosphorylation d'ADP en ATP qui peut alors être utilisé par la cellule comme source d'énergie. L'oxygène étant l'accepteur final des électrons de la CTE, il est essentiel à la synthèse d'ATP (Junge et Nelson, 2015). Bien que la production d'énergie soit le rôle principal que l'on attribue aux mitochondries, celles-ci sont aussi impliquées dans plusieurs autres processus tels que le métabolisme, l'apoptose et la signalisation cellulaire et ont aussi un impact sur le vieillissement ainsi que plusieurs cancers et maladies neurodégénératives (Nunnari et Suomalainen, 2012; Yee et al., 2014; Wang et Hekimi, 2015; Ladiges et al., 2010; Lin et Beal, 2006). Étant donné les multiples rôles que jouent les mitochondries au sein des cellules, il est important de bien comprendre leur fonctionnement et donc de bien connaître leur génome.

1.2 Génome mitochondrial

La mitochondrie est une ancienne bactérie qui est devenue la mitochondrie que nous connaissons aujourd'hui suite à son endosymbiose avec un groupe d'organismes unicellulaires : les archées d'Asgard. Cette endosymbiose est à l'origine des cellules eucaryotes (Roger et al., 2017 ; Wallace, 2018 ; Martijn et al., 2018 ; Gray, 2012). Il a longtemps été suggéré que la bactérie d'origine appartenait au groupe des α -protéobactéries, mais de récentes études suggèrent qu'elle pourrait plutôt appartenir à une lignée de protéobactéries qui aurait divergé avant la diversification de toutes les lignées d' α -protéobactéries connues à ce jour (Martijn et al., 2018). Étant entourées de deux membranes, les mitochondries possèdent un espace inter-membranaire ainsi qu'une matrice. C'est dans la matrice qu'est retrouvé le génome mitochondrial. En effet, de par leur origine exogène, les mitochondries possèdent leur propre génome, distinct du génome nucléaire de la cellule, sous forme d'ADN double brin circulaire. La publication de la première séquence complète de l'ADNmt (ADN mitochondrial) humain en 1981 indique qu'il est composé de 37 gènes dont 13 qui codent des protéines essentielles qui participent à la formation des complexes respiratoires de la CTE (complexes I-IV) ainsi que l'ATP synthase (complexe V) impliqués dans la production d'énergie par le processus de phosphorylation oxydative dans la membrane interne des mitochondries. L'ADNmt est également composé de 2 ARN ribosomiaux (ARNr 12S et 16S) et 22 ARNt (ARN de transfert) qui sont nécessaires à la traduction des 13 protéines mitochondriales (Figure 1) (Taanman, 1999; Anderson et al., 1981). Les autres protéines retrouvées dans la mitochondrie sont plutôt codées par le noyau, puis importées dans la mitochondrie. Ceci peut être expliqué par le fait qu'au cours de l'évolution de la cellule eucaryote, une grande partie des gènes du génome bactérien primitif de la mitochondrie a été transférée au génome nucléaire ou a tout simplement été remplacée par des gènes nucléaires exerçant une

fonction similaire (Adams et Palmer, 2003). Ainsi, au cours de l'évolution de la lignée des métazoaires, le génome mitochondrial s'est réduit pour être aujourd'hui, chez l'humain et les mammifères en général, très compact avec une taille d'environ 16 000 paires de bases. Il existe cependant certaines espèces d'invertébrés telles que des éponges, des cnidaires, le ver polychète, des isopodes et des nématodes qui ont un génome mitochondrial atypique, c'est-à-dire pouvant être beaucoup plus imposant que 16 000 paires de bases (jusqu'à 46 985 pb chez le mollusque *Scapharca broughtonii*) et pouvant présenter des introns normalement absents du génome typique de la mitochondrie animale (Breton et al., 2014). Contrairement au génome nucléaire, le génome mitochondrial animal typique ne possède presque aucune région non-codante. Les gènes sont arrangés les uns après les autres et sont séparés par de petites régions non-codantes de quelques nucléotides seulement (Anderson et al., 1981). De plus, presque tous les gènes (28 sur 37, ce qui représente environ 95% de l'ADN codant) sont situés sur le même brin d'ADN, le brin complémentaire étant considéré comme non-codant (à l'exception de 8 ARNt et du gène *nd6* qui représentent environ 5% de l'ADN codant) (Taanman, 1999). En plus d'être différente du génome nucléaire, l'organisation du génome mitochondrial diffère de celle des génomes bactériens, ancêtres des mitochondries. Les bactéries ont aussi un génome circulaire, mais beaucoup plus imposant dans lequel on retrouve des gènes sur les deux brins d'ADN et même des gènes à l'intérieur d'autres gènes (Meydan et al., 2018).

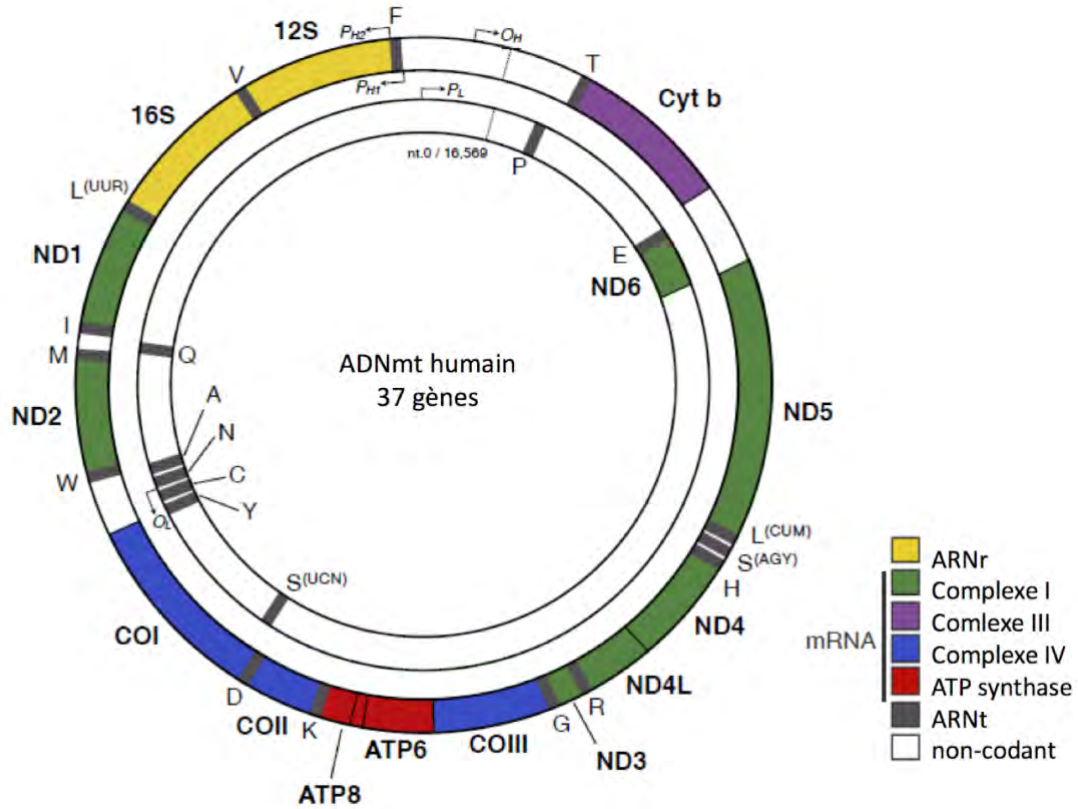


Figure 1. Carte du génome mitochondrial humain représentant les 37 gènes et leur organisation. Les 13 gènes codant pour des protéines sont représentés en vert, mauve, bleu ou rouge selon le complexe protéique qu'ils forment, les ARNr sont en jaune, les ARNt sont en gris foncé et les régions non-codantes en blanc. Figure tirée et modifiée de Amorim *et al.* (2019).

1.3 Initiation de la traduction

Tout comme les différences par rapport à la structure des génomes bactérien, nucléaire et mitochondrial, il est intéressant de noter que les mécanismes de réplication, de transcription et d'initiation de la traduction des gènes diffèrent entre les organismes procaryotes et eucaryotes ainsi qu'entre les génomes nucléaires et mitochondriaux des eucaryotes. Nous regardons ici plus en détails les distinctions entre les mécanismes de traduction des ARNm.

1.3.1 Initiation de la traduction chez les procaryotes

Chez les organismes procaryotes tels que les bactéries, le mécanisme d'initiation de la traduction le plus étudié implique une séquence particulière retrouvée sur l'ARNm. Cette séquence se nomme Shine-Dalgarno. Elle est typiquement retrouvée aux alentours de 7-8 nucléotides en amont du codon d'initiation de la traduction et a comme séquence consensus AGGAGG (Amin et al., 2018). L'ARNr 16S, qui constitue la petite sous-unité 30S du ribosome, possède à son extrémité 3' une séquence complémentaire à la séquence Shine-Dalgarno, nommée anti-Shine-Dalgarno, et permet l'appariement entre l'ARNm et l'ARNr. Cet appariement permet de positionner le ribosome sur le codon d'initiation afin d'initier la traduction efficace de l'ARNm (Rodina, 2018; Yang et al., 2016) (Figure 2). Grâce à ce mécanisme, le ribosome procaryote peut s'assembler n'importe où sur l'ARNm où l'on retrouve une séquence Shine-Dalgarno. De plus, comme les ARNm procaryotes possèdent souvent plus d'un codon d'initiation de la traduction, les ribosomes peuvent s'assembler à plusieurs endroits sur l'ARNm et ainsi produire plusieurs protéines à partir du même ARNm. Un ARNm qui code plusieurs protéines est appelé polycistronique (Cao et al., 2015) (Figure 3).

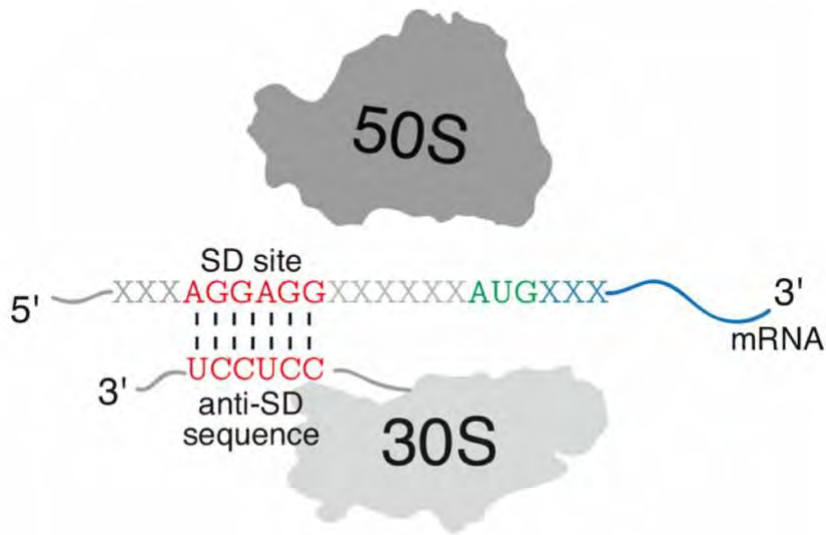


Figure 2. Liaison de la séquence SD (Shine-Dalgarno) de l'ARNm et de la partie complémentaire en 3' de l'ARNr 16S de la petite sous-unité ribosomique procaryote. La complémentarité des séquences permet de positionner le codon d'initiation de l'ARNm au site P du ribosome. Figure tirée de Yang *et al.* (2016).

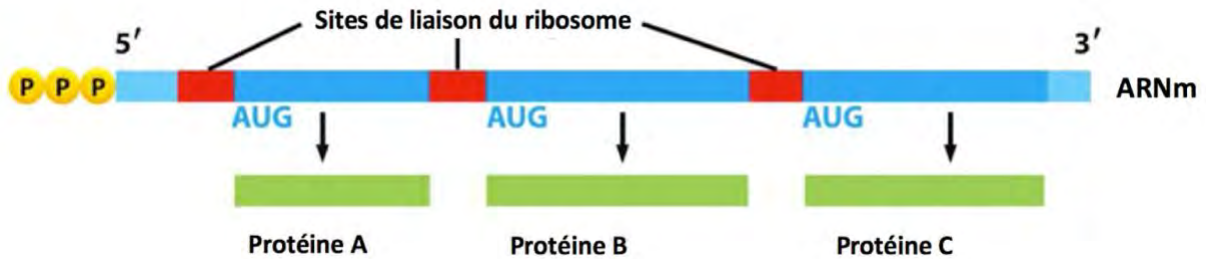


Figure 3. Multiples sites de liaison du ribosome avec un ARNm procaryote. L'ARNm code plusieurs protéines, c'est une molécule polycistronique. Figure tirée et modifiée de Alberts *et al.* (2011).

1.3.2 Initiation de la traduction chez les eucaryotes : génome nucléaire

Les ARNm nucléaires des organismes eucaryotes possèdent une coiffe à leur extrémité 5' ainsi qu'une queue polyA à leur extrémité 3' (Figure 4). La coiffe est une guanosine méthylée en position 7 et qui est liée au premier nucléotide de l'ARNm par un lien triphosphate 5' à 5'. La queue polyA est un ajout d'environ 200 adénosines à l'extrémité 3' de l'ARNm. Ces modifications permettent de protéger la molécule contre la dégradation induite par des exonucléases présentes dans le cytoplasme de la cellule (Ramanathan et al., 2016). Ces ajouts à l'ARNm servent aussi à l'initiation de la traduction. En effet, plusieurs facteurs protéiques nommés CBC pour « cap-binding-complex » s'associent spécifiquement à la structure de coiffe en 5' de l'ARNm. Parmi ceux-ci, on retrouve les facteurs eIF4E et eIF4G. Une fois le facteur eIF4E positionné à la coiffe en 5', le facteur eIF4G interagit avec lui ainsi qu'avec la protéine PABP1 qui est liée à la queue polyA de l'extrémité 3' de l'ARNm. Cette interaction va former une structure pseudo-circulaire de l'ARNm (Figure 5). La circularisation de la molécule pourrait être un moyen de s'assurer qu'uniquement les ARNm complets soient traduits en protéines (Ramanathan et al., 2016). La présence des facteurs associés à l'ARNm permet de recruter la petite sous-unité 40S du ribosome à l'extrémité 5' de la molécule. Plus précisément, le facteur eIF3 lié à la petite sous-unité ribosomale se lie au facteur eIF4G, ce qui permet de positionner la sous-unité à l'extrémité 5' de l'ARNm. Par la suite, l'ARNm défile à travers la sous-unité, ce qui lui permet de « scanner » l'ARNm dans la direction 5' - 3'. Lorsque la petite sous-unité du ribosome rencontre le premier codon d'initiation de la traduction AUG, elle s'arrête, recrute la grosse sous-unité ribosomale 60S et la traduction commence (Figure 5) (Aylett et Ban, 2017; Ramanathan et al., 2016). En raison de ce mécanisme d'initiation de la traduction, les ribosomes devraient en principe toujours commencer la traduction à partir du premier codon AUG. Cependant, le contexte dans lequel se

trouve ce codon, c'est-à-dire les nucléotides qui l'entourent, peut avoir un impact sur la reconnaissance du codon par le ribosome. En effet, on peut trouver sur l'ARNm au niveau du site d'initiation de la traduction une séquence nommée séquence de Kozak qui favorise l'initiation efficace de la traduction. La séquence peut varier selon les espèces, mais est en général assez bien conservée. Chez l'humain, cette séquence consensus est GCCGCC(A/G)(C/A)CAUGGCG (Kozak, 1987; Nakagawa, 2008). Afin que la séquence Kozak soit considérée forte, il faut, en plus de la présence essentielle du codon start AUG, une purine (A/G) en position -3 (3 nucléotides avant le A du codon start) et un G en +4 (3 nucléotides après le A du codon start) (Kozak, 1986). Ainsi, le ribosome peut "ignorer" un codon start AUG qui n'est pas dans un contexte favorable et privilégier l'initiation de la traduction à un codon AUG possédant une séquence de Kozak (Kozak, 1986). Ceci dit, contrairement aux ARNm polycistroniques des organismes procaryotes, les ARNm nucléaires eucaryotes ont jusqu'à tout récemment (voir plus bas) été décrits comme ne pouvant pas coder plusieurs protéines et donc considérés comme des molécules monocistroniques (Figure 4).

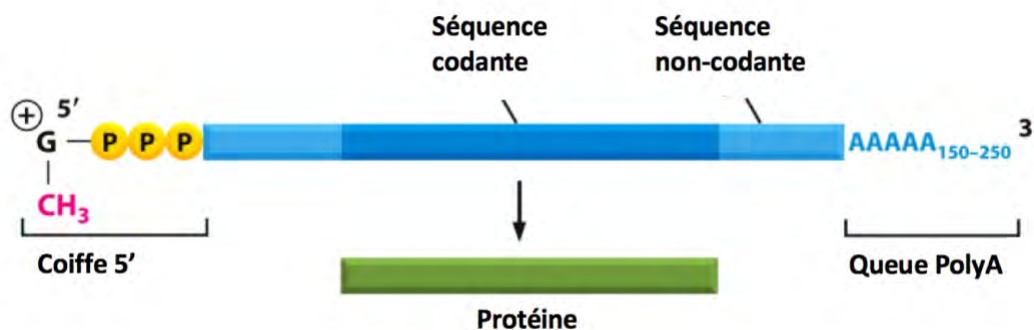


Figure 4. Coiffe en 5' et queue polyA en 3' d'un ARNm nucléaire eucaryote. La molécule est monocistronique. Figure tirée et modifiée de Alberts *et al.* (2011).

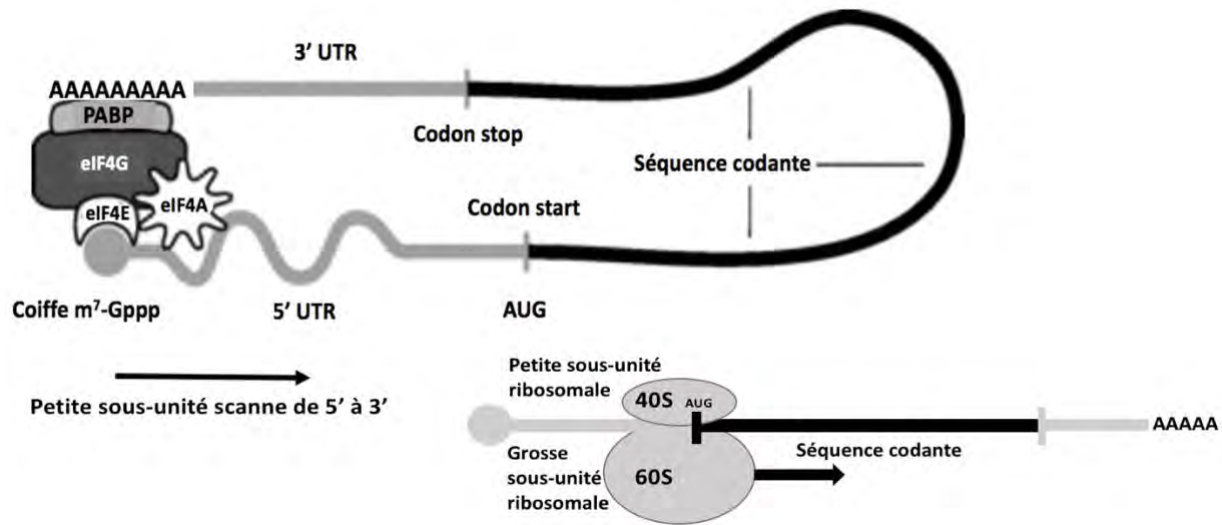


Figure 5. Circularisation de l'ARNm par l'interaction entre la coiffe en 5', les facteurs eIF4E et eIF4G et la protéines PABP liée à la queue polyA. La petite sous-unité ribosomale est recrutée à l'extrémité 5' de l'ARNm et scanne la molécule vers l'extrémité 3'. Une fois arrivée au codon AUG, la grosse sous-unité ribosomale est recrutée et la traduction commence. Figure tirée et modifiée de Kasinath *et al.* (2009).

1.3.3 Initiation de la traduction chez les eucaryotes : génome mitochondrial

Contrairement aux ARNm nucléaires, les ARNm mitochondriaux ne possèdent pas de coiffe à leur extrémité 5'. Cependant, chez les humains, ils sont tous polyadénylés en 3', sauf pour le gène *nd6* (D'Souza et Minczuk, 2018). Ce gène est d'ailleurs le seul qui code une protéine sur le brin antisens de l'ADNmt (Figure 1). Bien que les mitochondries soient d'anciennes bactéries, l'évolution rapide de leur génome a créé un système très divergent de leur ancêtre. En effet, les mécanismes d'initiation de la traduction sont très différents de ceux des organismes procaryotes ainsi que des ARNm nucléaires des organismes eucaryotes puisque les ARNm mitochondriaux ne possèdent pas de séquence Shine-Dalgarno et ne possèdent souvent presque aucun nucléotide en 5' avant le codon d'initiation de la traduction (Christian et Spremulli, 2012). Puisque les

mitochondries utilisent un code génétique légèrement différent du code génétique standard, les codons d'initiation de la traduction reconnus par les mitoribosomes des vertébrés sont AUG et AUA. Les mécanismes d'initiation de la traduction ne sont pas encore tout à fait bien compris, mais ils impliquent des facteurs d'initiation tels que mIF2 et mIF3. Le facteur mIF3 positionne le codon d'initiation de l'ARNm avec le site peptidyl (site P) de la petite sous-unité ribosomique, ce qui permet la traduction de l'ARNm (D'Souza et Minczuk, 2018). Les 13 protéines codées par le génome mitochondrial humain sont codées par 9 ARNm monocistroniques et 2 ARNm dicistroniques (un ARNm qui code pour deux protéines : ATP8 et ATP6, et un autre ARNm qui code pour deux autres protéines : ND4L et ND4). Il est intéressant de noter que ces ARNm dicistroniques possèdent des ORF qui se superposent de quelques nucléotides, c'est-à-dire qu'une partie du gène *atp8* (46 nucléotides) est retrouvée dans le gène *atp6* et une partie du gène *nd4l* (7 nucléotides) est retrouvée dans le gène *nd4* (Christian et Spremulli, 2012).

1.4 Annotation des génomes

1.4.1 ORF

Afin de comprendre comment les génomes sont annotés, il est important de comprendre ce que représente un ORF. ORF vient de « open reading frame » qui veut dire cadre de lecture ouvert. Un cadre de lecture ouvert est un cadre dans lequel est lue une séquence de nucléotides qui sera traduite en protéine. La lecture se fait par groupe de trois nucléotides (les codons), ce qui fait que chaque brin d'ADN ou d'ARN peut être lu selon trois cadres de lecture différents. Un ORF est donc simplement une séquence de nucléotides lue dans un certain cadre de lecture qui commence par un codon start et qui se termine par un codon stop et qui va être traduite en une protéine fonctionnelle. En raison du mécanisme d'initiation de leur traduction, les ARNm nucléaires

eucaryotes ont longtemps été considérés comme des molécules monocistroniques à 3 régions principales : la région 5'UTR (untranslated region) suivie d'un seul ORF (le plus long trouvé sur l'ARNm) qui code pour une protéine fonctionnelle, suivie par une région 3'UTR (Mouilleron et al., 2016). Lors de l'analyse de génomes complets, on remarque la présence d'une multitude de petits ORF de moins de 100 nucléotides. Ces ORF n'ont généralement pas été considérés dans l'annotation des génomes puisque leur taille est considérée comme trop petite pour que leur présence soit attribuée à autre chose qu'au hasard (Martinez et al., 2020). Ainsi, jusqu'à tout récemment (voir plus bas), seulement les ORF de plus de 100 nucléotides étaient considérés dans l'annotation des génomes et seulement un ORF était associé à chaque gène codant une protéine. Ces ORF annotés sont entre-autres appelés refORF pour ORF de référence ou refCDS pour "coding sequence of reference" (Mouilleron et al., 2016) (Figure 6).

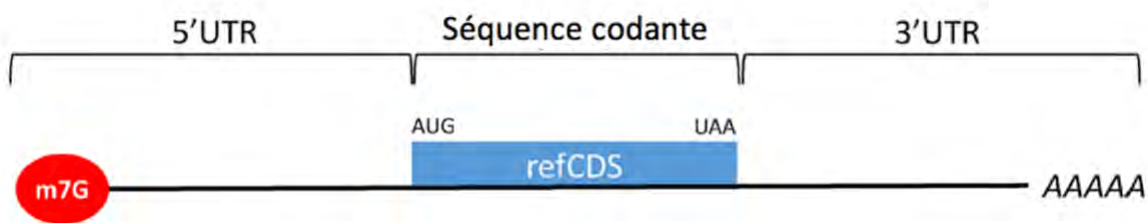


Figure 6. Structure d'un ARNm eucaryote typique en trois régions avec une seule séquence codante annotée (refCDS). Les régions 5'UTR et 3'UTR ne sont pas traduites. Figure tirée et modifiée de Mouilleron *et al.* (2015).

1.4.2 smORF et altORF

La vision fixe de la structure d'un ARNm et de la taille minimale que doit avoir un ORF afin d'être considéré comme codant suggère qu'une très grande portion des protéomes reste aujourd'hui cachée. En effet, pourquoi est-ce qu'un plus petit ORF ne pourrait pas coder une protéine ? De plus en plus d'études révèlent la présence de petites protéines codées par des ORF

de moins de 100 nucléotides. Ces smORF (pour small ORF) sont retrouvés dans plusieurs régions qui étaient considérées non-codantes ainsi que dans les régions non-traduites des ARNm (régions UTR) (Landry et al., 2015). En plus de ces smORF, d'autres types d'ORF ont été découverts et nommés altORF pour « alternative open reading frame ». Un altORF est un ORF qui peut chevaucher le refCDS ou bien se trouver à l'extérieur de la séquence et même sur le brin complémentaire supposé non-codant. Il est caractérisé par la présence de codons start et stop alternatifs à ceux de l'ORF principal. Un altORF peut être dans le même cadre de lecture que le refCDS ou bien dans un autre cadre de lecture (Figure 7). (Orr et al., 2020; Vanderperre et al., 2013). Les protéines codées par ces altORF sont appelées des protéines alternatives. La présence de telles protéines a surtout été décrite chez les virus et les bactériophages ainsi que chez les bactéries, et permettrait entre autres de minimiser la taille du génome afin d'accélérer la réplication (Normark et al., 1983; Orr et al., 2020).

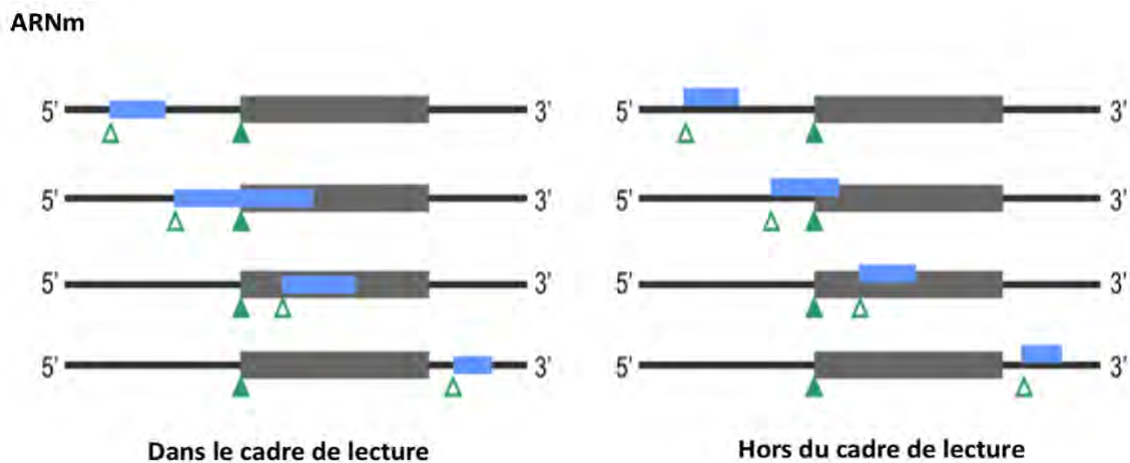
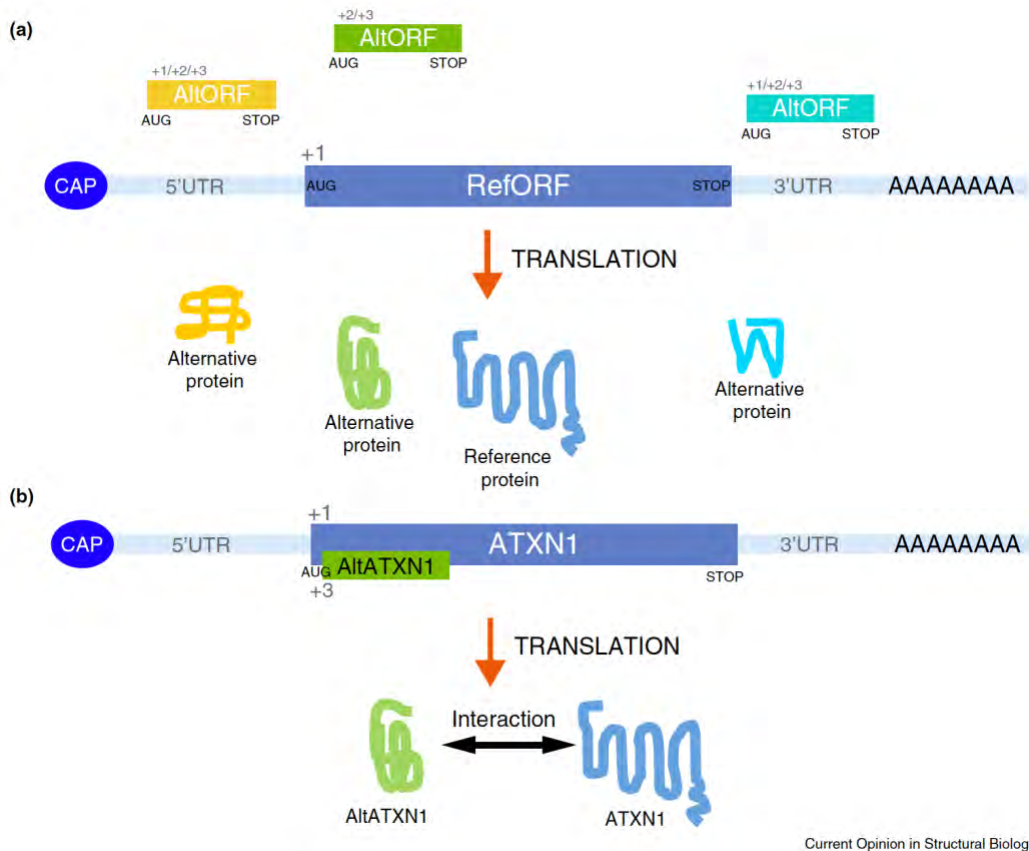


Figure 7. Organisation des altORF (en bleu) par rapport aux refCDS (en gris foncé) d'un ARNm. Les altORF sont dans le même cadre de lecture (à gauche) ou dans un autre cadre de lecture (à droite) que le refCDS. Les codons start des refCDS sont représentés par les triangles pleins et les codons start alternatifs par les triangles vides. Les altORF peuvent être dans les régions non-traduites des ARNm ou complètement à l'intérieur du refCDS ou encore partiellement à l'intérieur. Figure tirée et modifiée de Orr *et al.* (2020).

Ce protéome alternatif n'est cependant pas retrouvé uniquement chez les virus et les bactéries. En effet, malgré le fait que les ARNm nucléaires des eucaryotes ont longtemps été considérés comme des molécules monocistroniques en raison du mécanisme de sélection du site d'initiation de la traduction qui ne permet la lecture que d'un seul ORF, il a récemment été révélé que des altORF et smORF présents dans des ARNm nucléaires d'organismes eucaryotes, dont l'humain, peuvent être traduits en protéines fonctionnelles (Mouilleron et al., 2016; Martinez et al., 2020; Brunet et al., 2020). Dans plusieurs cas, les protéines alternatives et les protéines de référence interagissent ensemble. Ainsi, l'expression de deux ou plusieurs protéines à partir d'un même ARN pourrait être un mécanisme permettant d'assurer la co-expression de protéines qui interagissent entre-elles (Figure 8) (Mouilleron et al., 2016). Ces découvertes remettent en question les mécanismes d'initiation de la traduction connus chez les eucaryotes et montrent que le potentiel codant du génome nucléaire a été sous-estimé. Puisque le génome nucléaire contient des smORF et des altORF, il se pourrait que le même phénomène se produise dans l'ADNmt, ancien génome bactérien, et que plusieurs mtaltORF (altORF du génome de la mitochondrie) codent pour des protéines mitochondriales fonctionnelles encore inconnues.



Current Opinion in Structural Biology

Figure 8. (a) Représentation typique d'un ARNm eucaryote mature et ses altORF possibles. (b) Représentation du gène ATXN1 et de son altORF AltATXN1. AltATXN1 est dans un autre cadre de lecture que le refORF et il chevauche son côté N-terminal. Une fois les deux protéines traduites, la protéine alternative AltATXN1 interagit avec le domaine N-terminal de la protéine ATXN1. Figure tirée de Landry *et al.* (2015).

1.5 Des gènes codants supplémentaires dans les génomes mitochondriaux

Il y a une vingtaine d'années, des chercheurs ont découvert un nouveau petit peptide mitochondrial humain nommé Humanine (HN) dont la séquence codante se trouve à l'intérieur d'une autre séquence codante de l'ADNmt (Hashimoto *et al.*, 2001). Plus précisément, elle se trouve à l'intérieur du gène codant l'ARNr 16S de la mitochondrie (Hashimoto *et al.*, 2001). Suite à cette découverte, d'autres micropeptides mitochondriaux tels que MOTS-c et SHLP1-6 ont été

identifiés. Ces peptides sont codés respectivement dans les gènes ARNr 12S (MOTS-c), dans le gène ARNr 16S (SHLP6) ainsi que sur son brin complémentaire (SHLP1-5). Ces peptides exercent des fonctions importantes dans le métabolisme, la neuroprotection, l'apoptose et le vieillissement (Lee et al., 2015; Cobb et al., 2016). Un autre gène codant pour une protéine, mais dont la fonction demeure inconnue, a été identifié sur le brin complémentaire de Cox1 (Gau) de l'ADNmt humain (Figure 9) (Faure et al., 2011). La découverte de ces nouveaux peptides montre que tout comme le génome nucléaire, le potentiel codant du génome mitochondrial a été sous-estimé. Cela remet donc en question la complexité du génome mitochondrial ainsi que le dogme disant que l'ADNmt ne code que pour 13 protéines.

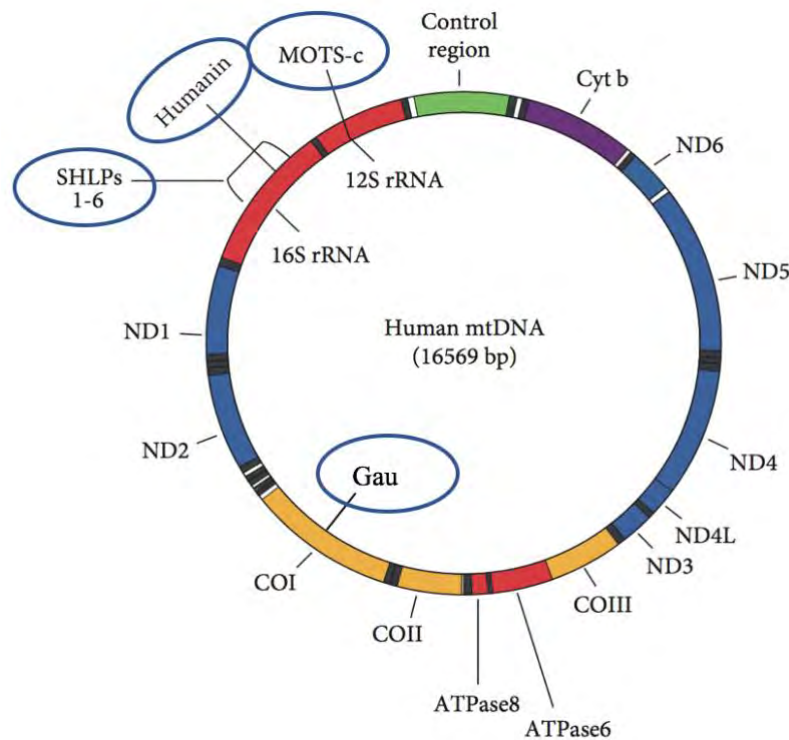


Figure 9. Emplacement des peptides Humanine, MOTS-c, SHLP1-6 et Gau dans l'ADNmt humain. Figure tirée de Minasyan *et al.* (2017).

Pour l'instant, les recherches afin d'identifier de nouvelles protéines alternatives dans le génome mitochondrial humain ont principalement été effectuées au niveau des ARNr 12S et 16S. Cependant, l'équipe de Sophie Breton et d'Annie Angers a effectué un examen approfondi du génome mitochondrial humain et a révélé la présence de 249 ORF de 60 nucléotides ou plus (Figure 10). Ces 249 ORF incluent les 13 protéines de la CTE et l'ATP synthase ainsi que Humanine, MOTS-c, SHLP1-6 et Gau, ce qui fait 227 nouveaux mtaltORF potentiellement traduits en protéines fonctionnelles (Angers et al., 2019).

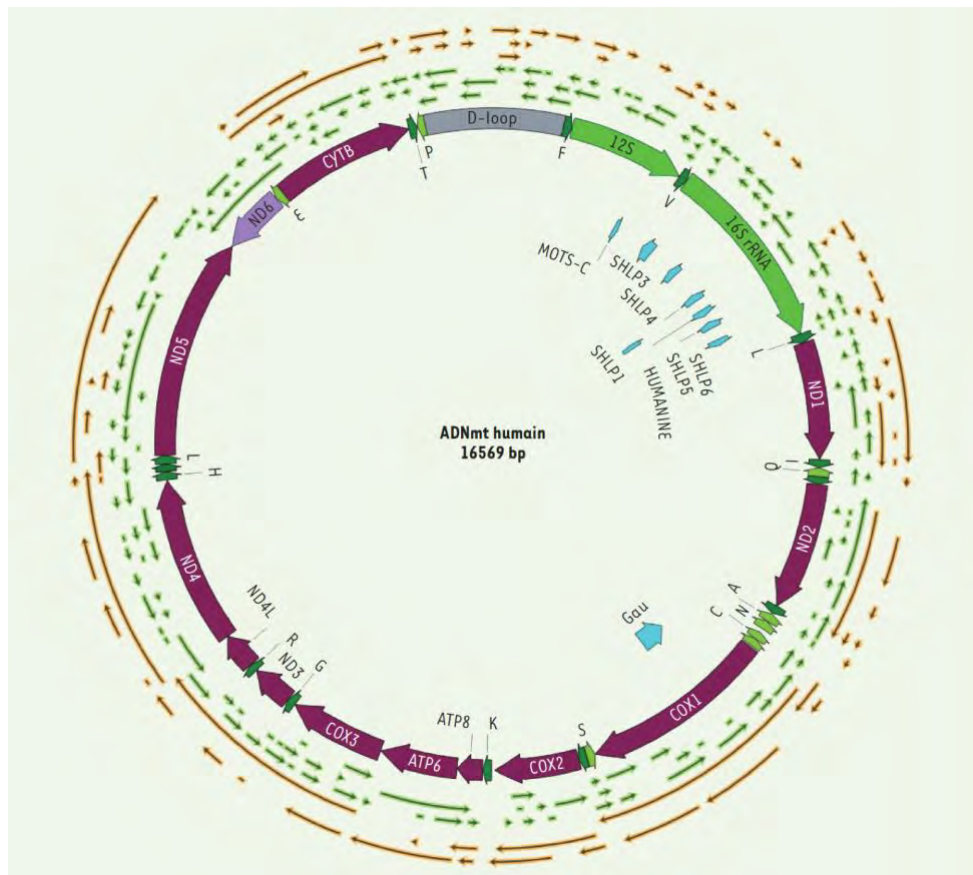


Figure 10. Carte du génome mitochondrial humain montrant les nouveaux cadres de lecture caractérisés (en bleu) et tous les cadres de lecture ouverts potentiels sur les brins sens (flèches orangées) et antisens (flèches vertes). Figure tirée de Angers *et al.* (2019).

1.6. Hypothèse et objectifs

L'hypothèse de ce projet de recherche est que d'autres peptides mitochondriaux humains fonctionnels sont codés par des séquences alternatives mitochondriales trouvées dans les gènes qui codent pour des protéines annotées et dans les régions non-codantes retrouvées notamment sur le brin complémentaire de l'ADNmt. L'objectif du projet de recherche est de déterminer l'ampleur de ce protéome alternatif et donc d'identifier des nouvelles protéines mitochondriales alternatives codées par certains des 227 ORF nouvellement caractérisés par Angers et al. (2019). Pour se faire, certaines séquences provenant des 227 ORF seront choisies et des anticorps dirigés contre ses séquences seront produits par immunisation de lapins. L'identification des nouvelles protéines mitochondriales potentielles se fera par immunobuvardage sur des cellules humaines HeLa et HEK293T. De plus, si de telles protéines sont identifiées, il faudra vérifier leur origine, c'est-à-dire si elles sont véritablement codées par le génome mitochondrial ou si elles sont plutôt codées par le génome nucléaire. En effet, en raison d'évènements d'intégration de fragments de séquences d'ADNmt dans le noyau, ce qu'on appelle des NUMTs, il est possible qu'une séquence mitochondriale soit transcrite par le noyau et traduite dans le cytoplasme plutôt que dans la mitochondrie. Afin de confirmer l'origine des protéines, d'autres expériences d'immunobuvardage seront réalisées sur des cellules dépourvues d'ADN mitochondrial, soit des cellules Rho-0. Finalement, des expériences d'immunofluorescence et d'immunoprécipitation seront réalisées afin de confirmer respectivement la localisation des protéines dans les cellules ainsi que leur séquence en acides aminés.

2. Article

The alternative human proteome of mitochondrial origin

Laura Kienzle¹, Marie Brunet², Xavier Roucou², Christian Landry³, Annie Angers¹, Sophie Breton^{1,*}

¹Department of Biological Sciences, University of Montreal, Montreal, Quebec, Canada

²Department of Biochemistry, University of Sherbrooke, Sherbrooke, Quebec, Canada

³Department of Biology, Laval University, Quebec, Quebec, Canada

Article en préparation

ABSTRACT

Small open reading frames (smORFs) and alternative open reading frames (altORFs) are emerging as a new class of important genes. Their discovery in the human genome shows that the proteome is more complex than previously anticipated. The study of these ORFs is important in order to get a better understanding of evolutionary processes, fundamental biological knowledge as well as new avenues for therapeutic discovery. These alternative proteins have also been identified in the human mitochondrial genome. These peptides include Humanin, MOTS-c, SHLP1-6 and Gau. A preliminary *in silico* examination of the human mitochondrial genome has revealed the presence of 227 unannotated ORFs of 60 nucleotides and more on both strands. In this study, we aimed to identify new human mtDNA-derived peptides from the 227 newly characterised ORFs. One protein, MTALTND4 has been identified by immunoblotting on HeLa and HEK293T cells. MTALTND4 is an alternative protein of 99 amino acids found inside the mitochondrial *nd4* gene. Immunofluorescence and immunoblotting experiments show that MTALTND4 is translated inside mitochondria and can be found in the cytoplasm as well as in plasma. More experiments are needed in order to understand the function of this novel protein. The discovery of this novel mtDNA-derived peptide confirms that the coding potential of the human mitochondrial genome is larger than anticipated and shows that many more mitochondrial alternative peptides have yet to be discovered.

INTRODUCTION

The recent identification in the human genome of thousands of unannotated translated small open reading frames (smORFs) and alternative ORFs (altORFs in reference genes) has brought to light a whole new class of important genes (Landry et al. 2015; Martinez et al. 2020). Although the elucidation of their function remains challenging, peptides derived from smORFs and altORFs reveal a new facet of eukaryotic proteomes, enabling a better understanding of key evolutionary processes such as the birth of novel genes and proteins, and providing novel fundamental biological knowledge as well as new avenues for therapeutic discovery. Mitochondria-derived proteomes should not escape this reality, given that the proteome of their bacterial ancestors has also been shown to be more complex than previously anticipated (Orr et al. 2020). Consistent with this idea, human mitochondrial alternative peptides have recently been identified. These peptides include Humanin, MOTS-c, SHLP1-6 and Gau whose coding sequences are respectively found inside the 16S rRNA, the 12S rRNA, the 16S rRNA and on its complementary strand and on the complementary strand of the *coxI* gene (Hashimoto et al. 2001; Lee et al. 2015; Cobb et al. 2016; Faure et al. 2011). Research aiming to identify the presence of mtDNA-derived peptides coded by altORFs or smORFs in other regions of the human mitochondrial DNA and on its complementary strand has yet to be done.

Despite the discovery of mtDNA-derived micropeptides, the vast majority of scientific papers and genetics textbooks still describe the human mtDNA as a small double-stranded circular molecule with only 37 genes, 13 of them coding for proteins involved in the production of energy, while the remaining 24 genes code for 22 transfer RNAs and 2 ribosomal RNAs necessary for the translation of the 13 protein coding genes (Taanman, 1999). Most of these genes (28 out of 37, representing approximately 95% of the coding DNA) are found on the same coding strand and are

separated by short non-coding regions, the complementary strand being non-coding (Anderson et al. 1981). Given its small size, the mitochondrial genome represents an excellent genetic system for a complete, comprehensive and accurate annotation of smORFs and altORFs. A preliminary *in silico* examination of its unresolved protein-coding potential has revealed the presence of 227 unannotated ORFs of >60 nucleotides on both strands (Angers et al. 2019) (**Fig. 1**).

This study aims to identify new human mtDNA-derived peptides from the 227 newly characterised ORFs to get a better understanding of the complexity of the mitochondrial genome and the contribution of these peptides to the mitochondrial proteome.

MATERIALS & METHODS

Identification of mitochondrial smORFs and altORFs and choice of candidates for antibody production

Three different approaches were used to search for potential smORFs and altORFs in the human mtDNA and to choose ORF-derived peptide sequences against which antibodies would be generated. First, an *in silico* search was performed for ORFs on both strands encoding peptides >20 amino acids using the vertebrate mitochondrial genetic code (Angers et al. 2019). However, instead of considering the previously designated stop codons AGA and AGG – which conventionally code for arginine but the human mtDNA does not possess a tRNA^{ARG} for these triplets – we considered that human mitochondria use only UAA and UAG as stop codons as recently suggested (Temperley et al. 2010; Ayyub et al. 2020), with the possible import of the cytosolic AGR tRNA into the mitochondria and/or the expression of mt antisense tRNAs with anticodons recognizing AGR codons according to an alternative genetic code that is induced by

the presence of suppressor tRNAs (Rubio et al. 2008 ; Faure et al. 2011). This analysis returned 249 sequences encoding 20 to 600 amino-acid-long peptides (the human mtDNA reference sequence was used; Angers et al. 2019). Of these sequences (i.e. 227 unannotated ORFs + the 13 typical protein-coding genes + humanin, Mots-c, gau and SHLPs 1-6), 14 had a minimal conserved Kozak sequence known to be favorable for efficient translation initiation (Harhay et al. 2005; Lee et al. 2015), i.e. with an A or G at position -3 preceding the initiation codon and a G at position +4 (table S2). Of these 14 smORFs or altORFs, one was SHLP2, a small humanin-like peptide encoded in the mtDNA 16S ribosomal RNA region identified by Cobb et al. (2016), and six were conserved outside the Homo/Pan (chimpanzees and bonobos)/Gorilla group, but two of them were too hydrophobic for immunisation according to MédiMabs (McGill University). We decided to generate an antibody against the one candidate with the lowest number of hits showing >70% similarity with sequences of the human nuclear genome (tables S1 and S2).

In the second approach, we interrogated the OpenProt database (Brunet et al. 2019). OpenProt is the first database that enforces a polycistronic model of eukaryotic genome annotations. Specifically, OpenProt retrieves transcripts from two well-used annotations (Ensembl and NCBI RefSeq), which constitutes an exhaustive transcriptome. A three-frame *in silico* translation then yields the ORFeome: any ORF longer than 30 codons in any frame of any transcript. This ORFeome is then filtered to categorize predicted ORFs. The first filter retrieves all known protein (all ORF already annotated in Ensembl, NCBI RefSeq, and/or UniProtKB), these are the RefProts. The second filter looks at the homology of the currently not annotated ORFs with the RefProt of the same gene (if applicable) and retrieves novel predicted isoforms. The remaining ORFs encode novel proteins, called AltProts (Brunet et al. 2019). For mitochondrial transcripts, OpenProt uses the vertebrate mitochondrial genetic code (with AGA and AGG considered as stop

codons). To increase confidence in ORF expression, OpenProt also cumulates several lines of evidence, such as (i) conservation evidence (for every ORF annotated, OpenProt identifies orthologs and paralogs across the 10 species currently supported by OpenProt), (ii) translation evidence (OpenProt retrieves publicly available ribosome profiling datasets and re-analyse them using the Price algorithm with a stringent 1 % FDR), and (iii) expression evidence (OpenProt retrieves publicly available mass spectrometry datasets and re-analyse them using multiple search engines, and a stringent 0,001 % FDR) (Brunet et al. 2019). The search in OpenProt returned 14 altORFs encoding 29-61 amino-acid-long peptides, eight of them were found within the 12S and 16S ribosomal genes and six in protein-coding genes (table S3). We choose to focus on these latter six, which were all conserved in *Pan troglodytes* except for one altORF. Moreover, all these six altORFs generated very few hits >70% similarity (i.e. less than 4) when searched against the nuclear genome (data not shown). We chose four of these altORFs for antibody production (table S3).

In the third approach, we interrogated mitochondria-derived mass spectrometry datasets to see if peptide sequences resulting from our approach 1 could be detected. Specifically, we used the novel tool PepQuery to interrogate spectrums from published datasets. This approach compares spectrum annotation with peptides from our novel sequences to annotation with any protein from the reference database (known proteins). The chosen dataset was CPTAC_Breast_Cancer as it is one of the largest in terms of MS/MS scans with more than 32 millions. We first identified all possible peptides from our 249 proteins (see approach 1). We realized an *in silico* digestion with Trypsin, allowing up to 2 miscleavages, a minimum peptide length of 7 amino acids and a mass from 400 to 6000 Da to mimic experimental conditions. Then each peptide was queried using the PepQuery tool (see Fig. S1).

From the PepQuery tool, we retrieved any Peptide Spectrum Match (PSM) that hits better with our queried peptide than with any protein from the reference database. Various filters were used to validate the results as confident or not. For a PSM to be considered as confident, a p value under 0.01 and a score higher than anyhow modified peptides from the reference database were needed. Using these criteria, we could be confident this spectrum corresponds to our queried peptide and not to a reference peptide with any PTM (post-translational modification). As such, our supplementary table S4 contains for each of the 249 proteins (the 227 novel predicted proteins, the 13 RefProts for positive control, Humanin, MOTS-c, SHLP1-6 and Gau) the number of PSMs with higher score than a referenced peptide, the number of confident PSMs, the best confident PSM (score), its corresponding p value, and its total mass. This analysis returned 46 protein sequences, 10 of which corresponded to RefProts and 36 to putative novel mitochondrial-derived proteins. Antibodies have been generated against 4 of these putative proteins (tables S1 and S4).

Analyses of MTALTND4 sequences

To test for the presence of nuclear mtDNA sequences (NUMTs) identical or very similar to MTALTND4 due to recent mitochondrial transfer (Ricchetti et al. 2004), we used the NCBI nucleotide basic local alignment search tool (BLASTN) (Altschul et al. 1997) against the human reference genome (Fig. S2). The degree of conservation of MTALTND4 was assessed by aligning the protein sequences from primate and mammal species obtained from the NCBI database. TBLASTn searches were used to ensure correct extraction of the protein sequences, which were aligned using ClustalW Multiple Alignment (Thompson et al. 1994) (Fig. S3). Putative transmembrane (TM) helices were searched using TMPred (Hofmann and Stoffel, 1993). Putative post-translational modifications (PTM) of MTALTND4 were searched using some of the PTM

prediction webservers recently reviewed by Audagnotto and Dal Peraro 2017, only scores with 80% probabilities were retained.

Rabbit anti-MTALTND4 antibody generation

Custom rabbit anti-MTALTND4 was ordered from MediMabs (and also our other antibodies; see Fig. 1 and table S1). High-titer polyclonal anti-sera against MTALTND4 was obtained.

Cell culture

HeLa and HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% calf bovine serum, penicillin, streptomycin and fungizone and were kept at 37°C with 5% CO₂. Chloramphenicol-treated HeLa cells were produced by adding 2ml of 50µg/ml chloramphenicol in a 10cm petri dish of 80% cell confluency. Cells were treated for 48 hours before being harvested for western blot analysis (Tian et al. 2016). Ethidium Bromide-treated HeLa- ρ 0 cell lysates (vs. control) were obtained from Abcam (ab154479).

Human plasma

Pooled human plasma (blood derived) was obtained from Innovative Research (IPLASK2E2ML). The whole blood is collected from donors in an FDA-approved collection center and processed into plasma via centrifugation. The plasma is frozen immediately after processing. Anticoagulant K2 EDTA was used to collect this product.

Western blotting

HeLa and HEK293T cells were washed in PBS, harvested and lysed with a sonic dismembrator sonicator (Fisher) in a lysing buffer containing 10mM HEPES, 150mM NaCl and a cocktail of protease inhibitors (benzamidine, PMSF, aprotinin and leupeptin). Triton X-100 10% was added after sonication. Total protein concentration was estimated using the Bradford dosage method. The samples were then mixed with a Laemmli buffer containing DTT and β -mercaptoethanol reducing agents and heated at 95°C for 5 minutes before being loaded at equal protein concentration (100 μ g) into a polyacrylamide gel. A tris-tricine SDS PAGE electrophoresis with a gel composed of three parts (separating, spacer and compaction gels) was made as suggested by Schagger and von Jagow (1987) to ease the separation and visualization of small proteins compared to the more common glycine SDS PAGE electrophoresis gel composed of two parts. To maximize the separation of low molecular weight proteins, the migration was first done for an hour at low voltage (30V). Then the voltage was increased to 43V, and the migration carried out overnight at room temperature. For the transfer, PVDF membranes of 2 μ m were used which allows better retention of small proteins. The transfer was done for one hour at 1000 mA at 4°C in a transfer buffer containing 194 mM glycine, 25 mM Tris-base and 20% methanol in H₂O. Before incubation with the primary antibodies, membranes were blocked for 30 minutes at room temperature in a blocking buffer containing 5% milk powder and 0.05% tween-20 diluted in PBS. Primary antibodies rabbit anti-MTALTND4 (1:1000), mouse anti-ATP5 (1:1000; Abcam [ab14748]), mouse anti-Cox1 (1:500; Abcam) and mouse anti-actin (1:2000; Abcam [ab14705]) were diluted in a PBS + 0.05% tween-20 solution. Membranes were incubated with the primary antibodies for two hours at room temperature. After incubation, membranes were washed 3 X 5 minutes in TBS-T at room temperature. Secondary antibodies goat anti-rabbit IgG (1:2000) and

goat anti-mouse IgG (1:2000) coupled to the horseradish peroxidase (HRP) were diluted in a PBS + 0.05% tween-20 solution. Membranes were incubated with secondary antibodies for 30 minutes at room temperature. Finally, membranes were washed 3 X 10 minutes with TBS-T and 1 X 5 minutes with TBS at room temperature. Protein signals were visualized by adding Montreal Biotech Inc.'s enhancer and substrate solutions and images were captured with a FUSION FX chemiluminescence imaging system.

Deglycosylation

Deglycosylation was done using NEB's protein deglycosylation mix II (#P6044S) and following the manufacturer's recommended protocol. HeLa cells were washed in PBS, harvested and lysed with a sonic dismembrator sonicator (Fisher) in a lysing buffer (10mM HEPES, 150mM NaCl and a cocktail of protease inhibitors: benzamidine, PMSF, aprotinin and leupeptin). 5 μ l of Mix Buffer 2 was added to 40 μ l of sample before incubating at 75°C for 10 minutes. After cooling down, 5 μ l of Mix II was added and incubated at room temperature (25°C) for 30 minutes. Samples were then incubated at 37°C for 1 hour. The loading buffer was then added for migration on SDS-PAGE followed by immunoblotting. Fetuin was used as a positive control and was visualized by coomassie blue staining.

Dephosphorylation

Dephosphorylation was done using NEB's Lambda Protein Phosphatase kit (#P0753S) and following the manufacturer's recommended protocol. HeLa cells were washed in PBS, harvested and lysed with a sonic dismembrator sonicator (Fisher) in a lysing buffer (10mM HEPES, 150mM NaCl and a cocktail of protease inhibitors: benzamidine, PMSF, aprotinin and leupeptin). 5 μ L of

10X NEBuffer for Protein MetalloPhosphatases (PMP) and 5 μ L of 10 mM MnCl₂ were added to 40 μ l of samples to make a total reaction volume of 50 μ l. 1 μ l of Lambda Protein Phosphatase was added before incubating at 30°C for 30 minutes. The loading buffer was then added for migration on SDS-PAGE followed by immunoblotting.

Immunofluorescence

HeLa cells were cultured on a 8-well PCA detachable microscope slide in DMEM supplemented with 10% calf bovine serum, fungizone, penicillin and streptomycin and were kept at 37°C with 5% CO₂. Each chamber was washed with PBS and fixed with 4% paraformaldehyde for 20 minutes. Cells were then washed 3 X 5 minutes in PBS before their membranes were permeabilized by adding 0.2% Triton for 4 minutes. Cells were washed 3 X 5 minutes with PBS and then blocked with 5% NGS in PBS for 30 minutes at room temperature before incubating for 1 hour with primary antibodies rabbit anti-MTALTND4 antibody (1:50) and mouse anti-ATP5 antibody (1:250; Abcam) diluted in 1% NGS in PBS. Cells were washed 3 X 5 minutes in PBS before incubating with the secondary antibodies (Alexa Fluor™ 488 goat anti-rabbit IgG [1:500] and Alexa Fluor™ 594 goat anti-mouse IgG [1:500]) in PBS containing 1% NGS for 1 hour at room temperature. Finally, cells were washed 3 X 5 minutes in PBS before adding a mounting drop containing DAPI and visualized with an EVOS M5000 microscope.

Immunoprecipitation and mass spectrometry analysis

HeLa cells were washed with PBS and harvested in a lysing buffer containing 10mM HEPES, 150mM NaCl and a cocktail of protease inhibitors (benzamidine, PMSF, aprotinin and leupeptin). Cells were then lysed with a sonic dismembrator sonicator (Fisher) 3 times for 3

seconds before adding 100µl of Triton X-100 10%. Cell lysates were incubated on ice for 20 minutes and then centrifuged for 15 minutes at 15000 RPM at 4°C. The supernatant was kept. 20µl of protein A agarose was added to 1ml of samples and incubated at 4°C for 1 hour. Samples were then centrifuged at 15000 rpm for 1 minute at 4°C and the supernatant was kept. 20µl of rabbit anti-MTALTND4 antibody and 20µl of protein A agarose were added to the samples before incubating at 4°C overnight. A control was also made by adding 20µl of rabbit pre-immune serum instead of rabbit anti-MTALTND4 antibodies. The next day, samples were washed 6 times by centrifuging at 15000 RPM for one minute, discarding the supernatant, adding some lysing buffer and centrifuging again. Samples were then given to the proteomics platform of the Research Institute of the McGill University Health Center (MUHC Research Institute) for mass spectrometry analysis. Data were analyzed with Mascot (Matrix Science) against a human database and a final report was generated using Scaffold5 (Proteome Software).

RESULTS & DISCUSSION

To identify potentially functional smORFs or altORFs in the human mtDNA, antibodies were generated against nine unannotated ORFs (**Fig. 1** and table S1), which were chosen using three different approaches: (i) we searched among the 227 unannotated ORFs of >60 nucleotides previously identified (Angers et al. 2019) for those having a minimal conserved Kozak sequence known to be favorable for efficient translation initiation (Harhay et al. 2005; Lee et al. 2015) (table S2), (ii) we interrogated the OpenProt database (table S3), which enforces a polycistronic model of eukaryotic genome annotations (Brunet et al. 2019), and (iii) we interrogated mitochondria-derived mass spectrometry datasets to see if peptides derived from our 227 predicted sequences (approach i) could be detected (table S4 and Fig. S1). These three approaches respectively returned 13, 14 and 36 candidates and we respectively chose 1, 4 and 4 of these candidates for antibody production (**Fig. 1** and tables S1-S4). Immunoblotting was used to assess their expression in human HeLa and HEK293T cells.

Despite multiple attempts, we were unable to obtain a western blot signal of expected size (or a disappearance of the signal in Rho-0 cells devoided of endogenous mtDNA) for eight of the nine antibodies (data not shown). Overall, only one unannotated ORF-derived peptide, which was identified in our mass spectrometry datasets interrogation approach (see tables S1 and S4), was clearly detected by western blotting (**Fig. 2C**) and is referred to as MTALTND4 (Mitochondrial Alternative ND4 protein) because the ORF sequence is found in the coding sequence of the mitochondrial *nd4* gene (**Fig. 1**).

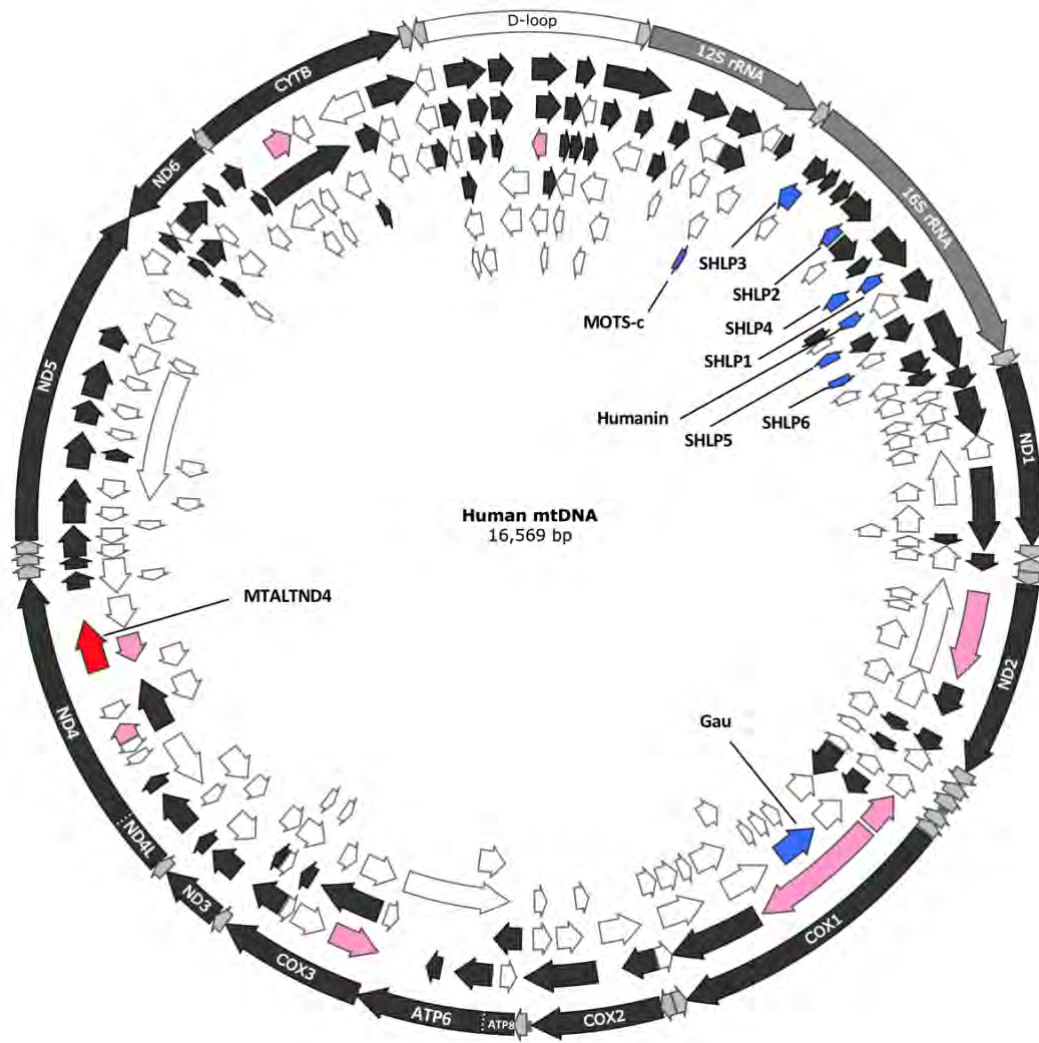


Figure 1. Map of the human mitochondrial genome. All 37 typical genes are represented on the outside circle of the map. New characterized alternative mitochondrial peptides such as Humanin, MOTS-c, SHLP1-6 and Gau are represented by blue arrows. 227 unannotated ORFs of >60 nucleotides are represented by black, white, pink and red arrows. ORFs in frames 1, 2 and 3 are represented by black arrows, ORFs in frames -1, -2 and -3 are represented by white arrows. The 9 ORFs chosen for this study are represented by pink arrows and the red arrow that codes for MTALTND4 alternative protein. Image was generated using SnapGene® software.

The altORF within the human mitochondrial *nd4* gene consists of 300 base pairs that translates into a 99 amino-acid peptide using the mitochondrial vertebrate genetic code and considering that human mitochondria use only UAA and UAG as stop codons as recently suggested (Temperley et al. 2010; Ayyub et al. 2020) (**Fig. 2A**). AGA and AGG, which

conventionally code for arginine, were thought to be recognized as termination codons in vertebrate mitochondria because of the absence of a mtDNA-encoded tRNA^{ARG} for these triplet codons, but it has now been shown that this is not the case at least in humans (Temperley et al. 2010; Ayyub et al. 2020). The only two human mitochondrial genes with AGA and AGG as putative terminators are *cox1* and *nd6* and it has been demonstrated that their two mRNAs would rather terminate at a UAG stop codon created as a result of a -1 frameshift of the mitoribosome (Temperley et al. 2010; Ayyub et al. 2020). With a similar situation for *mtaltnd4*, and knowing that there is no alternative start codon upstream the one that we first identified, the translated MTALTND4 peptide would be 59 amino-acids long with an expected size of 6,9 kDa. Alternatively, with AGA and AGG coding and a translation occurring either within mitochondria, as suggested for some of the SHLPs that contain putative AGA and AGG coding triplets (Cobb et al. 2016), or in the cytoplasm, as suggested for Mots-c because of its tandem ATGAGG start and “putative stop” codons using the typical mitochondrial genetic code (Lee et al. 2015), MTALTND4 would be 99 amino-acids long with an expected size of 11,5 kDa. Multiple peptide sequence alignments suggest that it is highly conserved in primates in general (**Fig. 2B** and Fig. S3). Our western blot results also indicate that *mtaltnd4* is encoded and translated within the mitochondrion because of the disappearance of the signal in Rho-0 cells as well as in HeLa cells after mitochondrial protein synthesis was inhibited with chloramphenicol (**Fig. 2C-E**), limiting the possibility of a nuclear origin due to a nuclear mtDNA transfer (NUMT; Richetti et al. 2004) (Fig. S2). Furthermore, the western blot signal, which was somewhat higher than expected (around 25 kDa) (**Fig. 2C**), suggest that the unassigned AGA and AGG codons may be coding in human mitochondria. This is in accordance with some of the SHLPs (Cobb et al. 2016) and would imply a possible import of the cytosolic tRNA^{ARG} into the mitochondria, an innate ability of mammalian

mitochondria (Rubio et al. 2008) and/or the expression of mt antisense tRNAs with anticodons recognizing AGR codons according to an alternative genetic code that is induced by the presence of suppressor tRNAs (Faure et al. 2011). This would also parsimoniously explain the previously reported intra-mitochondrial localization of endogenous MOTS-c (Lee et al. 2015; Kim et al. 2018), i.e. with a mitochondrial translation of the peptide instead of the exportation of its mRNA outside mitochondria for cytoplasmic translation followed by the importation of the peptide inside mitochondria. Although both possibilities are not mutually exclusive, further experiments, such as a treatment of cultured cells with chloramphenicol, will be needed to test this hypothesis.

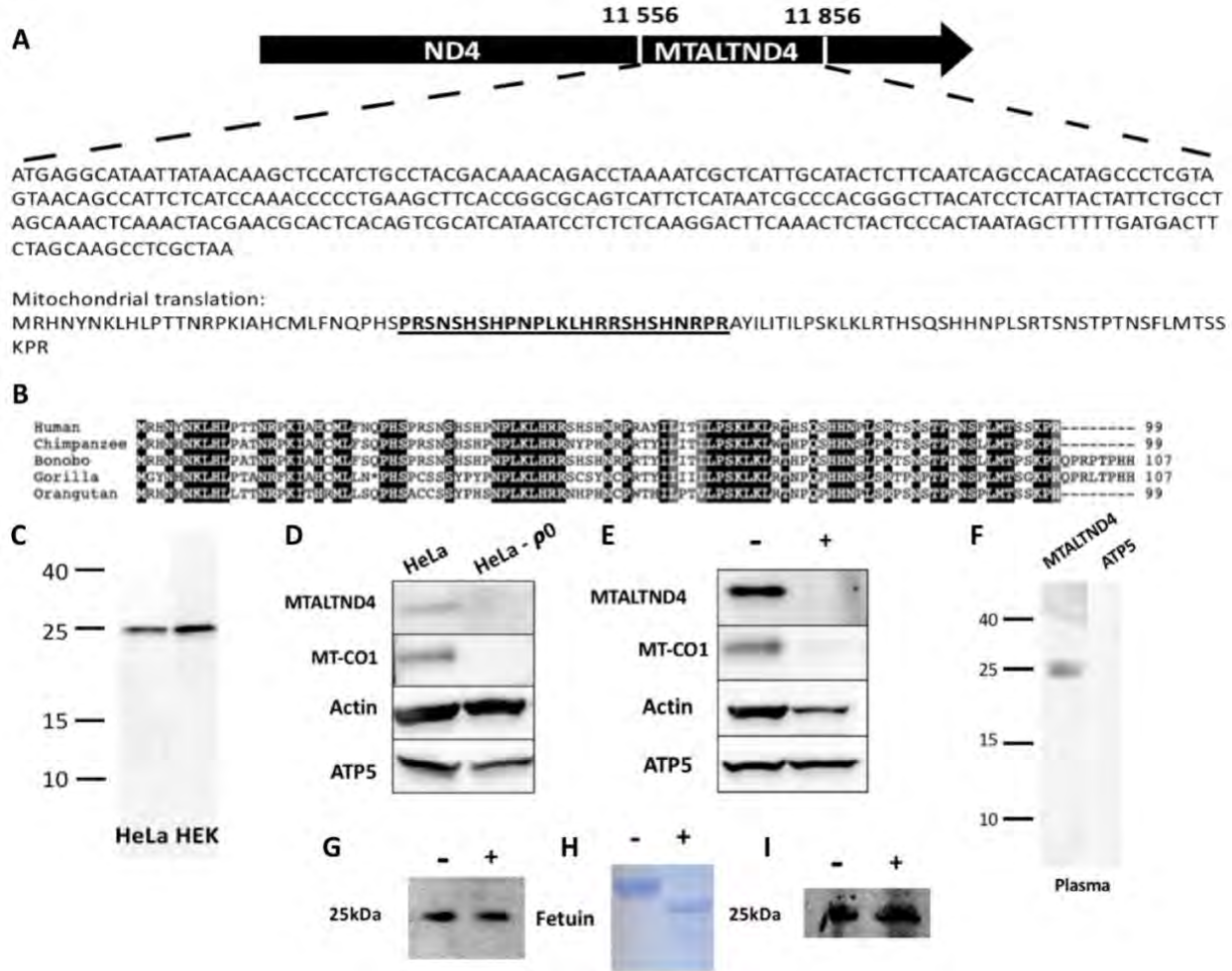


Figure 2. Identification of a novel altORF in the *nd4* gene of the human mitochondrial genome

- (A) Location of *mtaltnd4* ORF in the *nd4* gene and its translation in mitochondria considering AGA and AGG codons as coding. The antigenic sequence used for antibody production is in bold and underlined.
- (B) Peptide sequence alignment in 5 primate species: human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*).
- (C) Immunoblotting of MTALTND4 on HeLa and HEK293T cells. The apparent molecular weight is 25kDa.
- (D) Immunoblotting of MTALTND4, mitochondrial-encoded cytochrome C oxidase 1 and nuclear-encoded actin and ATP5 on normal HeLa cells and HeLa-*p0* cells
- (E) Immunoblotting of MTALTND4, mitochondrial-encoded cytochrome C oxidase 1 and nuclear-encoded actin and ATP5 on normal HeLa cells (-) and HeLa cells treated with chloramphenicol antibiotic (+).
- (F) Immunoblotting of MTALTND4 and ATP5 on human plasma. MTALTND4 appears at 25 kDa and ATP5 is absent.
- (G) Immunoblotting of MTALTND4 on non-treated HeLa cells (-) and HeLa cells treated with deglycosylation enzymes (+).
- (H) Coomassie Blue staining of non-treated Fetuin (-) and Fetuin treated with deglycosylation enzymes (+).
- (I) Immunoblotting of MTALTND4 on non-treated HeLa cells (-) and HeLa cells treated with Lambda phosphatase (+).

There are many possible reasons why the molecular weight of MTALTND4 is higher than the expected one. Since there is no predicted transmembrane helix in the protein that could affect the resolution and outcome of the western blots (e.g. Tsuji et al. 2020), the difference could, for example, be attributed to posttranslational modifications such as glycosylation and phosphorylation, or to SDS-resistant protein dimers (see Ouimet et al. 2020 for an extensive summary). Glycosylation events have been ruled out since treatment with enzymes that remove N-linked and O-linked oligosaccharides did not modify the apparent molecular weight of MTALTND4 but did have an effect on the positive control fetuin (**Fig. 2G-H**). As for phosphorylation, treatment with protein phosphatase Lambda did not modify the apparent molecular weight of MTALTND4, suggesting that the protein is not phosphorylated, but because no positive control for phosphatase Lambda was obtained, we cannot rule out the possibility of a phosphorylation event (**Fig. 2I**). It is also possible that the higher molecular weight could be attributed to the formation of a dimer resistant to SDS (Ettinger et al., 1998; Gentile et al., 2002), a programmed ribosomal frameshift that causes the elongation of the peptide (Dinman, 2006) or other untested postranslational modifications.

The mitochondrial localization of MTALTND4 in HeLa cells was confirmed by immunofluorescence (**Fig. 3**), with a certain degree of extra-mitochondrial localization of the peptide (**Fig. 3A**), suggesting that the peptide could be exported out of the mitochondria as shown also for humanin, MOTs-c and SHLPs (Hashimoto et al. 2001; Lee et al. 2015; Cobb et al. 2016). The absence of signal in HeLa-chloramphenicol treated cells confirmed again the mitochondrial origin of the alternative peptide (**Fig. 3B**). Since MTALTND4 could be exported outside the mitochondria and because some of the previously identified mtDNA-derived peptides such as Humanin, MOTS-c and SHLP2 have been found circulating in human plasma (Voigt et Jelinek,

2016; Lee et al., 2015; Cobb et al., 2016), immunoblotting experiments of MTALTND4 have been done on human plasma and revealed the presence of the protein but not the presence of the mitochondrial protein ATP5, which is a protein of the complex V that is also devoid of predicted transmembrane portion (**Fig. 2F**).

To confirm the identity of the signals on the western blots and to identify potential proteins that interact with MTALTND4, HeLa cells were immunoprecipitated and sent to the proteomics platform of the MUHC for MS analysis. The sequence of MTALTND4 was detected and confirms that the signal observed on the western blots corresponds to MTALTND4. More precisely, 5 unique peptide count representing a 21% coverage of the MTALTND4 sequence were detected (table S5). Other proteins were also found. These proteins could be potential interactive partners of MTALTND4 (table S5). Among these proteins the one that had the higher number of peptide count was glutathione peroxidase (GPx-1) an antioxidant enzyme that reduces hydrogen peroxide in water to limit oxidative damage to the cell (Lubos et al., 2011). More experiments such as Pull-down assays need to be done in order to confirm that this protein interacts with MTALTND4.

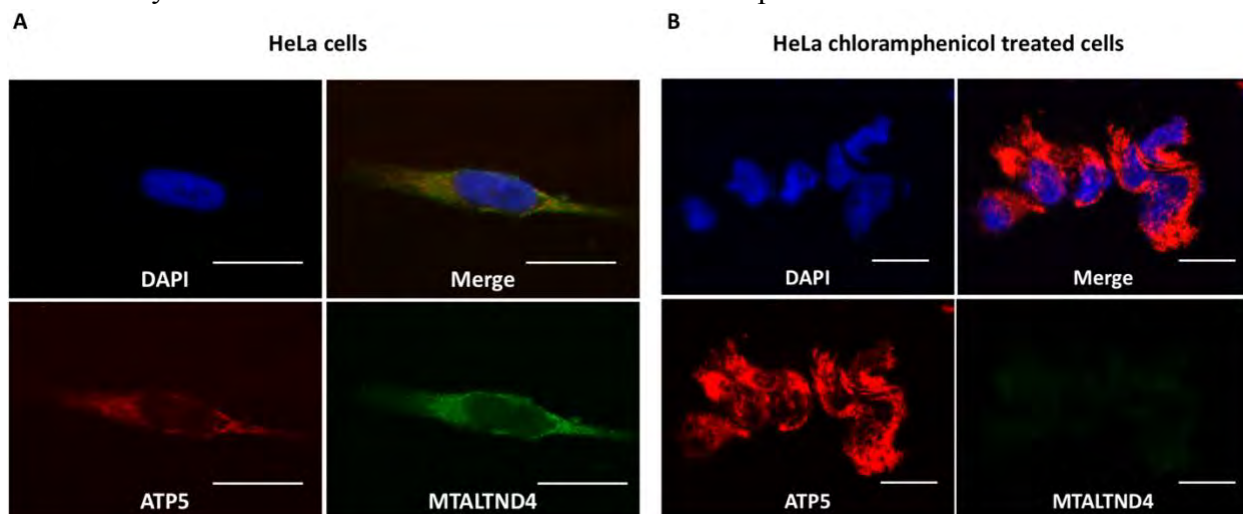


Figure 3. Localization of MTALTND4 by immunofluorescence

(A) Nuclear marker DAPI (blue), nuclear-encoded mitochondrial marker ATP5 (red) and MTALTND4 (green) by immunofluorescence in HeLa cells. Scale bar, 20 μ m.

(B) Nuclear marker DAPI (blue), nuclear-encoded mitochondrial marker ATP5 (red) MTALTND4 (green) by immunofluorescence in HeLa cells treated with chloramphenicol after 48 hours. Scale bar, 20 μ m.

To summarize, we have found a novel human mitochondrial alternative peptide of 99 amino acids whose sequence resides inside the *nd4* gene. This protein, MTALTND4, is translated inside mitochondria and can be found in the cytoplasm as well as in plasma and is well conserved in 5 other primate species. Western blot experiments show that the detected protein is migrating at approximately two times the expected molecular mass. This difference is not attributed to glycosylation and phosphorylation events but could be the result of the formation of a SDS resistant dimer or other PTMs. Finally, MS analysis of immunoprecipitated samples confirmed the presence of MTALTND4 and gave us an idea of putative interacting partners of MTALTND4 such as GPx-1. Pull-down assays need to be done to confirm this result. MTALTND4 is the first human mitochondrial alternative peptide discovered whose coding sequence is found inside the coding sequence of an annotated protein. This discovery shows that the mitochondrial proteome is more complex and larger than anticipated. Out of the 218 other newly characterised ORFs in the human mitochondria it is very probable that some, if not many, code for alternative proteins. Future researches need to be done to discover these peptides that could play major roles in cellular functions. The discovery of alternative peptides also calls into question the evolution of the mitochondrial genome as well as the selection pressures exerted on the mitochondria and the mechanisms allowing the translation of these alternative proteins. Future experiments on MTALTND4 will be done in order to get an understanding of its function. These experiments include the study of exogenous and endogenous effects of MTALTND4 on cell apoptosis, growth and gene expression, the analysis of MTALTND4 expression in different human and primate tissues and the analysis of MTALTND4 expression in cells exposed to different conditions such as hypoxia, anoxia, apoptosis and fasting.

3. Discussion

L'objectif de ce projet était de prouver l'existence d'un protéome mitochondrial alternatif chez l'humain plus vaste que ce que l'on pensait, donc d'identifier de nouvelles protéines mitochondriales codées par des altORF et des smORF. Ces séquences étaient principalement connues chez les virus et les bactéries, mais de récentes études révèlent de plus en plus de smORF et d'altORF codants dans les génomes nucléaires des organismes eucaryotes. Ces découvertes incitent à penser que de telles séquences sont aussi retrouvées dans les génomes mitochondriaux. En effet, plusieurs protéines dérivées du génome mitochondrial humain ont été découvertes dans les dernières années. Parmi celles-ci, on retrouve Humanine, MOTS-c, SHLPs1-6 et Gau. Les séquences de ces protéines sont trouvées à l'intérieur des gènes ARNr 16S (Humanine et SHLPs), ARNr 12S (MOTS-c) et sur le brin complémentaire du gène *cox1*(Gau). Aucune autre étude n'a été faite sur d'autres régions du génome mitochondrial humain. Une recherche approfondie *in silico* du génome mitochondrial humain effectuée par Angers et al. (2019) a révélé la présence de 227 ORF de 60 nucléotides et plus non annotés. L'hypothèse de notre projet de recherche était que parmi ces 227 ORF, certains codent pour des protéines mitochondriales fonctionnelles et que le protéome alternatif mitochondrial humain est beaucoup plus vaste que ce que l'on croyait. Afin de tester cette hypothèse, nous avons produit des anticorps dirigés contre 9 de ces 227 ORF et réalisé des western blot sur des cellules HeLa et HEK293T afin d'identifier des protéines mitochondriales potentielles.

3.1 Identification d'une protéine alternative par immunobuvardage

La détection de protéines par immunobuvardage (ou la technique de western blot) requiert l'utilisation d'anticorps spécifiques qui reconnaissent la séquence de la protéine d'intérêt. Avant

de pouvoir faire notre expérience il a donc fallu : (1) sélectionner les séquences protéiques d'intérêt, (2) choisir les séquences antigéniques utilisées pour la synthèse des anticorps, nous en avons choisi neuf, et (3) faire synthétiser les anticorps par MediMabs (Tableau S1). Une fois les anticorps reçus, nous avons pu faire les expériences d'immunobuvardage sur des cellules humaines HeLa et HEK293T. Parmi les 9 séquences à l'étude, les expériences de western blot ont permis d'identifier une bande protéique de manière hautement spécifique (Figure 2A) avec un des anticorps testés, celui dirigé contre MTALTND4. En effet, on peut voir la présence d'une seule bande sur la membrane avec un poids moléculaire d'environ 25kDa autant dans les cellules de type HeLa que les cellules de type HEK293T. Les expériences de western blot effectuées avec les 8 autres anticorps n'ont pas permis d'identifier clairement l'expression des nouvelles protéines putatives. En effet, soit les résultats ne montraient la présence d'aucune bande, signifiant l'absence de la protéine d'intérêt, soit on pouvait voir plusieurs bandes sur la membrane ce qui indique que la reconnaissance des anticorps n'était pas spécifique à nos séquences d'intérêt. Ainsi, nous avons pu identifier une protéine alternative mitochondriale humaine parmi les neuf séquences testées. La séquence codante de cette protéine est retrouvée dans le gène mitochondrial *nd4*. Plus précisément, elle se situe entre les nucléotides 11556 et 11856 du génome mitochondrial humain (voir tableau S1 et Figure 1, article). Puisque sa séquence est à l'intérieur du gène *nd4*, la protéine alternative a été nommée MTALTND4 pour « Mitochondrial Alternative ND4 protein ».

L'altORF de MTALTND4 est une séquence de 300 nucléotides traduite en un peptide de 99 acides aminés. Le code génétique utilisé pour la traduction est le code génétique mitochondrial des vertébrés et les codons AGA et AGG normalement reconnus comme codons stop sont considérés ici comme codants pour l'acide aminé arginine. Les codons AGA et AGG codent pour l'arginine dans le cytoplasme de la cellule, mais ils ont été considérés comme non-codants dans la

mitochondrie en raison de (1) l'absence d'ARNt^{ARG} codés par le génome mitochondrial qui reconnaissent les codons AGA et AGG et (2) puisque deux gènes mitochondriaux, *cox1* et *nd6*, ont des séquences se terminant respectivement par des triplets AGA et AGG (Anderson et al., 1981). Cependant, de récentes études ont montré que AGA et AGG ne seraient pas des codons stop dans le génome mitochondrial humain. En effet, la terminaison de la traduction des gènes *cox1* et *nd6* serait plutôt induite par un décalage de -1 nucléotide par le mitoribosome. Ce décalage placerait le codon UAG au site A du mitoribosome et serait le déclencheur de la terminaison de la traduction (Temperley et al., 2010; Ayyub et al., 2020). Ainsi, les codons AGA et AGG ne sont pas des codons stop dans le génome mitochondrial humain (Temperley et al., 2010; Ayyub et al., 2020). Puisque ces codons ne causent pas l'arrêt de la traduction et qu'ils sont trouvés dans les ORF de certains peptides dérivés de l'ADN mitochondrial tels que MOTS-c et certains SHLPs, il faut qu'ils codent pour un acide aminé et il serait logique que cet acide aminé soit l'arginine, comme c'est le cas pour le code génétique standard. Cela concorderait avec les recherches effectuées sur les peptides SHLPs (Cobb et al., 2016) et pourrait expliquer pourquoi le peptide dérivé du génome mitochondrial MOTS-c est retrouvé dans la mitochondrie alors qu'il avait été suggéré qu'il soit traduit dans le cytoplasme en raison d'un codon AGA considéré comme un codon stop positionné juste après le codon start. Si AGA est codant, MOTS-c pourrait donc être directement synthétisé dans la mitochondrie et non dans le cytoplasme de la cellule.

Comme le génome mitochondrial ne code pas pour un ARNt qui reconnaît AGA et AGG, il faudrait que les ARNt^{ARG} du cytoplasme de la cellule soient importés dans la mitochondrie. Or, les mitochondries de mammifères dont l'humain sont capables d'importer des ARNt du cytoplasme par un mécanisme inné dépendant de l'ATP (Rubio et al., 2008). Ce mécanisme est extrêmement intéressant et pourrait peut-être permettre d'exprimer les protéines mitochondriales

alternatives de manière indépendante de l'expression des 13 protéines mitochondriales de référence. En effet, les codons AGA et AGG sont trouvés chacun une seule fois dans les ORF des 13 protéines mitochondriales et uniquement à la fin des séquences de *cox1* et *nd6* respectivement (Richter et al., 2010). Ils ne sont donc pas utilisés pour la traduction des 13 protéines mitochondriales tandis qu'ils le sont pour certains peptides alternatifs tels que MOTS-c, SHLP et MTALTND4. La traduction de ces peptides étant dépendante de la présence d'ARNt^{ARG}, l'importation de ces ARNt pourrait permettre de traduire les peptides alternatifs sans avoir d'impact sur la traduction des 13 protéines mitochondriales ou peut-être pour une régulation indépendante de la traduction de ces 13 protéines. Il serait intéressant de faire des expériences telles que celles effectuées par Rubio et al. (2008) sur le transport de l'ARNt^{GLN} du cytoplasme à la mitochondrie afin d'examiner l'existence d'un mécanisme de transport similaire pour l'ARNt^{ARG}. La présence d'un tel mécanisme pourrait remettre en question l'évolution et les pressions de sélection exercées sur le génome mitochondrial. Si les codons AGA et AGG ont été démontrés comme codants dans le génome mitochondrial humain, il n'en est rien pour toutes les autres espèces vertébrées. Il serait donc intéressant d'effectuer d'autres recherches sur le décalage du mitoribosome telles que celles effectuées par Temperley et al. (2010) ou bien des expériences de profilage de mitoribosomes (Couvillon et Churchman, 2017) sur des espèces proches de l'humain telles que les primates afin de vérifier si AGA et AGG pourraient être codants chez d'autres espèces vertébrées. L'alignement de la séquence du peptide MTALTND4 chez cinq espèces de primates montre que la séquence est hautement conservée et que les codons AGA et AGG sont présents dans la séquence, (Figure 2B, article) ce qui renforce l'idée que AGA et AGG puissent aussi coder pour l'arginine chez les primates.

3.2 Détermination de l'origine de MTALTND4

Une fois le peptide MTALTND4 découvert, nous avons par la suite voulu vérifier l'origine du peptide, c'est-à-dire s'il est codé par le noyau ou par la mitochondrie. En raison de la présence de séquences mitochondriales transférées au noyau (NUMT Richetti et al., 2004) il est possible que des protéines dont les séquences sont trouvées dans la mitochondrie soient aussi trouvées dans le noyau et soient plutôt codées par ce dernier. Pour vérifier l'origine du peptide, nous avons effectué des western blot sur des cellules HeLa normales ainsi que sur des cellules HeLa Rho-0 (Figure 2D, article). Les cellules Rho-0 sont des cellules dont l'ADNmt a complètement été éliminé (Kukat et al., 2008). Dans de telles cellules, on ne retrouve normalement aucune protéine codée par le génome mitochondrial. Ces cellules peuvent être produites par exemple à l'aide de bromure d'éthidium ou d'enzymes de restrictions telles que EcoRI (Heitman, 1992).

Nos résultats (Figure 2D, article) montrent la présence de MTALTND4 et de la protéine mitochondriale Cox1 dans des cellules HeLa normales et leur absence dans les cellules HeLa Rho-0, tandis que les protéines ATP5 (mitochondriale) et actine (cytoplasmique) codées par le génome nucléaire ont été retrouvées autant dans les cellules HeLa normales que dans les cellules HeLa Rho-0. Ceci indique clairement que l'origine du peptide est bien mitochondriale et non nucléaire. Afin de confirmer que la traduction de MTALTND4 est effectuée dans la mitochondrie et non dans le cytoplasme, nous avons réalisé un western blot cette fois-ci sur des cellules HeLa normales ainsi que sur des cellules HeLa traitées au chloramphénicol. Le chloramphénicol est un antibiotique qui inhibe la synthèse de protéines traduites par les mitoribosomes sans affecter la synthèse de protéines traduites par les ribosomes cytoplasmiques (Leiter et al., 1999). Les résultats du western blot (Figure 2E, article) montrent la présence de MTALTND4 et de la protéine Cox1 (traduite dans la mitochondrie) dans des cellules HeLa non-traitées et leur absence dans les cellules

HeLa traitées au chloramphénicol, tandis que les protéines ATP5 et actine qui sont traduites dans le cytoplasme sont présentes autant dans les cellules HeLa non-traitées que les cellules HeLa traitées au chloramphénicol. Ceci indique clairement que MTALTND4 est synthétisée dans la mitochondrie, qu'elle n'est pas exportée dans le cytoplasme pour sa synthèse et donc supporte l'idée que les codons AGA et AGG ne sont pas des codons stop dans la mitochondrie chez l'humain.

3.3 Localisation de MTALTND4

Afin de comprendre où la protéine MTALTND4 est localisée dans la cellule, nous avons effectué des expériences d'immunofluorescence sur des cellules HeLa non-traitées et des cellules HeLa traitées au chloramphénicol. Les images obtenues à l'aide du microscope EVOS confirment les résultats obtenus par western blot quant à l'origine mitochondriale du peptide (Figure 3, article). En effet, les cellules traitées au chloramphénicol ne présentent aucun signal de MTALTND4 (Figure 3B, article). Les images des cellules HeLa non-traitées (Figure 3A, article) montrent une localisation mitochondriale du peptide avec un certain degré de localisation extra-mitochondriale. En effet, le signal vert de MTALTND4 superpose presque la totalité du signal mitochondrial de ATP5 en rouge. La couleur jaune résultante de la superposition des deux signaux témoigne de cette colocalisation. Cependant, on remarque certains endroits où le signal de MTALTND4 n'est pas situé à la mitochondrie. Ceci indique la possibilité que MTALTND4 ne soit pas trouvé uniquement dans la mitochondrie, mais qu'elle soit aussi exportée vers le cytoplasme de la cellule. D'autres peptides dérivés du génome mitochondrial tels que Humanine, MOTS-c et SHLP2 ont été retrouvés à l'extérieur de la mitochondrie et même dans le plasma (Voigt et Jelinek, 2016; Lee et al., 2015; Cobb et al., 2016). Cela signifie que non seulement ces

peptides sont exportés de la mitochondrie au cytoplasme, mais qu'ils sont aussi sécrétés à l'extérieur des cellules. Il serait donc possible que le peptide MTALTND4 soit aussi sécrété dans le plasma. Afin de tester cette hypothèse, nous avons réalisé des expériences de western blot sur des échantillons de plasma humain. Les résultats (Figure 2F, article) indiquent la présence de MTALTND4 dans le plasma humain. Il faut cependant rester prudent en interprétant ces résultats. En effet, des mitochondries ont déjà été détectées dans le plasma et auraient un rôle dans la communication cellule à cellule ainsi que dans le maintien de l'homéostasie (Song et al., 2020). Il se pourrait donc que le signal de MTALTND4 dans le plasma humain soit attribué à la présence de mitochondries. Bien que la protéine mitochondriale ATP5 n'ait pas été retrouvée dans le plasma (Figure 2F, article), il est possible qu'elle était tout de même présente, mais en trop petite quantité pour être détectée. Afin de confirmer la présence de MTALTND4 libre dans le plasma, il serait intéressant d'effectuer des expériences d'immunoessais par la technique d'ELISA telles que celles effectuées par Lee et al. (2015) et Cobb et al. (2016) sur les peptides mitochondriaux humains respectifs MOTS-c et SHLP2.

3.4 Poids moléculaire de MTALTND4

Afin de débiter des tests sur la fonction de la protéine, la séquence peptidique complète de MTALTND4 a été synthétisée par LifeTein. Ce peptide est donc formé des 99 acides aminés traduits de l'altORF et devrait avoir un poids moléculaire de 11,5 kDa (Tableau S1), soit près de deux fois moins que la masse suggérée par le signal de MTALTND4 obtenu par Western blot (Figure 2C, article). Les expériences effectuées sur le peptide synthétique en conditions dénaturantes révèlent pourtant la présence de deux bandes : une située environ à 11,5 kDa ce qui correspond à la taille attendue et une autre un peu plus haute aux alentours de 14 kDa (Figure 14).

La présence de cette deuxième bande à plus haut poids moléculaire amène à questionner la pureté du peptide synthétique. Cependant, la compagnie LifeTein qui a synthétisé le peptide a confirmé sa pureté par CLHP (chromatographie en phase liquide à haute performance). Si le peptide est pur, dans ce cas, on peut se demander si la deuxième bande ne serait pas causée par une réaction se produisant entre le peptide et le gel de polyacrylamide à travers lequel il migre. En effet, des recherches ont montré que la migration à travers un gel de polyacrylamide peut entraîner des modifications aux protéines telles que la désamination, l'oxydation des acides aminés méthionine et tryptophane, le clivage des liens aspartate-proline et finalement la formation d'adduits covalents avec des monomères d'acrylamide non polymérisés dans le gel (Bonaventura et al., 1994; Hachmann et Amshey, 2005). Ces modifications peuvent changer la mobilité électrophorétique des protéines, ce qui peut entraîner l'apparition de bandes floues ou multiples (Hachmann et Amshey, 2005). Plusieurs de ces modifications telles que la formation d'adduits d'acrylamide sont dépendantes du pH et plus fréquentes lorsque le pH est élevé (Hachmann et Amshey, 2005). L'électrophorèse sur un gel de Tris-tricine utilisé lors de nos expériences se fait à un pH assez élevé, soit 8,9. Il serait donc intéressant d'effectuer les expériences dans des conditions plus favorables, telles qu'en présence d'un tampon à pH plus neutre. Lorsque le peptide synthétique n'est pas mis en présence d'agents dénaturants, une troisième bande autour de 25 kDa (soit environ deux fois le poids moléculaire attendu) apparaît (Figure 14). L'apparition de cette troisième bande pourrait être expliquée par le fait que le peptide synthétique pourrait être capable de se dimériser en absence d'agents réducteurs.

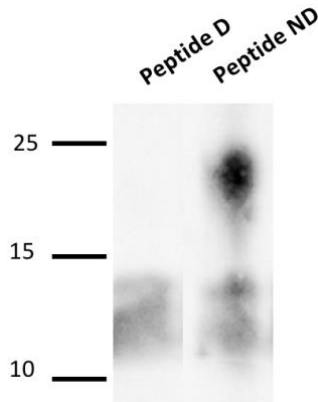


Figure 14. Immunobuvardage du peptide synthétique. Immunobuvardage de MTALTND4 sur le peptide synthétique en présence d'agents dénaturants (Peptide D) et en absence d'agents dénaturants (Peptide ND).

Contrairement au peptide synthétique qui migre aux alentours de 11,5 kDa, le signal de MTALTND4 obtenu par western blot lorsque l'on fait migrer les protéines extraites des cellules HeLa et HEK293T se situe à 25 kDa, ce qui correspond à peu près au double du poids moléculaire attendu. Cette différence significative de taille pourrait être causée par des modifications post-traductionnelles telles que la glycosylation et la phosphorylation. Des western blots effectués sur des cellules HeLa traitées avec des enzymes déglycosylantes ont éliminé la possibilité d'un événement de glycosylation puisque le traitement avec les enzymes n'a pas modifié le poids moléculaire de MTALTND4 (Figure 2G, article). La phosphorylation de la protéine a aussi été testée en traitant les cellules HeLa avec la protéine phosphatase Lambda. Tout comme les résultats obtenus pour la déglycosylation, la déphosphorylation n'a pas modifié le poids moléculaire apparent de MTALTND4 ce qui semble indiquer que la protéine n'est pas phosphorylée (Figure 2I, article). Toutefois, puisqu'aucun contrôle positif de la phosphatase Lambda n'a pu être obtenu, on ne peut pas complètement éliminer la possibilité d'un événement de phosphorylation. Étant

donné que le poids moléculaire apparent est près de deux fois celui du poids attendu, la formation d'un dimère résistant aux agents dénaturants serait une possibilité à considérer. Il existe en effet des dimères qui sont difficilement séparables malgré le traitement avec des agents dénaturants et réducteurs tels que le SDS, le DTT et le β -mercaptoéthanol (Fasshauer et al., 1998; Favreau et al., 2001; Ettinger et al., 1998; Gentile et al., 2002). La possibilité de la formation d'un dimère est renforcée par la présence dans la séquence peptidique d'acides aminés cystéines dont les résidus peuvent former des points disulfides favorisant la formation d'un dimère. De plus, l'apparition de la bande à 25 kDa lorsque le peptide synthétique migre en absence d'agents réducteurs renforce aussi la possibilité de la formation d'un dimère (Figure 14).

3.5 Immunoprécipitation et séquençage par spectrométrie de masse

Afin de vérifier que le signal présent dans les western blots corresponde vraiment à la séquence peptidique de MTALTND4, des immunoprécipitations ont été réalisées et les échantillons ont été envoyés à la plateforme de protéomique du centre universitaire de santé McGill pour analyse par spectrométrie de masse. Cette analyse avait aussi pour but d'identifier de potentiels partenaires d'interaction de la protéine. La séquence de MTALTND4 a été retrouvée dans les échantillons immunoprécipités avec l'anticorps anti-MTALTND4, mais pas dans les échantillons contrôles immunoprécipités avec le sérum pré-immun, ce qui confirme que ce qui est détecté par western blot correspond bien à MTALTND4. Plus précisément, les analyses de spectrométrie de masse ont permis de détecter 5 peptides représentant 21% de la séquence complète de MTALTND4. Plusieurs autres protéines absentes de l'échantillon contrôle ont été retrouvées dans l'échantillon immunoprécipité avec l'anticorps anti-MTALTND4 (Tableau S5). Parmi ces protéines, celle qui a été retrouvée en plus grand nombre est la glutathione peroxidase 1

(GPx-1). GPx-1 est une enzyme antioxydante qui réduit le peroxyde d'hydrogène en eau afin de limiter les dommages oxydatifs qu'il cause (Lubos et al., 2011). L'activité de cette enzyme est augmentée en condition d'hypoxie (diminution de la quantité d'oxygène), situation lors de laquelle la production de ROS générés par la mitochondrie est augmentée (Rousseau et al., 2006; Mansfield et al., 2004). Afin de valider quelles protéines parmi celles de la liste du tableau 1 sont réellement des partenaires d'interaction de MTALTND4, il serait intéressant d'effectuer des expériences de Pull-down qui complémenteraient les expériences d'immunoprécipitation (Louche et al., 2017) ainsi que des expériences de BioID qui permettraient de visualiser dans des cellules vivantes les protéines qui interagissent avec MTALTND4 ou qui sont à proximité de celle-ci (Roux et al., 2018).

3.6 Fonction de MTALTND4

En ce qui concerne la fonction de MTALTND4, mon collègue Stefano Bettinazzi a effectué quelques tests de respiration mitochondriale sur des cellules HeLa et HEK293T. Les résultats préliminaires indiquent que l'ajout exogène du peptide synthétique monomérique dans les cellules HeLa cause une diminution majeure de la respiration mitochondriale (Figure 15).

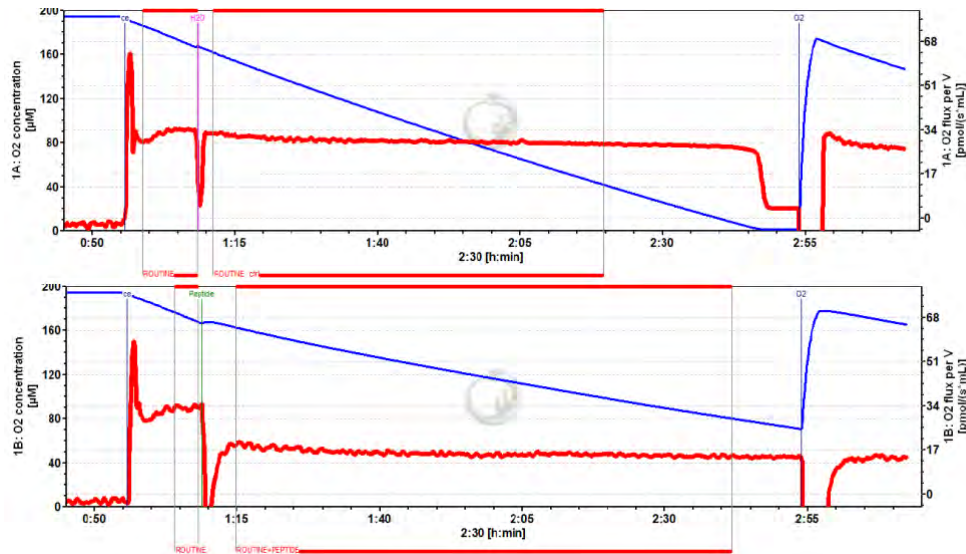


Figure 15. Respirométrie à haute résolution de cellules HeLa. La concentration d’oxygène est représentée en bleu et le flux d’oxygène est représenté en rouge. La figure du haut est le contrôle et montre que l’ajout d’eau dans la chambre cellulaire ne cause pas de diminution du flux d’oxygène, tandis que dans la figure du bas, l’ajout du peptide synthétique cause une nette diminution du flux d’oxygène.

Si MTALTND4 sous sa forme non-dimérisée cause une diminution de la respiration mitochondriale, on pourrait penser que ceci est néfaste pour les cellules et que son action est empêchée par la formation d’un dimère. Si c’est le cas et puisque la protéine est produite par la mitochondrie malgré son effet de diminution de la respiration, il doit y avoir des circonstances dans lesquelles l’action du peptide est bénéfique pour la cellule. Étant donné son effet de diminution de la respiration et son potentiel partenaire d’interaction GPx-1 impliqué dans l’hypoxie, nous avons posé l’hypothèse que MTALTND4 pourrait se dé-dimériser lorsque les cellules sont en hypoxie afin d’exercer son rôle. Des expériences de western blot sur des cellules HeLa en hypoxie ont été réalisées afin d’examiner l’impact de l’hypoxie sur la dimérisation de MTALTND4. Les résultats (Figure 16A) montrent en plus de la bande habituelle retrouvée à 25 kDa, deux autres bandes autour de 11,5 kDa et 14 kDa tout comme les bandes observées suite à la

migration du peptide synthétique. Les signaux étant très faibles, il est difficile de tirer conclusion de ses résultats. Cependant, les bandes à 11,5 kDa et 14 kDa n'ayant encore jamais été détectées par western blot sur les protéines extraites des cellules HeLa (seulement celle à 25 kDa était détectée), il semble que la condition hypoxique pourrait être impliquée dans la dé-dimérisation de MTALTND4. D'autres expériences effectuées par mon collègue Thierry Choquette sur des cellules HeLa en hypoxie et en normoxie commandées de Novus Biologicals montrent que le signal de MTALTND4 à 25 kDa est beaucoup moins fort dans les cellules en hypoxie qu'en normoxie (Figure 16B). Ces résultats suggèrent encore une fois que l'hypoxie a un effet sur MTALTND4, mais il est difficile de dire s'il implique un événement de dé-dimérisation puisque les bandes à 11,5 kDa et 14 kDa sont extrêmement faibles et semblent aussi être présentes en condition de normoxie. D'autres expériences sur des cellules en hypoxie et en normoxie sont nécessaires afin de confirmer notre hypothèse.

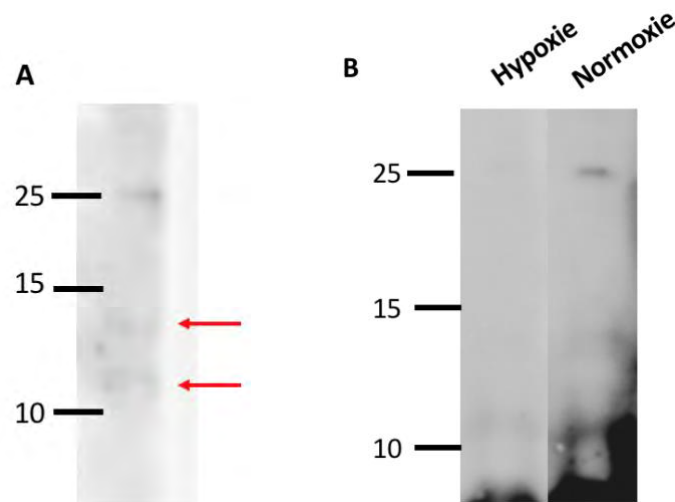


Figure 16. Immunobuvardage de MTALTND4 sur des cellules HeLa en hypoxie. (A) Une bande principale est observée à 25 kDa et deux autres bandes pointées par les flèches en rouge à environ 11,5 kDa et 14 kDa peuvent aussi être observées. **(B)** Comparaison du signal de MTALTND4 entre les cellules HeLa en hypoxie et en normoxie. Le signal est plus faible en hypoxie.

Conclusion et perspectives

Par ce projet de recherche, nous avons réussi à identifier une nouvelle protéine humaine dérivée du génome mitochondrial. Cette protéine, MTALTND4, est la première protéine mitochondriale alternative humaine qui a été découverte dont l'altORF est situé à l'intérieur et sur le même brin qu'une des 13 séquences codantes pour les protéines mitochondriales de référence. Cette découverte confirme que le potentiel codant du génome mitochondrial humain a été largement sous-estimé et ouvre la porte à un nouveau domaine de recherche, celui des protéines mitochondriales alternatives humaines. Les 8 autres séquences qui étaient à l'étude n'ont pas été détectées par western blot. Cela ne veut pas nécessairement dire qu'elles ne sont pas traduites en protéines. Afin d'améliorer les chances de détection, il serait intéressant d'augmenter la spécificité des anticorps. Pour ce faire, on pourrait par exemple produire des anticorps contre une plus grande séquence antigénique que ce qui a été fait et si possible contre la séquence entière. De plus, la purification des anticorps permettrait d'améliorer la spécificité du signal. Parmi les 218 autres ORF nouvellement caractérisés par Angers et al. (2019) il est fort probable que l'on retrouve plusieurs séquences qui codent des protéines fonctionnelles qui sont aujourd'hui encore inconnues. De futures recherches devraient donc être réalisées afin de découvrir ces nouvelles protéines. Afin de sélectionner des séquences dont la probabilité de traduction est élevée, il serait intéressant de choisir les séquences qui ont été détectées dans les bases de données de spectrométrie de masse. En effet, leur détection dans les bases de données renforce l'idée qu'elles sont traduites en protéines. Les peptides dérivés de l'ANDmt humain Humaine, MOTS-c, SHLP2 et SHLP3 découverts dans les dernières années et dont les séquences codantes sont trouvées sur l'ARNr 16S et 12S ont des implications majeures dans les fonctions cellulaires. Parmi ces fonctions on retrouve la neuroprotection (Humaine; Hashimoto et al., 2001), la régulation du métabolisme, la régulation

de l'insuline, la résistance à l'obésité (MOTS-c; Lee et al., 2015), la diminution de l'apoptose et de la génération de ROS ainsi que l'amélioration du métabolisme mitochondrial (SHLP2 et SHLP3; Cobb et al., 2016). De plus, les niveaux de Humanine et de SHLP2 diminuent avec l'âge, ce qui suggère que ces peptides sont impliqués dans la survie. Étant donné leurs effets importants, ils pourraient être utilisés en tant que molécules thérapeutiques. Il est aussi à se demander si certaines mutations silencieuses sur un gène de référence ne pourraient pas être responsables de maladies mitochondriales à cause de l'effet de cette mutation sur un altORF chevauchant la séquence. Ceci montre l'importance que peuvent avoir les peptides alternatifs mitochondriaux et donc la pertinence de continuer les recherches sur le protéome alternatif mitochondrial.

Bien que le rôle précis de MTALTND4 n'ait pas encore été déterminé, nos résultats préliminaires indiquent que la protéine a un effet sur la respiration mitochondriale et pourrait potentiellement agir en condition d'hypoxie. Sa présence dans le plasma pourrait permettre, lors de conditions spécifiques, d'enclencher rapidement une voie de signalisation pour répondre à un stress comme l'hypoxie. Afin de confirmer ces résultats et de comprendre plus en profondeur les implications de MTALTND4, plusieurs autres expériences seront nécessaires et seront réalisées très prochainement par l'équipe de laboratoire de Sophie Breton et Annie Angers. Ces expériences incluent: (1) l'effet sur l'apoptose, la croissance cellulaire et l'expression de certains gènes suite à l'exposition exogène et la sur-expression endogène de MTALTND4 dans des cellules HeLa et HEK293T en culture, (2) l'analyse par western blot de l'expression de MTALTND4 dans différents tissus humains tels que la peau, le foie, la rate, le rein et l'épithélium nasal (qui sont tous des tissus dans lesquels la protéine ND4 est exprimée) ainsi que dans les tissus d'autres espèces de primates telles que le chimpanzé, (3) l'analyse par western blot de l'expression différentielle de MTALTND4 dans différentes conditions telles que, l'hypoxie, l'apoptose, le jeûne et le traitement

au peroxyde (4) l'analyse des partenaires d'interaction de MTALTND4 à l'aide d'expériences de Pull-down et (5) la confirmation de la sécrétion de MTALTND4 dans le plasma par la technique d'ELISA.

Finalelement, la découverte de cette nouvelle protéine alternative du génome mitochondrial humain amène plusieurs questions telles que : quelle est l'ampleur de ce protéome alternatif mitochondrial humain ? Quelles sont les implications de ce protéome alternatif dans les fonctions mitochondriales et cellulaires, et dans les maladies mitochondriales ? Quels sont les mécanismes permettant la traduction de ce protéome ? Comment cette découverte remet-elle en question l'étude de l'évolution du génome mitochondriale et les pressions de sélection exercées sur la mitochondrie ? De futures recherches sur les protéomes alternatifs seront donc nécessaires afin de répondre à ces questions.

Références

Adams, K. L., & Palmer, J. D. (2003). Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution*, 29(3), 380-395.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2011). Biologie moléculaire de la cellule. (5ième Ed.). Paris: Médecine Sciences Publications.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402.

Amin, M. R., Yurovsky, A., Chen, Y., Skiena, S., & Futcher, B. (2018). Re-annotation of 12,495 prokaryotic 16S rRNA 3' ends and analysis of Shine-Dalgarno and anti-Shine-Dalgarno sequences. *PLoS One*, 13(8), e0202767.

Amorim, A., Fernandes, T., & Taveira, N. (2019). Mitochondrial DNA in human identification: a review. *PeerJ*, 7, e7314.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806), 457-65.

Angers, A., Ouimet, P., Tsyvian-Dzyabko, A., Nock, T., & Breton, S. (2019). L'ADN mitochondrial, un potentiel codant mésestimé [The underestimated coding potential of mitochondrial DNA]. *Medecine Sciences : M/S*, 35(1), 46-54.

Ardern, Z., Neuhaus, K., & Scherer, S. (2020). Are Antisense Proteins in Prokaryotes Functional? *Frontiers in Molecular Biosciences*, 7, 187.

Audagnotto, M., & Dal Peraro, M. (2017). Protein post-translational modifications: *In silico* prediction tools and molecular modeling. *Computational and Structural Biotechnology Journal*, 15, 307–319.

Aylett, C. H., & Ban, N. (2017). Eukaryotic aspects of translation initiation brought into focus. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 372(1716), 20160186.

Ayyub, S. A., Gao, F., Lightowers, R. N., & Chrzanowska-Lightowers, Z. M. (2020). Rescuing stalled mammalian mitoribosomes – what can we learn from bacteria? *Journal of Cell Science*, 133(1), jcs231811.

Bonaventura, C., Bonaventura, J., Stevens, R., & Millington, D. (1994). Acrylamide in polyacrylamide gels can modify proteins during electrophoresis. *Analytical Biochemistry*, 222(1), 44–48.

Breton, S., Milani, L., Ghiselli, F., Guerra, D., Stewart, D. T., & Passamonti, M. (2014). A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends in Genetics : TIG*, *30*(12), 555–564.

Brunet, M. A., Leblanc, S., & Roucou, X. (2020). Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. *Experimental Cell Research*, *393*(1), 112057.

Cao, J., Arha, M., Sudrik, C., Mukherjee, A., Wu, X., & Kane, R. S. (2015). A universal strategy for regulating mRNA translation in prokaryotic and eukaryotic cells. *Nucleic Acids Research*, *43*(8), 4353–4362.

Christian, B. E., & Spremulli, L.L. (2012). Mechanism of protein biosynthesis in mammalian mitochondria. *Biochimica and Biophysica Acta*, *1819*(9-10), 1035-54.

Couvillion, M. T., & Churchman, L. S. (2017). Mitochondrial Ribosome (Mitoribosome) Profiling for Monitoring Mitochondrial Translation In Vivo. *Current Protocols in Molecular Biology*, *119*, 4.28.1–4.28.25.

Cobb, L.J., Lee, C., Xiao, J., Yen, K., Wong, R. G., Nakamura, H. K., Mehta, H. H., Gao, Q., Ashur, C., Huffman, D. M., Wan, J., Muzumdar, R., Barzilai, N., & Cohen, P. (2016). Naturally occurring mitochondrial-derived peptides are age-dependent regulators of apoptosis, insulin sensitivity, and inflammatory markers. *Aging*, *8*(4), 796-809.

Dinman J. D. (2006). Programmed Ribosomal Frameshifting Goes Beyond Viruses: Organisms from all three kingdoms use frameshifting to regulate gene expression, perhaps signaling a paradigm shift. *Microbe (Washington, D.C.)*, *1*(11), 521–527.

D'Souza, A. R., & Minczuk, M. (2018). Mitochondrial transcription and translation: overview. *Essays in Biochemistry*, *62*(3), 309-320.

Ettinger, R. A., Liu, A. W., Nepom, G. T., & Kwok, W. W. (1998). Exceptional stability of the HLA-DQA1*0102/DQB1*0602 alpha beta protein dimer, the class II MHC molecule associated with protection from insulin-dependent diabetes mellitus. *Journal of Immunology (Baltimore, Md. : 1950)*, *161*(11), 6439–6445.

Fasshauer, D., Eliason, W. K., Brünger, A. T., & Jahn, R. (1998). Identification of a minimal core of the synaptic SNARE complex sufficient for reversible assembly and disassembly. *Biochemistry*, *37*(29), 10354–10362.

Faure, E., Delaye, L., Tribolo, S., Levasseur, A., Seligmann, H., & Barthélémy, R. M. (2011). Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. *Biology Direct*, *6*, 56.

Favreau, C., Bastos, R., Cartaud, J., Courvalin, J. C., & Mustonen, P. (2001). Biochemical characterization of nuclear pore complex protein gp210 oligomers. *European Journal of Biochemistry*, *268*(14), 3883–3889.

Gentile, F., Amodeo, P., Febbraio, F., Picaro, F., Motta, A., Formisano, S., Nucci, R., 2002. SDS-resistant active and thermostable dimers are obtained from the dissociation of homotetrameric beta-glycosidase from hyperthermophilic *Sulfolobus solfataricus* in SDS. Stabilizing role of the A-C intermonomeric interface. *Journal of Biological Chemistry*, 277, 44050–44060.

Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harb Perspect Biol*, 4(9), a011403.

Hachmann, J. P., & Amshey, J. W. (2005). Models of protein modification in Tris-glycine and neutral pH Bis-Tris gels during electrophoresis: effect of gel pH. *Analytical Biochemistry*, 342(2), 237–245.

Harhay, G. P., Sonstegard, T. S., Keele, J. W., Heaton, M. P., Clawson, M. L., Snelling, W. M., Wiedmann, R. T., Van Tassell, C. P., & Smith, T. P. (2005). Characterization of 954 bovine full-CDS cDNA sequences. *BMC Genomics*, 6, 166.

Hashiguchi, K., & Zhang-Akiyama, Q. M. (2009). Establishment of human cell lines lacking mitochondrial DNA. *Methods in Molecular Biology*, 554, 383-391.

Hashimoto, Y., Ito, Y., Niikura, T., Shao, Z., Hata, M., Oyama, F., & Nishimoto, I. (2001). Mechanisms of neuroprotection by a novel rescue factor humanin from Swedish mutant amyloid precursor protein. *Biochemical and Biophysical Research Communication*, 283(2), 460-468.

Hashimoto, Y., Niikura, T., Tajima, H., Yasukawa, T., Sudo, H., Ito, Y., Kita, Y., Kawasumi, M., Kouyama, K., Doyu, M., Sobue, G., Koide, T., Tsuji, S., Lang, J., Kurokawa, K., & Nishimoto, I. (2001). A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and Abeta. *Proceedings of The National Academy of Sciences of The United States of America*, *98*(11), 6336-6341.

Heitman, J. (1992). How the EcoRI endonuclease recognizes and cleaves DNA. *Bioessays*, *14*(7), 445-54.

Hofmann, K., & Stoffel, W. (1993). TMbase-a database of membrane spanning proteins segments. *Biological Chemistry Hoppe Seyler*, *347*, 166.

Junge, W., & Nelson, N. (2015). ATP synthase. *Annual Review of Biochemistry*, *84*, 631-657.

Karnkowska, A., Vacek, V., Zubáčová, Z., Treitli, S. C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L. D., Herman, E. K., Soukal, P., Hroudová, M., Doležal, P., Stairs, C. W., Roger, A. J., Eliáš, M., Dacks, J. B., Vlček, Č., & Hampl, V. (2016). A Eukaryote without a Mitochondrial Organelle. *Current Biology*, *26*(10), 1274–1284.

Kasinath, B. S., Feliers, D., Sataranatarajan, K., Ghosh Choudhury, G., Lee, M. J., & Mariappan, M. M. (2009). Regulation of mRNA translation in renal physiology and disease. *American Journal of Physiology. Renal Physiology*, *297*(5), F1153–F1165.

Kim, K. H., Son, J. M., Benayoun, B. A., & Lee, C. (2018). The Mitochondrial-Encoded Peptide MOTS-c Translocates to the Nucleus to Regulate Nuclear Gene Expression in Response to Metabolic Stress. *Cell Metabolism*, 28(3), 516–524.e7.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 44(2), 283-92.

Kozak, M. (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *Journal of Molecular Biology*, 196(4), 947–950.

Kukat, A., Kukat, C., Brocher, J., Schäfer, I., Krohne, G., Trounce, I. A., Villani, G., & Seibel, P. (2008). Generation of rho0 cells utilizing a mitochondrially targeted restriction endonuclease and comparative analyses. *Nucleic Acids Research*, 36(7), e44.

Ladiges, W., Wanagat, J., Preston, B., Loeb, L., & Rabinovitch, P. (2010). A mitochondrial view of aging, reactive oxygen species and metastatic cancer. *Aging Cell*, 9(4), 462-465.

Landry, C. R., Zhong, X., Nielly-Thibault, L., & Roucou, X. (2015). Found in translation: functions and evolution of a recently discovered alternative proteome. *Current Opinion in Structural Biology*, 32, 74–80.

Lee, C., Zeng, J., Drew, B. G., Sallam, T., Martin-Montalvo, A., Wan, J., Kim, S. J., Mehta, H., Hevener, A. L., de Cabo, R., & Cohen, P. (2015). The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metabolism*, 21(3), 443–454.

Leiter, L. M., Thatte, H. S., Okafor, C., Marks, P. W., Golan, D. E., & Bridges, K. R. (1999). Chloramphenicol-induced mitochondrial dysfunction is associated with decreased transferrin receptor expression and ferritin synthesis in K562 cells and is unrelated to IRE-IRP interactions. *Journal of Cellular Physiology*, 180(3), 334–344.

Lin, M. & Beal, M. (2006). Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature*, 443(7113), 787–795.

Louche, A., Salcedo, S. P., & Bigot, S. (2017). Protein-Protein Interactions: Pull-Down Assays. *Methods in Molecular Biology (Clifton, N.J.)*, 1615, 247–255.

Lubos, E., Loscalzo, J., & Handy, D. E. (2011). Glutathione peroxidase-1 in health and disease: from molecular mechanisms to therapeutic opportunities. *Antioxidants & Redox Signaling*, 15(7), 1957–1997.

Mansfield, K. D., Simon, M. C., & Keith, B. (2004). Hypoxic reduction in cellular glutathione levels requires mitochondrial reactive oxygen species. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, 97(4), 1358–1366.

Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. (2018). Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature*, 557(7703), 101-105.

Martinez, T. F., Chu, Q., Donaldson, C., Tan, D., Shokhirev, M. N., & Saghatelian, A. (2020). Accurate annotation of human protein-coding small open reading frames. *Nature Chemical Biology*, 16(4), 458-468.

Minasyan, L., Sreekumar, P. G., Hinton, D. R., & Kannan, R. (2017). Protective Mechanisms of the Mitochondrial-Derived Peptide Humanin in Oxidative and Endoplasmic Reticulum Stress in RPE Cells. *Oxidative Medicine and Cellular Longevity*, 2017, 1675230.

Mittwoch, U., Kirk, D. J., & Wilkie, D. (1974). Effects of mitochondrial inhibition by chloramphenicol on the mitotic cycle of human cell cultures. *Journal of Medical Genetics*, 11(3), 260–266.

Meydan, S., Va'zquez-Laslop, N., & Mankin, A.S. (2018). Genes within genes in bacterial genomes. *Microbiology Spectrum*, 6(4).

Moulleron, H., Delcourt, V., & Roucou, X. (2016). Death of a dogma: eukaryotic mRNAs can code for more than one protein. *Nucleic Acids Research*, 44(1), 14-23.

Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H., & Miura, K. (2008). Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic Acids Research*, *36*(3), 861–871.

Normark, S., Bergström, S., Edlund, T., Grundström, T., Jaurin, B., Lindberg, F. P., & Olsson, O. (1983). Overlapping genes. *Annual Review of Genetics*, *17*, 499-525.

Nunnari, J., & Suomalainen, A. (2012). Mitochondria: in sickness and in health. *Cell*, *148*(6), 1145–1159.

Ouimet, P., Kienzle, L., Lubosny, M., Burzyński, A., Angers, A., & Breton, S. (2020). The ORF in the control region of the female-transmitted *Mytilus* mtDNA codes for a protein. *Gene*, *725*, 144161.

Orr, M. W., Mao, Y., Storz, G., & Qian, S. B. (2020). Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Research*, *48*(3), 1029-1042.

Ramanathan, A., Robb, G. B., & Chan, S. H. (2016). mRNA capping: biological functions and applications. *Nucleic Acids Research*, *44*(16), 7511-26.

Ricchetti, M., Tekaia, F., & Dujon, B. (2004). Continued Colonization of the Human Genome by Mitochondrial DNA. *PLoS Biol* *2*(9): e273.

Richter, R., Pajak, A., Dennerlein, S., Rozanska, A., Lightowlers, R. N., & Chrzanowska-Lightowlers, Z. M. (2010). Translation termination in human mitochondrial ribosomes. *Biochemical Society Transactions*, 38(6), 1523–1526.

Rodina, M. V. (2018). Translation in Prokaryotes. *Cold Spring Harbor Perspectives in Biology*, 10(9), a032664.

Roger, A. J., Muñoz-Gómez, S. A., & Kamikawa, R. (2017). The Origin and Diversification of Mitochondria. *Current Biology*, 27(21), R1177–R1192.

Rousseau, A. S., Richer, C., Richard, M. J., Favier, A., & Margaritis, I. (2006). Plasma glutathione peroxidase activity as a potential indicator of hypoxic stress in breath-hold diving. *Aviation, Space, and Environmental Medicine*, 77(5), 551–555.

Roux, K. J., Kim, D. I., Burke, B., & May, D. G. (2018). BioID: A Screen for Protein-Protein Interactions. *Current Protocols in Protein Science*, 91, 19.23.1–19.23.15.

Rubio, M. A., Rinehart, J. J., Krett, B., Duvezin-Caubet, S., Reichert, A. S., Söll, D., & Alfonzo, J. D. (2008). Mammalian mitochondria have the innate ability to import tRNAs by a mechanism distinct from protein import. *Proceedings of The National Academy of Sciences of The United States of America*, 105(27), 9186–9191.

Schägger, H., & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry*, *166*(2), 368–379.

Song, X., Hu, W., Yu, H., Wang, H., Zhao, Y., Korngold, R., & Zhao, Y. (2020). Existence of Circulating Mitochondria in Human and Animal Peripheral Blood. *International Journal of Molecular Sciences*, *21*(6), 2122.

Taanman, J.W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta*, *1410*(2), 103-23.

Temperley, R., Richter, R., Dennerlein, S., Lightowers, R. N., & Chrzanowska-Lightowers, Z. M. (2010). Hungry codons promote frameshifting in human mitochondrial ribosomes. *Science (New York, N.Y.)*, *327*(5963), 301.

Tian, F., Wang, C., Tang, M., Li, J., Cheng, X., Zhang, S., Ji, D., Huang, Y., & Li, H. (2016). The antibiotic chloramphenicol may be an effective new agent for inhibiting the growth of multiple myeloma. *Oncotarget*, *7*(32), 51934–51942.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680.

Tsuji, Y., (2020). Transmembrane protein western blotting: Impact of sample preparation on detection of SLC11A2 (DMT1) and SLC40A1 (ferroportin). *PloS one*, 15(7), e0235563.

Vanderperre, B., Lucier, J. F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., Wisztorski, M., Salzet, M., Boisvert, F. M., & Roucou, X. (2013). Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS One*, 8(8), e70698.

Voigt, A., & Jelinek, H. F. (2016). Humanin: a mitochondrial signaling peptide as a biomarker for impaired fasting glucose-related oxidative stress. *Physiological Reports*, 4(9), e12796.

Wallace, D. C. (2018). Mitochondrial genetic medicine. *Nature Genetics*. 50(12), 1642-1649.

Wang, Y., & Hekimi, S. (2015). Mitochondrial dysfunction and longevity in animals: Untangling the knot. *Science*, 350(6265), 1204-7.

Yang, C., Hockenberry, A. J., Jewett, M. C., & Amaral, L. (2016). Depletion of Shine-Dalgarno Sequences Within Bacterial Coding Regions Is Expression Dependent. *G3 (Bethesda, Md.)*, 6(11), 3467-3474.

Yee, C., Yang, W., & Hekimi, S. (2014). The intrinsic apoptosis pathway mediates the pro-longevity response to mitochondrial ROS in *C. elegans*. *Cell*, *157*(4), 897-909.

Annexes

Supplementary Material

Supplementary Figures

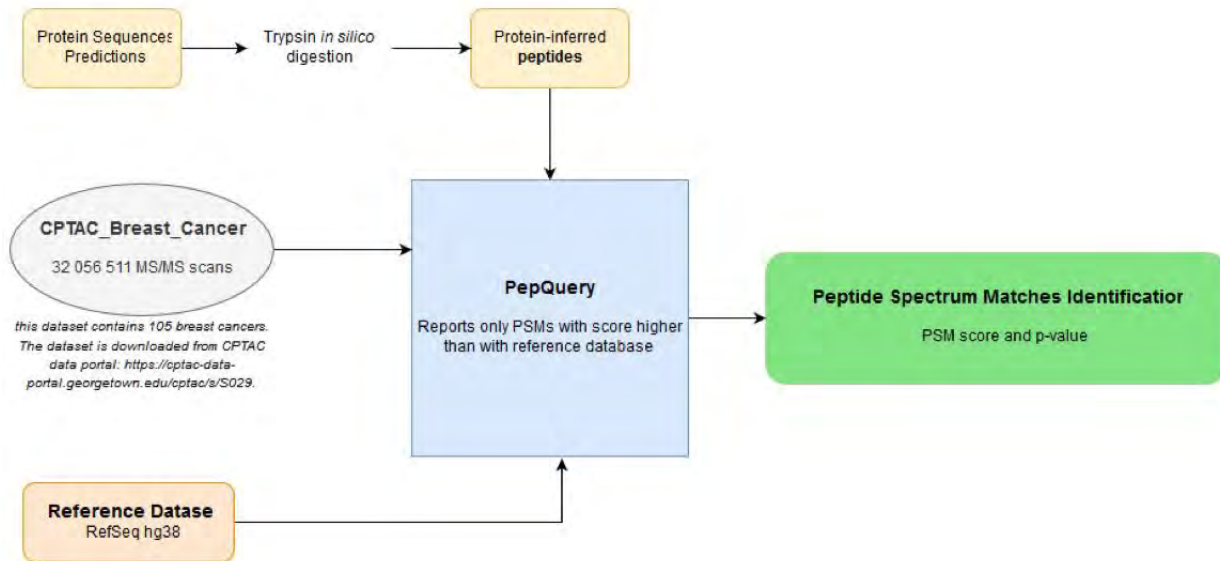


Figure S1. Interrogation of mitochondria-derived mass spectrometry datasets using the novel PepQuery tool

Predicted protein sequences are digested *in silico* with trypsin. Then each peptide is queried using the novel PepQuery tool from which any Peptide Spectrum Match (PSM) that hits better with the queried peptide than with any protein from the reference database is retrieved.

```

Query: None Query ID: lcl|Query_3741 Length: 99

>Homo sapiens mitochondrion, complete genome
Sequence ID: NC_012920.1 Length: 16569
Range 1: 11557 to 11853

Score:152 bits(383), Expect:3e-42,
Method:Compositional matrix adjust.,
Identities:98/99(99%), Positives:99/99(100%), Gaps:0/99(0%)

Query 1      MRHNYNKLHLPTTNRPKIAHCMLFnqphsprsnshshpnpLKLHRRSHSHNRPRAYILIT 60
             MRHNYNKLHLPTTNRPKIAHC+LFNQPHSPRSNSHSHPNPLKLHRRSHSHNRPRAYILIT
Sbjct 11557  MRHNYNKLHLPTTNRPKIAHCILFNQPHSPRSNSHSHPNPLKLHRRSHSHNRPRAYILIT 11736

Query 61     ILPSKCLKLRTHSQSHHNPLSRTSNSTPTNSFLMTSSKPR 99
             ILPSKCLKLRTHSQSHHNPLSRTSNSTPTNSFLMTSSKPR
Sbjct 11737  ILPSKCLKLRTHSQSHHNPLSRTSNSTPTNSFLMTSSKPR 11853

>Homo sapiens chromosome 5, GRCh38.p13 Primary Assembly
Sequence ID: NC_000005.10 Length: 181538259
Range 1: 134926946 to 134927239

Score:132 bits(331), Expect:3e-35,
Method:Compositional matrix adjust.,
Identities:87/98(89%), Positives:91/98(92%), Gaps:0/98(0%)

Query 1      MRHNYNKLHLPTTNRPKIAHCMLFnqphsprsnshshpnpLKLHRRSHSHNRPRAYILIT 60
             +RHNYNKLHLP TNRPKIAHC+LFNQPHSPRSNSH HPNPLKLHRRSHSH+RPR YILIT
Sbjct 134927239 VRHNYNKLHLPATNRPKIAHCILFNQPHSPRSNSHHPNPLKLHRRSHSHSRPTYILIT 134927060

Query 61     ILPSKCLKLRTHSQSHHNPLSRTSNSTPTNSFLMTSSKPR 98
             ILPSKCLKLRT+ QSHHN LSRTSN TPTNS LMTSSKPR
Sbjct 134927059 ILPSKCLKLRTYPQSHHNLSRTSNPTPTNSLLMTSSKPR 134926946

>Homo sapiens chromosome 2, GRCh38.p13 Primary Assembly
Sequence ID: NC_000002.12 Length: 242193529
Range 1: 120214413 to 120214703

Score:74.7 bits(182), Expect:3e-15,
Method:Compositional matrix adjust.,
Identities:49/97(51%), Positives:68/97(70%), Gaps:0/97(0%)

Query 1      MRHNYNKLHLPTTNRPKIAHCMLFnqphsprsnshshpnpLKLHRRSHSHNRPRAYILIT 60
             MR+ Y+KL+L TTN+ KI +C+L ++P+S +S+ HP PLKL+R SH +N PR Y ++
Sbjct 120214413 MRNGYDKLYLSTTNQSKITYCLLLHKPYSTCYHSYPHPKPLKLYRCSHYPYNCPRTYFILI 120214592

Query 61     ILPSKCLKLRTHSQSHHNPLSRTSNSTPTNSFLMTSSK 97
             ILPSK +LR + +H R SN+ TNS LMTSSK
Sbjct 120214593 ILPSKFELRANYPYNNHITYPRPSNTASTNSLLMTSSK 120214703

```

Figure S2. Comparison of MTALTND4 peptide encoded in mitochondrial DNA (mtDNA) and hypothetically in nuclear DNA that have been transferred from mtDNA (NUMT) through evolution. Only sequences with >50% identities and complete without stop codons are shown.

Color Align Properties results

G, A, V, L, I
 F, Y, W
 C, M
 S, T
 K, R, H
 D, E
 N, Q
 P

Human	MRHNYNKLHLP	TNRPKIAHCMLFNQPHSPRSNSHSHPNPL	LHRRSHSHNRPRAYILITILPSKCLKLRT	HSQSHHNPLS	80
Chimpanzee	MRHNNKLHLP	TNRPKIAHCMLFSQPHSPRSNSHSHPNPL	LHRRNYPHNRPTYILITILPSKCLKLW	THPQSHHNSLP	80
Bonobo	MRHNNKLHLP	TNRPKIAHCMLFSQPHSPRSNSHSHPNPL	LHRRSHSHNRPTYILITILPSKCLKLRT	HPQSHHNSLP	80
Gorilla	MGYNNKLHLP	TANRPKIAHCMLLN*PHSPCSSSYYPNP	LHRRSCSYNCPRTYIIIIILPSKCLKLRT	NPQPHHNSLS	79
Orangutan	MRHNNKLHLL	TNRPKIHRMLLSQPHSACSSYPHNP	LHRRNHPHNCPTHILPTVLPKCLKLRT	NPQPHHNSLP	80
Zebrafish	MRSNHNRI	ISLSTTNRKIANCLFISRPYGTSCRNFDSNPL	IYRSNYFNRPRTNLRPILLSHI*VWTNS*PDHNPRS	78	
Bear	MGHNYNKFY	LPTPNRFKVPNCLFLC*SHSPSHCSSINPNPM	LYRSNSLNNRPRTNILLTILSCKLQLWTNP*PNYDRA	78	
Lion	MRNSHDKFY	LSTPDGPEILNRMLISKPHSPSNCSTNP	NALLYRSYPYNSRTNLLSTILSCKLKLRTSP*PNNNSST	79	
Dog	MRHNNKLYL	PPSNRSKIPNCMLLS*SYGTGYRSGSYNTM	LYRCNSNSNCPWFNILLNTILLSQLQLRTNP*PYYNSRT	78	
Rat	MRHNYN*L	NLPTPNRSKITNCLLLS*PHSTCYCINHNPN	SMELHRSYNTNRPRLNLLTILPSKHQLRTNSQPNNYSS	79	
Mouse	MRNNYN*L	NLLTPNRFKITNCLLLS*PHSTCYCINHNPN	SMELHRSNNINNRTWPHIITPILPSKLQLRTDPQPYNHGP	78	
Horse	VRDNHN*F	HLPATNRPEITYCMLLCQPHSPSNNRSPH	PTMELYRSYPNNRPRPIINTILPSKCLKLRTYPQPNYNPSP	79	
Naked Mole Rat	MRNSYNELH	LPTANRPKIPYCLLLHQPHSPCNHSYPH	SNTMELHRSNHLNRRPRPHIIPPILPSKLQLRTHPQPHNSPSS	80	
Dolphin	MGNHNNQF	YLPSTSNRPKITNRMLLRQPHSTRHCSCPYP	NPLLHRSYRLNNCPRPYILHTILPSKCLKLWTHPQPNYDPST	80	
Cow	MRHNYNQL	NLPPSNGPKIHRMLLCKPHSTRYRSHYPDTL	LHRSNRPYDCPRPHILHTFLPSKCLKLRTNPQPHNSSS	80	
Human	RTSNSTPTNS	SFLMTSSKPR-----	99		
Chimpanzee	RTSNSTPTNS	LMLTPSKPR-----	99		
Bonobo	RTSNSTPTNS	LMLTPSKPRQPRTPHH-----	107		
Gorilla	RTPNPTPTNS	SPLMTSGKPRQPRLTPHH-----	106		
Orangutan	RPSNSTPPNS	SPLMTSSKPH-----	99		
Zebrafish	RTPD	SATASNSMVIYC-----	94		
Bear	RTD	TSPTYGSLMTTGKPH-----	97		
Lion	R	PANHPPSNSCLMTTSQPREPGPAPYH-----	106		
Dog	RTSN	SPSPNSLMTISKPHKSGSPSNN-----	105		
Rat	RT	TNNLSIDSNMMTISKLSQPSTTPN-----	106		
Mouse	RTS	NGLPTYSHMMTDSKSS-----	97		
Horse	RTS	NTSSPYSSLMTISQPNQPGPPSQH-----	106		
Naked Mole Rat	RT	TNHPSPD	SALMTPS-----	96	
Dolphin	RT	NLFSTNSYLMTISMLNKPCPTSYH-----	107		
Cow	RT	NAPSTNSHLMTTSKSNQLSSTPNNQLNWR	TICSNVNLFMI 123		

Figure S3. Multiple peptide sequence alignment in 15 mammal species: human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), naked mole rat (*Heterocephalus glaber*), dog (*Canis lupus familiaris*), cow (*Bos taurus*), zebrafish (*Danio rerio*), lion (*Panthera leo*), bear (*Ursus arctos*), horse (*Equus caballus*), dolphin (*Tursiops truncatus*).

Supplementary Tables

Table S1. Chosen candidates for antibody production.

Approach used	Amino acid sequence (antigenic sequence in bold underlined)	Length	Predicted MW	Position in the reference human mtDNA
In silico & Kozak	MGCSGSSVSQCYRVHTPQTKMPNAWRAPVSG	31aa	3,5kDa	non-coding region anti-sense 117 to 24
OpenProt	IFYLSRPRNKHASFYSSSNQKNKPSFHRSCHQVFPHASNRHNPNSNSYPLQQYTLRTMNHNQYYQSMLIHNHNSYSNKTRNSPLSLSPRGYPRHPDIRPASSHMTKTS HLNHMPNLSLTKRK SPHSLNLIHHSRQLRWIKPNPATQNL SMLLNYPHRMNNSSSTVQP	172aa	20kDa	within nd2 sense, frameshift 4547 to 5063
OpenProt	ICNNLLHSNTHHNRRLWQLTSSPNNRCPRYGVSPHKQHKLTLTSLSPTPARICYSGGRSRNRLNSLP SLSRELLPPWSLRRPNHLLLTPSRCLLYLRGHQFHHNNYQYKTPCHNP MPNAPLRLIRPNHSSPTSPISPSPCWHHYTTNR PQPQHLLRPRRRRRPHSMPTPILIFRSPWSLYSYPTLRNRLPYCNLLLRKKRTIWMHRYGLSYDINWLP RVYRVSTPYIYSRNRRRHTSMFHLYHNHRYPHRRQSI	259aa	31,2kDa	within cox1 sense, frameshift 6089 to 6866
OpenProt	MSSSKPHLSPPWLSSPDEATSONAWTQAHTSYSTP	35aa	3,8kDa	within nd4 sense, frameshift 11115 to 11219
OpenProt	MAESSATFTPMAPQYLSASSYTS GEAYITDHFSTQKPETS ALSSCLQL	39aa	5,2kDa	within cytb sense, frameshift 14970 to 15116
MS/MS	MGLSRIEGLFGQVVCVGGGLGMCFLVLHRAIIGMWLVCWLVLVWGALWSG SEITWLGRRSLGGLRGP LLGVMGWVLLYDRHVIGGSLCVVVQVEAY	95aa	10,3kDa	cox3 anti-sense 9462 to 9177
MS/MS	MWSLPR RLPGWPSSARMRRLRAVPRTPAHAPNNRY SVPMSLWVENSQRSANISGGEVKWLSEALDCKSKDRG	73aa	8,3kDa	non-coding region anti-sense 6061 to 5842
MS/MS	IMRMTAPVKLQGVWVRMAVTTTRAMWLIEEYAMSDFRSVCRRQ MELVMIMPHRDSTRK	57aa	6,9kDa	nd4 anti-sense 11709 to 11535
MS/MS	MRHNYNKLHLPTTNRPKIAHCMLFNQPHS PRNSHSHPNPLKLHRRSHSHNRPR AYILITILPSKCLKLRTHSQSHHPLSRTSNSTPTNSFLMTSSKPR	99aa	11,5kDa	within nd4 sense, frameshift 11557 to 11854

Table S2.1 Identification of mitochondrial smORFs and altORFs - in silico + kozak approach

Vertebrate mitochondrial genetic code

Start: ATG, ATT, ATA

Stop: TAA, TAG

AGA, AGG coding

START	STOP	LONG.	CODON I	CONTEXT	Kozak -3	Kozak +4	CODON F	FEATURE	PROTEIN SEQUENCE
49	199	50	ATT	TCTCCATGC ATT G	T	G	TAA		IWYFRLGGMHAMALRDAGAGAPYVAVSVFDSC LILLFIAPTFNITGEHTY
232	334	34	ATA	TTGTAGGAC ATA T	G	T	TAA		MMMTIECLHSHFPHRHHNKKFPPNPPSPASGH ST
448	832	128	ATT	AACTAACAC ATT T	C	T	TAG		IIFPSHSHTTNLINTTPAHPTQHTHTAANPMPRT NQTPKTPPTVYVAYLLKAMHWKCLDGLTSPHKQ MGLVLAFLALSKITHASIPVPVSSPSKSPRSKGTSI KHAAMQLKTLATPPRETAVINL
967	1195	76	ATA	CCCTCCCCA ATA A	C	A	TAG	RNR1	MKLKLTWVVKNSSWHKMDYESGFNMSEHTMA KTQTGIRYPTMLSPKQQLNQNCSPHEHYEQL KTQRTWRCFMSL
1258	1426	56	ATA	TCAGCCTAT ATA C	T	C	TAG	RNR1	MPPSSANPDEGYKVSASTHVKTLGQGVAHEVAR NGLHFLPQKTTMALMKLKGRRWI
1540	1609	23	ATT	CCCCTACGC ATT A	C	A	TAG	RNR1	IYMEETSRNMVSVLESALGRTRV
1963	2038	25	ATA	TGTAGCAAA ATA T	A	T	TAG	RNR2	MVGRFMGRGDKPTEPGDSWLSKMES
2323	2563	80	ATT	CATGAAAAC ATT T	A	T	TAA	RNR2	ILLRMSLRQIKTLNWQLTAQYLQSTNKSLLPSLST QHRHAHKERLKKVKGTRQILPRLFTKNITSSITSIR GTACPVTHV
2617	2800	61	ATA	GTTCTTAA ATA G	T	G	TAA	RNR2	MGTCMNGSTRVQLSLTFNQWNWPAREEAGM TQQDEKTLWSFNLLMQTVPNKPTGPKLPNLH
2893	3232	113	ATT	CTATACTCA ATT A	T	A	TAA	RNR2	IDPMTWPTEQVTLGMTAQSYSRVHINNRYVDL DVGSGHPDGAAGAAIKGSFVQRLKSYVIWVQTGVI QVGFYLSNSSLYERTREMRPTSQSAPPRKWHYL NLVLYPHPPKNRVC

3289	4261	324	ATT	AGAGGTTCA ATT C	T	C	TAA	TRNL1	IPLLNNMPMANLLLLLIVPILIAMAFMLMLTERKILGY MQLRKGNVVGOPYGLLQPFADAMKLFTKEPLKP ATSTITLYITAPTALTIALLLWTPMPNPLVNLN LGLLFILATSSLAVYSILWSGWASNSNYALIGALRA VAQTISYEVTLAIILLSTLLMSGSFNLSTLITTQEHL WLLLPSWPLAMMWFISTLAETNRTPFDLAEGES ELVSGFNIEYAAGPFALFFMAEYTNIIIMMNTLTTT IFLGTTYDALPELYTTYFVTKLLLLSLFLWIRTAY PRFRYDQLMHLLWKNFLPLTLALLMWWYVSMPIITI SSIPPQT
5425	5500	25	ATA	AGTTTGAAC ATA A	A	A	TAA	ND2	MQNPPHSSPHSSPLPRYSYLSPLLY
5857	5989	44	ATT	TGTCTTTAG ATT A	T	A	TAG		IYSPMLHSAILPHPHWCSPVDYSLQTTKLEHYT YSAHELES
7540	7738	66	ATA	AACCATTTC ATA C	T	C	TAA	TRND	MTLSKLNRYRLNPMYLNGTCSASRSTRRYFPYHRR AYHLSWSRPHNHFPYLLPSPVCFPNTHNKTN
7894	8311	139	ATG	TGGCCACCA ATG T	C	T	TAG	COX2	MVLNLRVHRLRRTNLQLLHTSPIIPRTRRPATPW RWQSSSTPDWSPHSYNNYITRRLALMSCPHIRLK NRCNSRTSKPNHFHRYTTGGMLRSMLWNLWSK PQFHAHRPRINSPKNLWNRARIYPMAPPLPLEP TVKLT
8335	8506	57	ATT	AAGTTAAAG ATT A	A	A	TAA	TRNK	IKRTNTSLQWNAPTKEYYRMAHHNYPHTPYTIPH HPTKNIKHKLPPTSLTKAHKNKKL
8527	9205	226	ATG	AGAACCAAAA ATG A	A	A	TAA	ATP8	MNENLFASFIAPTILGLPAAVLIILFPPLLIPTSKYLI NNRLITTQQWLIKLTQMMTMHNTKGRTWSL MLVSLIIFIATTNLLGLLPHSFTPTTQLSMNLAMAI PLWAGTVIMGFRSKIKNALAHFLPQGTPTPLIPM LVIIETISLLIQPMALAVRLTANITAGHLLMHLIGSA TLAMSTINLPSTLIIFTILILLTILEIAVALIQAYVFTLL VSLYLHDNT
9721	9802	27	ATT	TGGGTCTCT ATT T	T	T	TAG	COX3	ILPSYKQSTSSLPSPFPTASTAQHFL
9943	10003	20	ATG	ATTTTGTAG ATG G	T	G	TAA	COX3	MWFDYFCMSPSIDEGLTLLV
10789	10861	24	ATT	AACAATTAT ATT C	T	C	TAG	TRNC-comp	ITTTDMTFQKTHNLNQHNHPQPNY

11086	11410	108	ATT	AATTATAAC ATT A	A	A	TAA	ND4	IHSRTNHILYLLRNHTYPHLGYHHPMRQPARTP ERRHMLPILHPSRLPSPTHRNLHSQHPRLTKHST THSHCPRTIKLLSQQLNMTSLHNSFYSKDTSRTP LMTP
11557	11854	99	ATG	ACTATCCCT ATG G	C	G	TAA	ND4	MRHNYNKLHLPTTNRPKIAHCMLFNQPHSPRSN SHSHPNPLKLHRRSHSHNRPRAYILITILPSKLLRT HSQSHHNPLSRTSNSTPTNSFLMTSSKPR
12046	12139	31	ATT	AAAACCCTC ATT A	C	A	TAA	ND4	IHTRKHPHVHTPIPHSPPIQPQRHHYRVFLL
12310	14146	612	ATT	GCCCCAAA ATT T	A	T	TAA	TRNL2	ILVQLQMKVMTMHTTMTTLTSLIPPILTTLVNP NKKNSYPHYVKSIVASTFIISLFPTTMFMCLDQEVI ISNWHWATTQTTQLSLSFKLDYFSMMFIPVALFV TWSIMEFSLWYMNSDPNINQFFKYLLIFLITMLIL VTANNLFQLFIGWEGVGIMSFLISWWYARADA NTAAIQAILYNRIGDIGFILALAWFILHSNSWDPQ QALLNANPSLTPLLGLLAAAGKSAQLGLHPW LPSAMEGPTPVSALLHSSTMVVAGIFLLIRFHPLA ENSPLIQTTLCLGAIITLFAAVCALTONDIKKIVAF STSSQLGLMMVTIGINQPHLAFHICHTHAFFKAM LFMCSGSIHNLNNEQDIRKMGGLLKTMPSTSTS LTIGSLALAGMPFLTGFYSKDHIETANMSYTNA WALSITLIATSLTSAYSTRMILLTLTGQPRFPTLTNI NENNPILLNPIKRLAAGSLFAGFLITNNISPASPFQ TTIPLYLKLTA LAVTFLGLLTALDLNYLTNKLKMK PLCTFYFSNMLGFYPSITHRTIPYLGLLTSQNLPLLL LDLTWLEKLLPKTISQHQISTSIITSTQKGMIKLYFL SFFFPLILLLLIT
14152	14218	22	ATT	ACATAACCT ATT C	C	C	TAA		IPPSNLNYNMYTNKQCSTSNYY
14230	14386	52	ATA	TCAACGCC ATA T	C	T	TAA		MIMQSPRTNRILPNQPWPLSFMNYSASYTIKVY HNHHPIMLFHPQHQSYLHR
14524	14584	20	ATA	TAAACCCAT ATA C	C	C	TAA		MTSPKIQNNTPDHTANNQY
14665	14767	34	ATA	AAACAAAGC ATA A	A	A	TAA		MHHYSRTDYNHDQWYEKPSLYFNYKNTNDPNT QN

14797	15394	199	ATT	TAACCACTC ATT A	C	A	TAA	CYTB	IHRPPHIQHLRMMKLRLTPWRLPDPNHHRTI PSHALLTRRLNRLFNRPHHSRRKLWLNHPLPSR QWRLNLYLPLPTHRARPMLRIISLLRNLKHRHYP PACNYSNSLHRLCPPVRPNILRGHSNYKLTIRHP MHWDRPSSMNLRRLLSRQSHPHILYLSLHLALH YCSPSNTPPPILARNGIKQPPRNHLPFR
15469	15601	44	ATT	CTTAATGAC ATT A	G	A	TAA	CYTB	INTILTRPPRRPRQLYPSQPLKHPSPHQARMMFP IRLHNSPIRP
15628	15895	89	ATT	CCTTGCCCT ATT C	C	C	TAG	CYTB	ITIHPHPSNNPHPPYIQTTKHNISPTKPITLLTPSRR PPHSNLNRRTTSKLPFYHHWTSSIRTMLHNNPN PNTNYLPNWKQNTQMGLSL
15907	15997	30	ATA	TATAAACTA ATA A	C	A	TAA	TRNT	MHQCKPEMKTFFQGGQIREKVFNSTISTQS
16087	16315	76	ATG	ACAACCGCT ATG A	G	A	TAA		MYFVHYCQPPWMLYGTMTWPVHVHKNPIHI KTPSPCLQASTAINPQLSHINCNSKATPHPLGYQ QTYPPLTVHST
16333	16462	43	ATA	TTACCGTAC ATA C	T	C	TAA		MAHYSQIPSRPHGWPPSDRGPLTTILREINIPHS ATLLAPGP
116	209	31	ATG	GAGCACCT ATG C	C	C	TAA		MSQYLSLIPASSYYLSHLRSMQLANMLTKVC
281	383	34	ATA	ACAGACATC ATA C	A	C	TAA		MTKNFHQTPPPPLLATALKHISAKPQKQRTLTPA
470	584	38	ATA	CCCACTCCC ATA T	C	T	TAG		MLLISSMQPPPILPSTHTPLLTPYPEPTKQRHPP QFM
695	794	33	ATG	AGATTACAC ATG A	C	A	TAG	RNR1	MQASPFQWVHPLNHHDDQKEQASSTQQCSSKRL A
860	968	36	ATA	AACTAAGCT ATA T	G	T	TAA	RNR1	MLTPGLVNFVPATAVTRLTQVNRSRKKECFRSP PQ
1751	1862	37	ATA	GTATAGGCG ATA A	G	A	TAA	RNR2	MEIETWRNRYSTARERWKIMTKHNMARTNPYT FCMMN
1871	1934	21	ATA	TAACTAGAA ATA C	G	C	TAA	RNR2	MTLQGEPLRPPKPDELKNS
2192	2336	48	ATT	CAGCCACCA ATT A	C	A	TAA	RNR2	IKKAFKLNTHYLKKNPKHMTPELLTPNWTNLSPYRR TNVSMNMKTFSSA
2633	2705	24	ATG	CCTGTATGA ATG C	T	C	TAA	RNR2	MAPRGFSCLLLLTSEIDLVPKRRRA
2798	2903	35	ATT	CAAACCTGC ATT A	T	A	TAA	RNR2	IKNFGWGDGAEPNLRAVHAKTSPVKANYYTQLI Q
2951	3050	33	ATT	CGCAATCCT ATT T	C	T	TAA	RNR2	ILESMTMGFTTSMMLDQDIPMVQPLLKVRLFND

3236	3347	37	ATG	TTTGTTAAG ATG C	A	C	TAA	TRNL1	MAEPGNRMKCLKTLQSEVQFLFLTTPWPTSYS SSLYPF
4217	4277	20	ATG	TTATATGAT ATG C	G	C	TAA	ND1	MSPYPLQSPAFPLKPKKYVW
4325	4418	31	ATT	AACCCCTT ATT C	C	C	TAA	TRNI	ISRTMRIPIPENPKFSVPPITPHPKVRS AK
4547	5063	172	ATT	CTCGCACTG ATT T	C	T	TAA	ND2	IFYLSRPRNKHASFYSSSNQKNKPSFHRS SCHQVFP HASNRIHNPSNSYPLQQYTLRTMNH NQYQSM LIINNHSYSNKTRNSPLSLLSPRGYPR HPSDIRPA SSHMTKTSPHLNHMPNLSLTKRKPS PHSLNLIHH SRQLRWIKPNPATQNLMSLLNYPHR MNNSSSTV QP
5114	5291	59	ATT	TACTACCGC ATT C	C	C	TAG	ND2	IPTTQLKLQHDPTTISHLKQANMTNT LNSIHPP LPRRPAPANRLFAQMGHYRRIHKKQ
5411	5519	36	ATG	AAAAATAAA ATG C	A	C	TAG	ND2	MTVWYKTHPIPPHTHRPYHATPTYL PFYTNN LMEI
5573	5642	23	ATT	CAATACTTA ATT C	T	C	TAA	TRNW	ISVTAKDCKTPLCINWTQISHFN
5807	6050	81	ATG	CAATTCAAT ATG A	A	A	TAA		MKITSELVKRGLTPVFRFTVQCFTQ PFYLTPTD VRRPLTILYK PQRHWNT MPIIRMS WSPRHSS KPPYSS RAGPAR QPSR
6089	6866	259	ATT	AGCCCATGC ATT G	T	G	TAG	COX1	ICNLLHSNTHHNRRLLWQLTSSPNN RCPRYG VSPHKQH KLLTLT LSPTPA RICYS GGRSR NRLNS LPSLS RELLP PWSL RRPN HLLTP SRCL LYLR GHQF HHNNY QYK TPCH NPM PNAP LRL LIR PNH SSPT SPIS PSP SCW HHY TTNR PQP QHLL RPR RRR RPH SMP TPIL FRS PWS LYS YPT RLR NNL PYC NLL LRK KRTI WMH RYG LSY DIN WL PRV YRV STP YIY SR NRR RH TSM FHL RYH NH RYP HRR QSI
6902	7442	180	ATG	CAATATGAA ATG T	G	T	TAG	COX1	MICCSALS PRIHLS FHR RWD WHC ISK LIT RH RTT RHL VLR CS PL PC PIN RSC ICH HRR LH SL IP LRL HP RPN LR QN PF HY H HRR KSN FL PT TL SR PI RN AP TL LGL PR CM HH MK HPI CRL IH FS N SS NIN NF HD LRS LR FEA KSP NS RRT LHK PG VT MW MP PT LPH IRRT RM HKI

7586	8267	227	ATG	TATATCTTA ATG C	T	C	TAG	COX2	MAHAAQVGLQDATSPIMEELITFHDHALMIIFLI CFLVLYALFLTLLTKLTNTNISDAQEMETVWTILP AIIIVLIALPSLRILYMTDEVNDPSLTIKSIHQWY WTYEYTDYGGGLIFNSYMLPPLFLEPGDLRLLDVD NRVVLPIEAPIRMMITSQDVLHSAWVPTLGLKTD AIPGRLNQTTFTATRPGVYYGQCSEICGANHSFM PIVLELIPLKIFEMGPVFTL
8366	8570	68	ATG	TACAGTGAA ATG C	G	C	TAG	ATP8	MPQLNTTVWPTMITPMLLLTFLITQLKMLNTNY HLPPSPKPMKMKNYNKPWEKWTKICSLHSLPP QS
9251	9653	134	ATG	ACCCAGCCC ATG C	C	C	TAG	COX3	MTPNRGPLSPPNDLRPSHVISLPLHNAPHTRPTN QHTNHMPMMARCNTRKHMPPRPHHTTCPKRPS MRDNPIYYLRSFFLRRIFLSLLPLQSPYPPIRRALA PNRHHPAKSPRSPTPKHIRITRIRSINHLSSP
9794	9998	68	ATT	CGGCTCAAC ATT T	A	T	TAG	COX3	IFCSHRLPRTSRHYWLNFPHYLLHPPTNISLYIQTS LWLRSRRLMLAFRCRCGLTISVCLHLLMRVLLF
10121	10304	61	ATT	AATTATTAC ATT T	T	T	TAA	ND3	ILTTTTQRLHRKIHPLRVRLRPYIPRPRPFLHKILLSS YLLIIWSRNCPPFTPTMSPTNN
10337	10412	25	ATT	ATCCCTCTT ATT A	C	A	TAG	ND3	INHHPSPKSGLWVTTKRIRLNIGM
10502	10742	80	ATT	TATACTAGC ATT A	A	A	TAA	ND4L	IYHLTSRNTSMSLTPHILPTMPRRNNTIAPHYSYS HNPQHPLPLSQYCAYCHTSLCRLRSSGGPSPTSL NLQHMWPRLRT

10760	12278	506	ATG	CCTACTCCA ATG T	C	T	TAA	TRNC-comp	MLKLIVPTIMLLPLTWLSKKHMIWINTTTHSLIISII PLFFNQINNNLFSCSPTFSSDPLTTPLLMLTTWLL PLTIMASQRHLSSEPLSRKKLYLSMLISLQISLIMTF TATELIMFYIFFETTLIPTLAIITRWGNQPERLNAG TYFLFYTLVGSPLLIALIYTHNTLGSLNILLTLTAQ ELNSWANNLMWLAYTMAFMVKMPLYGLHL WLPKAHVEAPIAGSMVLA AVLKGGYGMMRL TLILNPLTKHMAYPFLVLSLWGMIMTSSICLRQTD LKSLIAYSSISHMALVVTAILIQTPWSFTGAVILMI AHGLTSSLLFCLANSNYERTHSRIMILSQGLQTLLP LMAFWWLLASLANLALPPTINLLGELSVLVTTFS WSNITLLLGLNMLVTALYSLYMFTTTQWGS LTH HINNMKPSFTRENTLMFMHLSPIILLSLNPDIITG FSSCKYSLTKTSDCESDNRGLRPLIYRESSQELLTH APMSNNMAFSTFKG
12824	12893	23	ATG	CCCGAGCAG ATG C	C	C	TAG	ND5	MPTQQPFKQSYTTVSAMS VSSSP
14171	14231	20	ATT	GCAATCTCA ATT C	T	C	TAA		ITMYTPTNNVQPVT TTNQRP
14297	14498	67	ATT	CCTTCATAA ATT T	T	T	TAA		IIQLPTLLKFTTTTTPSYSFTHSTNPTS IANPTKTLTK TSTPDPHASGYSSMAIAVVYPKTTIIPPK
14597	15905	436	ATA	CCCCATAA ATA G	T	G	TAA		MGEGLEENPTNPITKPTLN RNKAYIIILARTTTTTN DMKNHRCISTTRTPMTPMRKTNPLMKLINHSFI DLPTPSNISAWWNFGSLLGACLILQIT TGLFLAM HYSPDASTAFSSIAHITRDVNYGWIIRYLHANGAS MFFICLFLHIGRGLYYGSFLYSETWNIGIILLATM ATAFMGYVLPWGQMSFWGATVITNLLSAIPYIG TDLVQWIWGGYSVDSPTLTRFFTFHFILPFIIAALA TLHLLFLHETGSNNPLGITSHSDKITFHPYYTIKDAL GLLLFLLSLMTLTLFSPDLLGDPDNYTLANPLNTPP HIKPEWYFLFAYTILRSV PNKGGVLALLSILILAM IPILHMSKQQSMMFRPLSQSLYWLLAADLLITW IGGQPVSYPFTIIGQVASVLYFTTILILMPTISLIENK MLKWACPCSMN
16016	16136	40	ATT	ATTTAAACT ATT T	A	T	TAA		ILCSFMGKQIWVPPKYWLTHQQPLCISYITASHH EYCTVP

16160	16271	37	ATA	CTGTAGTAC ATA A	T	A	TAG		MKTQSTSKPPPHAYKQVQQSTLNYHTSTATPKP PLTH
16340	16475	45	ATT	ACATAGCAC ATT C	C	C	TAG		ITVKSLLVPMDDPPQMGVPWPPSSVKSMSRTR VLLSSLRAHNTWG
21	213	64	ATT	TATCACCT ATT A	C	A	TAA		INHRELSMHLVFSSGGYARDSIARRWSRSTLCRS ICLWFLPHPIIYRTYVQYYRRTYLLKCVN
219	282	21	ATG	AATTAATTA ATG T	T	T	TAA		MLVGHNNNNNWMSAQPLSTQTS
291	372	27	ATT	TAACAAAAA ATT C	A	C	TAA		ISTKPPLPRFWPQHLNTSLPNPKNKEP
390	477	29	ATT	CCTAACCAG ATT C	C	C	TAA		ISNFIFWRYALLTVTPQLTHYFPLPLPY
924	1035	37	ATA	CCCAAGTCA ATA A	T	A	TAA	RNR1	MEAGVKSVDHPLPNKAKTHLSCKKLQLTQNR RKWL
1218	1407	63	ATA	CTGTAATCG ATA A	T	A	TAA	RNR1	MNPDQPHHLLLSLYTAIFSKPWWRLQSKRKYPR KDVRSRCPWGGKKWATFSTPENYDSPYET
2052	2226	58	ATT	CAACTTTAA ATT G	T	G	TAA	RNR2	ICPQNPLNPLVNLTVSPKRNSSLDRKKPCRESKK FNTHSRPKSSHQLRKRSSSTPTT
2370	2445	25	ATT	GAACTGACA ATT A	A	A	TAA	RNR2	INSPMSTINQQVIITLVNPTQACS
2598	2667	23	ATA	AAAGGTAGC ATA T	A	T	TAA	RNR2	MITCSLNRDLYEWLHEGSAVSFY
2805	2955	50	ATT	GCATTAATA ATT C	A	C	TAG	RNR2	ISVGATSEQNPTSEQYMLRLHQSKRTTMLNWSN NLTNGTSYPRDNSAILF
3048	3174	42	ATT	TGTTCAACG ATT A	A	A	TAA	RNR2	IKVLRDLSSDRSNPGRFLSXFKFLPVRKDKRNKAY FTKRLPP
3177	3258	27	ATG	CCCCCGTAA ATG T	T	T	TAA	RNR2	MMSSQLSIMPTPTQEQGLLRWQSPVIA
3360	3633	91	ATT	CGCAATGGC ATT C	G	C	TAG	ND1	IPNAYRTKNSRLYTTTQRPQRCRPLRATTTLRWR HKTQHQRAPKTRHIYHHPLHHRPDLSSHRSSTM NPPPHTQPPGQPQRPIIYSSHL
3825	4299	158	ATT	ACACCTCTG ATT C	C	C	TAA	ND1	ITPAIMTLGHNMIYLHTSRDQPNPLRPCRRGVRT SLRLQHRMRRRPLRPILHSRMHKHYYNKHPHY NLPRNNMWRTLPLWTLHNMFCQDPTSINLPVL MNSNSMPPPIPLRPTHTPPMKKLPPTHPSITYMM CLHTHYNLQHSPSNLRNMSDKRVTLM

4470	5511	347	ATT	CCCGTACTA ATT A	C	A	TAG	ND2	INPLAQPVIYSTIFAGTLITALSSHWFFTWVGLEM NMLAFIPVLTKKMNPRSTEAAIKYFLTQATASMIL LMAILFNNMLSGQWTMTNTTNQYSSLMIMMA MAMKLGMAPFHFVWPEVTQGTPLTSGLLLLTW QKLAPISIMYQISPSLNVSLLLLSILSIMAGSWG LNQTQLRKILAYSSITHMGWMMAVLPYNPNMT ILNLTIIILTTTAFLLNLNSSTTTLLSRTWNKLTW LTPLIPSTLLSLGGLPPLTGFLPKWAIIEEFTKNNSLI IPTIMATITLLNLYFYLRLIYSTSITLLPMSNNVKMK WQFEHTKPTPFLPTLIALTLLLPISPFMLMIL
5904	7563	553	ATG	CCCCCACTG ATG T	C	T	TAG	COX1	MFADRWLFSTNHKDIGTLYLLFGAWAGVLGTAL SLLIRAELGQPGNLLGNDHIYNVIVTAHAFVMIFF MVMPIMIGGFGNWLVPLMIGAPDMAFPRMN NMSFWLLPPSLLLLASAMVEAGAGTGWTVYPP LAGNYSHPGASVDLTIFSLHLAGVSSILGAINFITTI INMKPPAMTQYQTPLFVWSVLITAVLLLLSLPVL AGITMLLTDRLNNTFFDPAGGGDPILYQHLFWF FGHPEVYILILPGFGMISHIVTYYS GKKEPFGYMG MVWAMMSIGFLGFIVWAHMHMFTVGMDVDTR AYFTSATMIIAIP TGVKVFSWLATLHGSNMKWSA AVLWALGFIFLFTVGGLTGIVLANSSLDIVLHDTYY VVAHFHYVLSMGAVFAIMGGFIHWFLPFSGYTL DQTYAKIHFTIMFIGVNLTFPQHFLGLSGMPRR YSDYPDAYTTWNILSSVGSFISLTAVMLMIFMIW EAFASKRKVLMVEEPSMNLEWLYGCPPPYHTFE EPVYMKSQRKRKESNPPKLVSSQPHGLHDFFKK LEKPFHNFVKVKL
8502	8715	71	ATT	AAATAAAAA ATT T	A	T	TAA	ATP8	IMTNPENQNERKSVRFIHC PHNPRPTRRSTDHSI SPSIDPHLQMSHQPTNHHPTMTNQTNLKNTND NHTQH
8799	8883	28	ATT	GCCTCACTC ATT A	C	A	TAA	ATP6	IYTNHPTIYKPSHGHP LMSGHSDYRLSL

9204	10764	520	ATA	CGACAACAC ATA T	C	T	TAA	ATP6	MMTHQSHAYHMKPSPWPLTGALSALLMTSGL AMWFHFHSM TLLMLGLLTNTLTMYQWWRDV TRESTYQGHHTPPVQKGLRYGMILFITSEVFFAG FFWAFYHSSLAPTPQLGGHWPPTGITPLNPLEVP LLNTSVLLASGVSITWAHHS LMENNRNQMIQAL LITILLGLYFTLLQASEYFESPFTISDGIYGSTFFVAT GFHGLHVIIGSTFLTICFIRQLMFHFTSKHHFGFEA AAWYWHFVDVVWFLYVSIYWWGSYSFSMNS TVNFQLTSFDNIQKRV MNFALILMINTLLALLMII TFWLPQLNGYMEKSTPYECGFDPMSPARVPFS MKFFLVAITFLLFDLEIALLLPLP WALQTTLNPLMV MSSLLLLILALSLAYEWLQKGLDWTELVYSLNKTN DFDSLNYDNHIYQMPLIYMNIMLAFTISLLGMLV YRSHLMSSLLCLEGMMLSLFIMATLMTLNTHSLL ANIVPIAMLVFAACEAAVGLALLVSISNTYGLDYV HNLNLLQC
11115	11220	35	ATA	TCATATTTT ATA C	T	C	TAG	ND4	MSSSKPHLSPPWLSPDEATSQNAWTQAHTSYS TP
12165	12228	21	ATT	AAACATCAG ATT T	C	T	TAA	TRNH	IVNLTTEAYDPLFTEKAHKNC
12234	12408	58	ATG	TGCTAACTC ATG C	C	C	TAA	TRNQ-comp	MPPCLTTWLSQLLKDNSYPLVLGPKNFGATPNKS NNHAHYYNHPNPDFNSPHYHPR
12429	12678	83	ATA	AAAAAACTC ATA C	C	C	TAA	ND5	MPPLCKIHCRILHYQSLPHNNIHVPRPRSYYLELT LSHNPNNPALPKLQTRLLLHNIHPCSIVRYMVHH RILTVMYKLRPKH
12747	12954	69	ATT	TAACAACCT ATT C	C	C	TAA	ND5	IPTVHRLRGRNYILLAHQLMMRPSRCQHSSHSS NPMQPYRRYRFHPRLSMIYPTLQLMRPTTNSPS KR
13014	13143	43	ATT	ATCAGCCCA ATT G	C	G	TAG	ND5	IRSPPLTPLSHRRPHPSLSPTPLKHYS CSRNLLTHPL PPPSRK
13170	13386	72	ATG	TCTAACACT ATG T	A	T	TAA	ND5	MLRRYHHSVRSSLRPYTKWHQKNRSLHFKSTRT HNSYNRHQPTTSPIPAHL YPRLLQSH TIYVLRVHH PQP
13473	13602	43	ATT	CAGCCTAGC ATT G	A	G	TAG	ND5	ISRNTFPHRFLQRP HHRNRKHIMHKRLSPIYYSH RYLPDKRL
14145	14208	21	ATA	CCTAATCAC ATA C	C	C	TAA	ND5	MTYSPEQSQLQYMHQQTMFNQ

14703	14775	24	ATG	CCACGACCA ATG T	C	T	TAA		MMWKTIVVFQLQEHQWPQYAKLTP
14967	15117	50	ATT	GAGACGTAA ATT T	T	T	TAG	CYTB	IMAESSATFTPMAPQYSLSSASSYTSGEAYITDHFS TQKPETSALSSCLQL
15387	15462	25	ATT	TCACCTCCC ATT C	C	C	TAA	CYTB	IPMKSPSTLTTQSKTPSAYFSSFSP
16062	16161	33	ATT	CACCCAAGT ATT A	A	A	TAA		IDSPINNRYVFRLLPATMNIVRYHKYLTTCTST
16194	16308	38	ATG	CCCCTCCCC ATG T	C	T	TAG		MLTSKYSNQPSTITHQLQLQSHSPSTRMPTNLPT LNST
16314	16386	24	ATA	ACATAGTAC ATA A	T	A	TAG		MKPFTVHSTLQSNPFSSPWMTPLR
-16315	-16087	76	ATG	AATGGCTTT ATG A	T	A	TAG		MYYVLLRVGRFVGLVGEGLWCSWCVMVEG WLLYLLVSMGRGFWCGLGFYVLQVVKYLWYRT MFMVAGSNVRNT
-16075	-16006	23	ATG	CGGTTGTTG ATG G	T	G	TAG		MGESMLGWYPNLLPHERTENSLN
-15139	-15052	29	ATA	CGGGAGGAC ATA C	G	C	TAG	CYTB-comp	MAYEGCCYSCKQEDNADVSGFWVEKWSVM
-15001	-14899	34	ATT	TGAGGCGCC ATT G	G	G	TAG	CYTB-comp	IGVKVADDSAMIYVSSDVGDWWKGGWGVWW VVHG
-14719	-14611	36	ATG	AATACAACG ATG T	A	T	TAA		MVFHIIGRGCSPCENNDVCFVSVECGFSNGVCG VFF
-14470	-14320	50	ATA	TGTCTTTGG ATA A	T	A	TAA		MYYSDGYWGVSWGMGVRGWGLGECFSGVSD GGRIGAVGERVWWGGGCGKL
-14221	-14152	23	ATT	TGGGCGTTG ATT G	T	G	TAG		ISSSYWLNIVCWCMYCNWDCSGE
-13564	-13387	59	ATA	GAGTAATAG ATA G	T	G	TAA	ND5-comp	MGLRRLCMMCLRFRWCGLWSRNLWGKVFLLM LGCQWWGRLKWEVWFVVLLFFEYLVHC
-13354	-13282	24	ATA	AGCACATAA ATA T	T	T	TAA	ND5-comp	MVWLWRRRRGYRCAGMLGVVGWCR
-13171	-13090	27	ATA	CGCCTAAGC ATA T	A	T	TAG	ND5-comp	MVLEFGLVGYFLLGGGSGWVRRFLLQL
-13015	-12955	20	ATT	GGAGACCTA ATT G	C	G	TAG	ND5-comp	IGLICLLLGGGLVVGWGLD
-12898	-12784	38	ATG	GGATAAATC ATG T	A	T	TAA	ND5-comp	MLRRGWNRYRRYGCMLLEWLLCWHLGRIN WWARRM
-12616	-12523	31	ATG	TAACGAACA ATG T	A	T	TAA	ND5-comp	MLQGWMLWRSSLVWSLGRAGLFLWLSVSSR
-10711	-10483	76	ATT	GTGTTGGAG ATT A	G	A	TAA	TRNY	IETSRARPTAASQAAKTSMAMGTMLAKREWVL RVMRVAMMNSDSIIPSRHSREDMRCERYTSIPR SEMVNASMMFM
-10348	-10168	60	ATG	AGGGCTAGG ATG T	A	T	TAA	ND3-comp	MMINKRDDMTISGRLLVCRAHGRGKRRRAISRN NKKVMATKKNFMEKGTRAGDMGSKPHS

-10117	-9991	42	ATA	CAAAATGTA ATA T	G	T	TAA	ND3-comp	MIISSKARRVLIKIKAKFITLFWMLSKLVNWKLTV LFMLKE
-9973	-9784	63	ATG	CATCAATAG ATG A	T	A	TAG	COX3-comp	METYNRSQTTSTKCYQAAASKPKWCLDVKWN ISWRMKQMRKVEPMMTWSPWKPVATKNVE P
-9781	-9721	20	ATG	GAGCCGTAG ATG C	T	C	TAG	COX3-comp	MPSEMVKGDSKYSEACRRVK
-9706	-9511	65	ATT	CCCAGTAAA ATT T	A	T	TAA	COX3-comp	IVMSSAWIIFRLFSIRLWVAQVIDTPDASNTD VFRSGTSRGSFSGVMPVGGQCPPNWGVGARLE W
-9466	-9229	79	ATA	TCTGAGGTA ATA A	G	A	TAG	COX3-comp	MNRIIPYRRPFWTGGVWVWPWYVLSRVTSRHH WYMVSVLVSRSRSMRSVMEWKWNHMARPEVI RRAERAPVRGHGLGFTMW
-9208	-9127	27	ATT	TGGTGGGTC ATT T	G	T	TAG	COX3-comp	IMCCRAGRGLLEVWKRRLGLRRQRFLG
-8668	-8470	66	ATT	TGATTAGTC ATT T	G	T	TAG	ATP6-comp	IVGWVWLVGCWWDIWRWGSMEGEMEWSVLR RVGLGLWGQWMKRTDFRSFWFSGFVMIFYFYG LWWGR
-7372	-7231	47	ATG	TCCAGGTTT ATG A	T	A	TAG	COX1-comp	MEGSSTIRTFRFEAKASQIMKIINITAVREMNEPT DDRMFHVYASG
-7210	-7129	27	ATT	CGTCGGGGC ATT C	G	C	TAG	COX1-comp	IPDRPRKCCGKKVRFTPMNMMVKWILA
-6997	-6685	104	ATG	TGTAGTACG ATG C	A	C	TAG	COX1-comp	MSSDEFANTMPVRPPTVKKRKMNPRAQSTAADH FMLLPWSVASQLNTLTPVGMAMIMVAEVKYAR VSTSIPTVNMWCAHTMNPRKPIDIMAQTMPM YPNGSFFPE
-6589	-6310	93	ATG	TGGTATAGA ATG G	A	G	TAG	COX1-comp	MGSPPPAGSKKVLLRLRSVSSMVMMPAARTGRD RRSRTAVIRTDQTKRGVWYWVMAGGFMLMIV VMKLMAPKMEETPARCKEKMVRSTEAPGWE
-6253	-6064	63	ATA	GCCTCCACT ATA C	A	C	TAG	COX1-comp	MADASRSRREGGKSQKLMLFMRGNAMSGAPII RGTSQLPKPPIMMGITMKKIITNAWAVTMTL
-6061	-5842	73	ATG	ACGTTGTAG ATG G	T	G	TAG	COX1-comp	MWSLPRRLPGWPSSARMRRLRAVPRTPAHAPN NRYSVPMSLWFVENSQRSANISGGEVKWLSEAL DCKSKDRG
-5806	-5710	32	ATT	GATTTTCAT ATT A	C	A	TAG		IELQIRRSSFKPAGASPAFFPGGGRSRLKPVD
-5641	-5572	23	ATT	CTTAGCTTA ATT A	T	A	TAA		IKVADLRSVDAEWGFAVLSCYRN

-5506	-5362	48	ATT	TTCTATAAG ATT T	A	T	TAG	ND2-comp	IISMKGEMGRSSVVRAMSVGRNGVGFVCSNCH FIFLLDMGSSVIEVE
-5323	-5092	77	ATG	AGGAGGGTG ATG T	G	T	TAA	ND2-comp	MVAMMVGMMRLLFFVNSSMMAHLGKKPVSG GRPPRERRVDGKGVSHVSLFQVRDSSRVVLEF KLSSRNAVVVRMM
-5089	-4498	197	ATA	ATAATATAA ATA T	T	T	TAG	ND2-comp	MVKLRMVMLGLYGRTAIHPMWVIEEYAKILRS WVWFNPPQLPAMMDKIERVRRRLTFSEGEIWW MIEMGASFCHVRRSRPDVRGVPWVTSQTQKW KGAIPSIAMAIMIINDEYWLVLVLMVHCPESML LKRMAIRRIMDAVACVRKYLMAASVERGFIFLVR TGMKASMFISRPTQVKNQCELSAVMSVPAKMV E
-4396	-4315	27	ATG	TACTTTAGG ATG G	A	G	TAA		MGCDRWHGEFWILRDGFDSHSPRNKGV
-4306	-4234	24	ATT	TAAGCTCCT ATT T	C	T	TAA	TRNI-comp	IISIKVTLTLLSDMFLRFEGECWRL
-4204	-4024	60	ATG	ATATAAGTA ATG T	G	T	TAG	ND1-comp	MLGWVVGSGFFMGGVWVGRSGIGGMLFEFMR TGRLEVGSWWQNMLCRVQGRVRHMLFLGRL
-3988	-3865	41	ATT	TGTTTGTGT ATT G	T	G	TAA	ND1-comp	IRLWRMGRRGLRRIRCWSLRLVRTPLRQGRGF GWSLLVWR
-3766	-3634	44	ATG	TTATTAGTA ATG T	G	T	TAG	ND1-comp	MLMVEWWLWGLHMRFLGLLAVRRSGRSLSL MLTLIRGLSKRLG
-3616	-3505	37	ATA	CTAGAATAA ATA G	T	G	TAG	ND1-comp	MGGLGWGWPGGWVWGGGFMVEERWWELR SGRWCRGWW
-3502	-3373	43	ATG	TGATGGTAG ATG G	T	G	TAA	ND1-comp	MWRVLGALWWRVLRQRRVVARRGLQRW GLCVVVYSLEFFVR
-3361	-3301	20	ATG	GCATTAGGA ATG C	G	C	TAA	ND1-comp	MPLRLEWVQWGVGGWPWVCC
-3253	-3151	34	ATT	TTTTATGCG ATT C	G	C	TAG	TRNL1-comp	ITGLCHLNKPCSWVGVGMMLSWDDIYGRRFV K
-2887	-2782	35	ATA	CAATTGAGT ATA T	A	T	TAG	RNR2-comp	MVVRFDWWSLSMYCSEVGFCEVAPTEIFNAGL VV
-2731	-2668	21	ATA	TAAAGCTCC ATA G	T	G	TAA	RNR2-comp	MGSSRLAVLCPPLHGQVNFTG
-1861	-1705	52	ATT	TTCTAGTTA ATT A	T	A	TAG	RNR2-comp	IHYAEGMGVSPCYIMLGYNFSSFCGTMSIAPGF NFYRLYFIWVNLAKVWV
-1267	-1159	36	ATG	TTGCTGAAG ATG C	A	C	TAA	RNR1-comp	MAVYRLSKRWWGWSGFIDYRTGSSRGMWSTA RSFEF

-673	-532	47	ATA	TAAGAGCTA ATA A	C	A	TAG	RNR1-comp	MERLGNLNFVYGVMMWARLNIFSVLLWGGKLHKL WGVSLGFGWFGVWG
-292	-214	26	ATT	TTGGTGGAA ATT T	G	T	TAA		IFCYDVCVESGCADIQLLLLCPTSIN
-157	-13	48	ATA	GTGCGATAA ATA T	T	T	TAG		MMGWGRNQRQMLRHRVLRQLRAMLRSRYP PDENTKCMESSREWLMGW
-16500	-16326	58	ATG	AGGAACCAG ATG C	C	C	TAA		MSDTVHFSYPQVLWARSEESSTLVRDIDFTEDGG QGTPIWGGSSMGTRRDLTVMCYVR
-16161	-16011	50	ATG	TGGGTTTTT ATG A	T	A	TAA		MYYRWSSIYGTQYSWWLAVMYEMHSGCWW VSQYLGGTQICFPMKEQRMV
-16008	-15906	34	ATT	ATAGTTTAA ATT G	T	G	TAG		IRILALGANGGVKDFSDLSLEKGFHLRFTRLVY
-15900	-15780	40	ATA	TATTAGTTT ATA T	T	T	TAG	TRNT-comp	MLQGOAHLNILSIREMVGIRIRIVVKYSTDATCP MMVKG
-15693	-15579	38	ATT	GGGCGAAAT ATT T	A	T	TAG	CYTB-comp	IMLCCLDMWRMGIIARMRMDSNRARTPPSLLG TDRRIV
-15399	-15249	50	ATT	TGGAAGGTG ATT T	G	T	TAG	CYTB-comp	ILSEWEVIPRGLFDPVSCKNRRWSVARAAMMKG KMKWKVKNRVRVGLSTE
-15237	-15138	33	ATT	CCTCCTCAG ATT A	C	A	TAG	CYTB-comp	IHWTRSVPMYGMADSKFVITVAPQNDIWPGR T
-15132	-15072	20	ATG	ACATAGCCT ATG A	C	A	TAG	CYTB-comp	MKAVAMVASRRMMPMFQVSE
-14904	-14139	255	ATG	GAGTAGTGC ATG C	T	C	TAG	CYTB-comp	MARNSPVVIWRIRQAPRSEPKFHHAEMLDGVG RSMNEWLINFIRGLVLRIGVIGLVVEMQRWFF MSLVVVVVRARMMMYALFLLSVGLVMGFVGF SKPSPIYGLVLIVSGVVCVILNFGGGYMGLMV FLIYLGGMVVFYTTAMAIEEYPEAWGSGVEV LVSVLVGLAMEVGLVLWVKEYDGVVVVNFNSV GSWMIYEGEGSGLIREDPIGAGALYDYGRWLVV VTGWTLFVGVYVIVIEIARGNRLCD
-13953	-13887	22	ATA	AAGGCCTAG ATA G	T	G	TAG	ND5-comp	MGDCAVCDARVESEYVGEMKCA
-12696	-12582	38	ATA	GATGAGTAG ATA T	T	T	TAG	ND5-comp	MFEELINVVVVVYMSQWFEYDGPCNEQCYRD EYYGEVV
-12441	-12381	20	ATA	GGATTTTAC ATA T	T	T	TAG	ND5-comp	MMGVVWFFVRVNEGKDGGN
-12291	-12204	29	ATG	CTAAGACCA ATG A	C	A	TAA	TRNL2-comp	MDSCYPLKVEKAMLLDMGAWVSSSCELSR

-12201	-11985	72	ATA	TCTCGGTAA ATA G	T	G	TAG	TRNQ	MRGRKPLLSQSDVLVKLYLQEENPVMMSGLR DRRRMGDRCMNMRVFSRVNEGFMLLMWWV SEPHCVVNM
-11958	-11760	66	ATG	GTGACTAGT ATG T	A	T	TAG	ND4-comp	MLSPVSRVMFDQENVVTSTESSPSRLMVGGA RLARLARSHQKAISGSRVWSPWERIMMRLWVR S
-11709	-11535	58	ATT	CCGTGGGCG ATT T	G	T	TAG	ND4-comp	IMRMTAPVKLQGVWMRMAVTTRAMWLIEEYA MSDFRSVCRRQMEVMIMPHRDSTRKG
-11343	-11256	29	ATT	GCTAGTCAT ATT A	C	A	TAA	ND4-comp	IKLLAQEFDSSWAVRVSSRMFSEPRVLWV
-11109	-11043	22	ATG	ATATAAAAT ATG T	A	T	TAG	ND4-comp	MISSVAVNVMIKEICREISMER
-10992	-10650	114	ATG	TGGCTTGCC ATG T	G	T	TAG	ND4-comp	MIVRGRSQVVSIRRGVVRGSEEKVGELNRLLI WLKNSRGMMLMIRLWVVVLIQIMCFLESHVSG SNMIVGTISFSIGVGLGYVRS LGHMCWRLRLVGL GPPLRRRQRLVWQ
-10584	-10506	26	ATA	TGAACAGCG ATA T	G	T	TAA	ND4L-comp	MVLFLLGMVGRMWGVSDMLVFLEVRW
-10443	-10323	40	ATG	CATAATTTA ATG G	T	G	TAA	TRNR-comp	MSRNHSFCLNYMPIRFSLILFVVTHRPDLGLGW WLMRGMT
-9957	-9888	23	ATA	CATACAGAA ATA T	G	T	TAA	COX3-comp	MVKPHLQNASIRRRRSQSDVWM
-9795	-9726	23	ATG	CTACAAAAA ATG T	A	T	TAA	COX3-comp	MLSRRRCRRKWWRETRSTLRLVGG
-9633	-9531	34	ATA	AGGTGATTG ATA T	T	T	TAG	COX3-comp	MLLMRVMRMCLGVLLGLDLAGWCLLGASALLI GG
-9462	-9177	95	ATA	AGGTAATAA ATA G	T	G	TAG	COX3-comp	MGLSRIEGLFGQVVCGLGMCFLVLHRAIIGMW LVCWLVLVW GALWSGSEITWLGRRSLGGLRG PLLGVMGWVLLYDRHVIGGSLCVVQVEAY
-8448	-8358	30	ATT	GTGTTAAT ATT T	A	T	TAA	ATP8-comp	IFSWVMRNSVRSMGVIMVGH TVVFSWGISL
-8247	-8163	28	ATT	ACGGGCCCT ATT C	C	C	TAG	COX2-comp	ISKIFRGINSRTMGMKLFAPQISEHWP
-8109	-7947	54	ATT	CGTCCGGGA ATT C	G	C	TAG	COX2-comp	IASVFKPNVGTAECKTSCDVIIIMRMGASIGSTTR LSTSRSRSPGSRNNGGSM
-7833	-7569	88	ATG	ATGTAAAGG ATG G	A	G	TAG	COX2-comp	MRRDGRAMRTRMMAGRMVQTVSISWASEML VLVSFVSVRKRAYRTRKQMRKMIMRAWSWKV MSSSMMGEVASCRIPTCAACA IKMYRI
-7536	-7443	31	ATG	AGTTATGAA ATG T	G	T	TAG	TRND-comp	MVFLMPFWKSHGGHGVGLKPALGG SIPSFFV

-7440	-7149	97	ATT	TTTGTCTAG ATT T	T	T	TAG	COX1-comp	ILCMRVLRMCGRVGGIHMVTPGLWRVLLLLGLF ASKRRLKSWKLLMLLLEKWMQLMMGCFM WCMHRGSPSNVGAFRMGRESVVGKLDLRRW MW
-7101	-6867	78	ATA	AGCCTGAGA ATA G	A	G	TAA	COX1-comp	MGEISEWSLLWWQMQLLLMGHSGSGLQRSTC RVVRCLVMSLLMQCQSGHLRWKERWILGLRAL QQIISYCFRGVWRVS
-6705	-6642	21	ATG	TGTATCCAA ATG T	C	T	TAA	COX1-comp	MVLFFRSSKLQYGRFRSLVG
-6606	-6525	27	ATA	AAAATCAGA ATA G	A	G	TAG	COX1-comp	MGVGMEWGLLLRRRGRWWCWGCGLLVV
-6495	-6408	29	ATA	CTGGGAGAG ATA G	G	G	TAA	COX1-comp	MGEVGLLWLGRIRRRGAFGIGLWQGVLYW
-6246	-6072	58	ATG	CTATAGCAG ATG G	C	G	TAA	COX1-comp	MRAGVGEREVRVRSCLCCLGETPYRGHRLLGELV SCQSLRLWVLLWRRLLQMHGLWR
-5814	-5694	40	ATT	TCCGAGGTG ATT T	G	T	TAA		IFMLNCKFEEAASNLPGLLPPFFPAAGEVDWSQL IRVLSC
-5628	-5520	36	ATT	AAGTGGCTG ATT G	C	G	TAA		ICVQLMQSGVLQSLAVTEIKYCNLLRALKALGLYL T
-5259	-5109	50	ATT	TAATGGCCC ATT G	C	G	TAG	ND2-comp	IWAKSRLAGAGLLGRGGWMELRVLMMLACFRC EMVVGSWCWLSWVVGMR
-5010	-4824	62	ATG	TTGAGGAGT ATG T	A	T	TAA	ND2-comp	MLRFCVAGFGLIHLNCLLWWMRLREWGEGLRL VRERFGMWLRWGLVFMWEEAGRMSEGCLG
-4668	-4590	26	ATG	GGATTATGG ATG G	T	G	TAA	ND2-comp	MRLLAGWNTWWQLLWNEGLFFWLELE
-4434	-4287	49	ATG	TTTCGGGGT ATG G	G	G	TAA	TRNM-comp	MGPmayLADLTGwGVMGGTENFGFSGMGSI LMVLEMrgfKLLlFTLSK
-3732	-3639	31	ATA	GGTGACTTC ATA G	T	G	TAG	ND1-comp	MWDCLGYCSQCADQGVVWVWCSPWSEDWV NG
-3186	-3105	27	ATG	TAAGTTGAG ATG T	G	T	TAG	RNR2-comp	MMSFTGEGALWSRPYFSCPFVQGGIWX
-3006	-2931	25	ATG	ACCATCGGG ATG C	G	C	TAG	RNR2-comp	MSWSNIEVVNPIVDMDSRMGLRCYP
-2679	-2487	64	ATT	GGCAGGTCA ATT C	T	C	TAA	RNR2-comp	ISLVKSKRQLNPRGAIHTGPYLRNKWLCYLCTVRV PRPLNMCHWAGGASNTGDARGDVFGKQAG
-1824	-1731	31	ATA	GTCCTTGCT ATA T	G	T	TAA	RNR2-comp	MLCLVMIFHLSLAVLYLLRQVSISIAATLFG
-1506	-1374	44	ATA	CTTTGAAGT ATA T	A	T	TAG	RNR1-comp	MLEEGDGRCVRASGPCSTKHSTLSLLLNPSTLKF HKGYSFLG
-1257	-1116	47	ATA	TGGCGGTAT ATA G	T	G	TAA	RNR1-comp	MGWARGGEVDRGLSITEQAPLEGYEAPPGLSF KLWLWVFWRAVLLI

-582	-402	60	ATA	GTAAGCTAC ATA A	T	A	TAA	TRNF-comp	MNCGGCLWGLVGSYGVS SGVCVLGRMGGGC IDEISSMGVGGENNVLVGGWLLKVHTAKR
-342	-210	44	ATG	TTGGCAGAG ATG G	G	G	TAA		MCLSAVARSGGGGVVWKFVMMMSVWKVAVQ TFNCYYYVLQALIN
-117	-24	31	ATA	TACTGCGAC ATA G	G	G	TAA		MGCSGSSVSQCVRVHTPQTKMPNAWRAPVSG
-16562	-16343	73	ATG	TCCATCGTG ATG C	G	C	TAA		MSYLRGTCGLFRLYDPEVGTTRCRMQFTLATPKCY GPGARRVALLCGMLISRRMVVKGPLSEGGHPW GREGIWL
-16238	-16103	45	ATG	TTGCAGTTG ATG G	T	G	TAA		MCDSWGLI AVLACKHGEGLVMWIGFLCTTGGQ VFMVPYNIHGGWQ
-16088	-16016	24	ATA	ACGAAATAC ATA C	T	C	TAG		MAVVDGWVNTWVVVPKSASPWKNRE
-15983	-15851	44	ATG	TGGGTGCTA ATG T	C	T	TAG		MVELKTFSLICPWKKVFISGLQDWCSISLYYKDRPI WVFCFQLGR
-15713	-15602	37	ATT	AATAAAGTG ATT G	G	G	TAG	CYTB-comp	IGLVGEMLCFVWVMYGGWGLLLGWGWMVM GQGRLLVC
-15572	-15311	87	ATA	TGTAGGCGA ATA G	C	G	TAA	CYTB-comp	MGNIIRAWCGEGCLRGWLGYNCLGRLGGLVRM VLMSLRREGREVSRLWLCSKGRWFYRNGR WFLGGCLIPFRARMGGGVLLGLQ
-15014	-14789	75	ATA	AGATAAAGA ATA T	A	T	TAA	CYTB-comp	MLRRHWREGSGWFSHNLRLWCGRLMKRRLR RLVSSAWLGMVLWWFGGSGRRQGVSRSFIMR RCWMGWGGRWMSG
-14468	-14396	24	ATA	TCTTTGGAT ATA T	G	T	TAG		MLQRWLLRSILRHGGQGLRSWWVF
-14294	-14219	25	ATG	GAATAATTT ATG A	T	A	TAG		MKERGQGWFGRIILLVRGLCMIMGVD
-14087	-13952	45	ATT	TAAAGTTTA ATT T	T	T	TAG	ND5-comp	IMPFWVEVMMEVEIWCCEIVLGNFSQVRSRR SRGRFWLVRPR
-13781	-13601	60	ATT	TAGAGGGGG ATT T	G	T	TAG	ND5-comp	IVVWKG DAGEMLLVMRNPANRLPAARRLMGFS RVGLFSLMLVRVGRGWVVRRIIRVL
-13577	-13505	24	ATG	GAGGTAGCG ATG G	G	G	TAG	ND5-comp	MRVMDRAQAFVYDMFAVSMMSWSLE

-13484	-12587	299	ATT	AGGAAAGGT ATT C	G	C	TAG	ND5-comp	IPANARLPMVREVEVRGMVLSSPPIFRMSCSLLR LWMMDPEHMNSMALKKAWVQMCRNARCG WLMPIVTIMSPSWLEVEKATIFLMSFCVRAQTA ANRVVMAPKHSVRVWISGLFSARGWKRMSKKI PATTMVLEWSRAETGVGPSMAEGSQGWRPN WADLPAAARRRPPSSGVRLGLAFRRRAICCGSHELE CRMNHAKARMKPMSPMRLYRIAWMAAVLASA RAYHQLMSKKDMIPTSPQMNSWNRLAVTKIS MVIRKMSRYLKNWLMFGSEFMYHSENSMMDH VTNNATGMNIMEK
-12524	-12440	28	ATA	CAGTTCGAG ATA T	G	T	TAA	ND5-comp	MMTSWSRHMNIVVGKRLMMKVDATMDFT
-12392	-12299	31	ATG	GTGGTAAGG ATG G	A	G	TAA	ND5-comp	MGGIREVRVRVVMVVCMVITFIWSCTKIFGA
-12101	-12029	24	ATA	GGTTGAGGG ATA G	G	G	TAA	ND4-comp	MGGEWGMGVWTWGCFLVWMRVLCC
-11558	-11411	49	ATA	TTATGCCTC ATA G	C	G	TAG	ND4-comp	MGMVQGRGRLCVLSGGWEVWVWGVLYHSRLV LRVLRQVLLTQRWGLRHGL
-11345	-11153	64	ATA	AAGCTAGTC ATA T	G	T	TAG	ND4-comp	MLSCWLRSLMVLGQWEVWVECLVSLGCCECKL VRWVGEGLLGCRCMGMCLRSGLAGCLIGW W
-11027	-10967	20	ATA	TTTTTCGTG ATA T	G	T	TAG	ND4-comp	MVVHWMMSGVGLPWLWGVGVR
-9935	-9875	20	ATG	ATCTACAAA ATG C	A	C	TAG	COX3-comp	MPVSGGGFEAKVMFGCKVKY
-9152	-8420	244	ATT	TAGGCTTGG ATT A	T	A	TAA	ATP6-comp	IKATAISRMVSRIRIVKMMSVEGRMLMVDIARVAL PIRCMSRWPVAVMLAVRRTARAIGWMSRLMVS MMTSMGMRGVGVPCGKKWARAFLILERKPMI TVPAHKGMAMARFMDSWVGVNEWGRSPRR LVVAMKMIKDTSMRDQVRPLVLCMVIICFEVSLI SHCWVVISRLLMRYLEVGINRGGNRMISTAAGR PRIVGAMNEANRFSFILVLRVCYNFLFLWALVRE VGGSLCLMFLVGVWGMV
-8258	-8144	38	ATA	ATAGGGTAA ATA G	T	G	TAG	COX2-comp	MRALFQRFLGELILGRWAWNCGLLHRFQSIDRS MPPVV
-7964	-7871	31	ATA	GTTCTAGGA ATA T	G	T	TAA	COX2-comp	MMGEVCRSWRLVRRSRCTRRFSTIGGQLIWW
-7505	-7388	39	ATG	GAAAAAGTC ATG A	G	A	TAG		MEAMGLAWNQLWGVRFPLFSRFVYVYGFECV VGWGASM

-6896	-6752	48	ATT	TCATTTTCAT ATT C	C	C	TAG	COX1-comp	IASVECGESAKYFDAGGSDDDYGSGGEMCSCVY VYSYCKYMCVSHDKP
-5801	-5729	24	ATT	TCATATTGA ATT C	T	C	TAG		IANSKKQLQTCRGRFSRRLFSRRREK
-4742	-4682	20	ATT	TGAGTATTG ATT G	T	G	TAG	ND2-comp	IGSIGYGSLSGEYIVEEDSY
-4247	-4151	32	ATG	TGAGGGGGA ATG T	G	T	TAG	ND1-comp	MLEIVMGMEYHMSNARVSGRKFHRRCMSW S
-4028	-3686	114	ATT	CCTAGGAAG ATT T	A	T	TAG	ND1-comp	IVVVRVFIMMMFVYSAMKNRAKGAAYSMLKP ETSSDPSARSKGVRLVSASVEMNHIMAKGHDG RSNQRCSCVVMRVERLKEPLISNVDSRMMARVT SYEIVWATARSAPIRA
-3620	-3518	34	ATA	GTGGCTAGA ATA A	A	A	TAG	ND1-comp	MNRRPRLRLTRGLMGRGVHSRRAMVRAKVG AVM
-3509	-3434	25	ATG	TAGAGGGTG ATG T	G	T	TAG	ND1-comp	MVDVAGFRGSLVKSFMASAKGCSSP
-3386	-3266	40	ATT	TAGCCTAGA ATT T	A	T	TAA	ND1-comp	IFRSVSIRNAIAIRMGTMRSRRLAMGMLLRGIE PLTVKF
-2750	-2615	45	ATT	ACTGTTTGC ATT A	T	A	TAA	RNR2-comp	INKLKLHRVFSSCCVMPASSRAGQFHWLKVRS WTLVEPFMQVPI
-2531	-2303	76	ATA	GTGCCTCTA ATA T	C	T	TAA	RNR2-comp	MLVMLEVMFLVNRGKICRVPTFFNLSLWACL CWVDSEGNNDLLVDCRYWAVNCQFSVLIWRR L MRRRMFSCYLY
-2279	-2195	28	ATA	TATAGGGTG ATA A	G	A	TAA	RNR2-comp	MDWSNWWVWGVQLYVWDFLGSGCWAWTLS
-2168	-2090	26	ATG	AGGCCTACT ATG G	A	G	TAA	RNR2-comp	MGVKFFTLSTRFFPSVQRAVPLWTNS
-1520	-1427	31	ATG	TTTAGTTAA ATG C	T	C	TAA	RNR1-comp	MSFEVYLRRVTGGVYALQGPVQLSTLLLVYC
-1400	-1292	36	ATA	TTAAGTTTC ATA G	T	G	TAG	RNR1-comp	MRAIVVFWGRKCSPLATSWATPWPVFTWVL ALTL
-1091	-1022	23	ATA	GGGCTAAGC ATA T	A	T	TAG	RNR1-comp	MVGYLIPVWVLAIVCSDMLKPLS
-1016	-950	22	ATT	TCGTAGTCT ATT T	T	T	TAA	RNR1-comp	ILCQLEFFTTQVSFSFIGEVI
-773	-590	61	ATT	TTGAGCTGC ATT C	T	C	TAA	RNR1-comp	IAACLMLVPFDRGDLEGELTGTGMLACVILLRAN RKARTKPICLWGDVSPSKHFQCIARR
-539	-443	32	ATG	GTTCGGGGT ATG G	G	G	TAG		MGLAAVCVCWVGWAGVVLMLRVVWEWEGK MMC
-416	-335	27	ATA	TAAAAGTGC ATA C	T	C	TAA		MPPKDKIWNLVRLVLGFFVFGVWQRCV
-179	-116	21	ATA	TCGCCTGTA ATA T	G	T	TAG		MLNVGAMNNRMRQESKTDAT

Table S2.2 14 chosen candidates from the kozak approach

PROTEIN SEQUENCE	FEATURE	tBLASTn excluding Homo/Pan/Gorilla	tBLASTn against Mus musculus	Nuclear hits with >70% similarity	Antibody comments
MRIEPIPENPKFSVPPITPHPKVRSAK	Non-coding region + strand	yes	no	25	
MWRTLPLWTLHNMFCHQDPTSNLPVLMNS NSMPPIPLRPTHTPPMKKLPTTHPSITYMMC LHTHYNLQHSPSNLRNMSDKRVTLM E	Within nd1 + strand	yes	no	3	
MGSPPPAGSKKVLRRLRSVSSMVMPAA	Cox1 complementary strand	no	no	2	
MSVGRNGVGFVCSNCHFIFLLDMGSSVIEV E	Nd2 complementary strand	yes	no	6	Too hydrophobic for immunisation
MGEISEWLLWWQMQLLLMGHSGSGLQR STCRVVRCLVMSLLMQCQSGHLRWKERWIL GLRALQQIISYCFRGVWRVS	Cox1 complementary strand	yes	no	3	
MGVGM EWGLLLRRRRRWCWGCLLVV	Cox1 complementary strand	no	no	2	
MGEVGLLWLGRIRRRGAFGIGLWQGVLYW	Cox1 complementary strand	no	no	3	
MGPMAYLADLT LGWGV MGGTENFGFSG MGSILMVLEM RGFKLLLF T LSK	trnM complementary strand	yes	yes	14	Too hydrophobic for immunisation
MGGGCIDEISSMGVGGENNVLVGGWLLKV HTAKR	Non-coding - strand	no	no	0	

MCLSAVARSGGGGVVWKFFVMMSVWKV AVQTFNCYYYVLQALIN	Non-coding - strand	no	no	1	Too hydrophobic for immunisation
MGCSGSSVSQCYRVHTPQTKMPNAWRAP VSGR	Non-coding - strand	yes	no	1	
MGGEWGMGVWTWGCFLVWMRVLCC	Nd4 complementary strand	no	no	1	
MGVKFFTLSTRFFPSVQRAVPLWTNS	16Srna complementary strand SHLP2 (Cobb et al. 2016)	yes	no	SHLP2	
MGLAAVCVCWVGWAGVVLMLRVVWEWE GKMMC	Non-coding - strand	no	no	0	

**ANTIBODY
PRODUCTION**

START	LONG.	CODON I	CONTEXT	Kozak -3	Kozak +4	CODON F	FEATURE
-117	31	ATA	TACTGCG AC ATA G	G	G	TAA	MGCSGSSVS QCYRVHTPQT KMPNAWRAP VSGR

Table S3. Identification of mitochondrial smORFs and altORFs - OpenProt approach and sequences selected for antibody production

Genomic regions without an associated transcript will not be in OpenProt (i.e. regions with no gene name, or -comp)

Genes	Present_in_OP
ATP6	1
ATP8	1
COX1	1
COX2	1
COX3	1
CYTB	1
ND1	1
ND2	1
ND3	1
ND4	1
ND4L	1
ND5	1
RNR1	1
RNR2	1
TRND	1
TRNH	1
TRNI	1
TRNK	1
TRNL1	1
TRNL2	1
TRNQ	1
TRNS1	1
TRNT	1
TRNW	1
TRNY	1

OpenProt predictions for genes above - AltProt (novel protein from non canonical ORF) are in green - sorted by MS detection - more info on each protein on OpenProt website

Identification of mitochondrial smORFs and altORFs - OpenProt approach

OpenProt release 1.3 - August 30 2018

Protein accession	Protein Type	Species	Protein length (a.a.)	Molecular weight (kDa)	Isoelectric point	Gene symbol	Transcript accession	Type	MS score	TE score	Domains	Orthology Across 10 Species = Species name : id %
P00403	RefProt	Homo sapiens	227	25.56	4.44	COX2	COX2	CDS	335	0	25	SC:44.09,DR:67.73,MM:73.64,RN:74.09,DM:58.45,CE:44.5,BT:74.55,PT:97.8,OA:73.64
P03915	RefProt	Homo sapiens	603	67.03	9.32	ND5	ND5	CDS	110	0	62	DR:55.18,MM:66.13,RN:64.68,DM:37.2,CE:35.31,BT:71.23,PT:93.27,OA:71.6
P03928	RefProt	Homo sapiens	68	7.99	10.56	ATP8	ATP8	CDS	88	0	7	MM:46.27,RN:47.76,BT:55.17,PT:94.12,OA:49.23
YP_003024035.1	RefProt	Homo sapiens	459	51.58	9.67	ND4	ND4	CDS	72	0	45	DR:60.22,MM:67.03,RN:67.9,DM:42.96,CE:32.93,BT:74.56,PT:94.99,OA:75.66
YP_003024026.1	RefProt	Homo sapiens	318	35.66	6.53	ND1	ND1	CDS	71	0	32	DR:67.2,MM:78.33,RN:77.89,DM:49.34,CE:35.33,BT:78.55,PT:94.65,OA:77.99
P00395	RefProt	Homo sapiens	513	57.04	6.7	COX1	COX1	CDS	66	0	56	SC:66.05,DR:85.57,MM:91.02,RN:90.62,DM:76.24,CE:60.67,BT:90.82,PT:98.83,OA:91.99
YP_003024031.1	RefProt	Homo sapiens	226	24.82	10.68	ATP6	ATP6	CDS	57	0	30	SC:35.58,DR:52.86,MM:75.66,RN:75.66,DM:37.95,CE:32.69,BT:77.88,PT:94.25,OA:77.43
YP_003024032.1	RefProt	Homo sapiens	261	29.95	7.34	COX3	COX3	CDS	49	0	28	SC:43.85,DR:80.84,MM:86.97,RN:87.74,DM:65.25,CE:43.31,BT:87.69,PT:97.32,OA:86.97

YP_003024027.1	RefProt	Homo sapiens	347	38.96	10.3	ND2	ND2	CDS	37	0	39	DR:44.8,MM:58.06, RN:57.76,DM: 41.64,BT:63.93,PT: 96.81,OA:63.85
YP_003024038.1	RefProt	Homo sapiens	380	42.72	8.22	CYTB	CYTB	CDS	28	0	37	SC:50.27,DR:70.26, MM:78.57,RN: 78.63,DM:62.81,CE: 44.87,BT:78.89,PT: 93.67,OA:77.84
YP_003024033.1	RefProt	Homo sapiens	115	13.19	4.08	ND3	ND3	CDS	17	0	13	DR:57.52,MM: 70.79,RN:73.03, DM:47.62,BT:73.91, PT:94.78,OA:73.04
IP_306405	AltProt	Homo sapiens	61	6.38	11.88	RNR1	RNR1	rRNA	1	0	8	PT:90
IP_306387	AltProt	Homo sapiens	32	3.66	11.93	ND2	ND2	CDS	0	0	0	PT:65.62
IP_306389	AltProt	Homo sapiens	33	3.7	9.51	COX1	COX1	CDS	0	0	0	
IP_306392	AltProt	Homo sapiens	39	4.56	10.98	ATP8	ATP8	CDS	0	0	0	PT:79.49
IP_306398	AltProt	Homo sapiens	35	3.82	6.5	ND4	ND4	CDS	0	0	0	PT:85.71
IP_306399	AltProt	Homo sapiens	59	6.91	12.81	ND4	ND4	CDS	0	0	0	PT:82.76
IP_306403	AltProt	Homo sapiens	49	5.22	4.14	CYTB	CYTB	CDS	0	0	0	PT:74.29
IP_306404	AltProt	Homo sapiens	31	3.43	10.22	RNR1	RNR1	rRNA	0	0	0	
IP_306406	AltProt	Homo sapiens	37	4.35	9.95	RNR1	RNR1	rRNA	0	0	0	PT:86.49
IP_306407	AltProt	Homo sapiens	32	3.27	6.5	RNR1	RNR1	rRNA	0	0	0	PT:93.75
IP_306408	AltProt	Homo sapiens	33	3.81	8.56	RNR1	RNR1	rRNA	0	0	0	
IP_306409	AltProt	Homo sapiens	32	3.71	9.44	RNR2	RNR2	rRNA	0	0	0	
IP_306410	AltProt	Homo sapiens	33	3.83	10.9	RNR2	RNR2	rRNA	0	0	0	
IP_306411	AltProt	Homo sapiens	29	3.33	4.31	RNR2	RNR2	rRNA	0	0	0	BT:93.1,PT:100,OA: 93.1
YP_003024034.1	RefProt	Homo sapiens	98	10.74	6.2	ND4L	ND4L	CDS	0	0	12	DR:55.68,MM: 66.33,RN:69.15, DM:36.67,BT:73.47, PT:98.98,OA:76.53

ANTIBODY PRODUCTION (antigen in red)

IP_306387	AltProt	Homo sapiens	32	3.66	11.93	ND2	ND2	CDS	0	0	0	PT:65.62	MTKTSPHLN HMPNLSLT KRKPSPHSL NLIHHS
IP_306389	AltProt	Homo sapiens	33	3.7	9.51	COX1	COX1	CDS	0	0	0		MPNAPLRLIR PNHSSPTSPIS PSPSCWHHY TTN
IP_306398	AltProt	Homo sapiens	35	3.82	6.5	ND4	ND4	CDS	0	0	0	PT:85.71	MSSSKPHLSP PWLSSPDEAT SQNAWTQA HTSYSTP
IP_306403	AltProt	Homo sapiens	49	5.22	4.14	CYTB	CYTB	CDS	0	0	0	PT:74.29	MAESSATFTP MAPQYLSA SSYTSGEAYI TDHFSTQK PETSALSSCL QL

Table S4. Identification of mitochondrial smORFs and altORFs - MS approach and sequences selected for antibody production

Legend

Tables : PepQuery outputs

Prot_Accession	Accession number for the protein (unique ID)
Description	Accession number for the protein (output by pepquery)
Prot_Seq	Protein sequence
PepQuery_hits	Number of peptide spectrum matches (PSMs) reported by PepQuery = number of PSMs with higher score than with any reference peptides without PTMs
Confident_hits	Number of confident PSMs reported by PepQuery = number of PSMs with higher score than with any reference peptides with or without any PTMs
Best_Score	Score of the best confident PSM
pvalue	p value of the best confident PSM
Peptide_Sequence	Peptide sequence yielding the best PSM
Peptide_Check_Mass	Theoretical mass of the peptide (<i>may vary from experimental mass based on precursor charge, PTMs, and error</i>)
Exp_Pep_Mass	Observed mass of the peptide (experimental)
Error_mass_ppm	Error of the peptide mass in ppm
Validation	#N/A if no confident PSMs; No if no confident PSMs with an error less than 4,5 ppm; Yes if confident PSM with an error less than 4,5 ppm

Table 1	PepQuery results_all PSMs reported by PepQuery
Table 2	PepQuery results_best PSMs for each protein
Table 3	PepQuery results_Filtered list based on validation (only "Yes")
Table 4	PepQuery results_For all 14 previously selected candidates

Best PSM for each protein

Prot_Accession	Description	Prot_Seq	PepQuery_hits	Confident_hits	Best_Score	pvalue	Peptide_Sequence	Peptide_Check_Mass	Exp_Pep_Mass	Error_mass_ppm	Validation
SB_0001	SB_0001	IWYFRLGGMHA MALRDAGAGAPY VAVSVFDSCLILLFI APTFNITGEHTY	37	4	18.28136369	0.001020408163	LGGMHAMALR	1055.536813	1215.632296	1.251744129	Yes
SB_0002	SB_0002	MMMMTIECLHSHF PHRHHNKKFPPN PPSPASGHST	31	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0003	SB_0003	IIFPSHSTTNLINT TPAHPTQHTHTA ANPMPRTNQTPK TPPTVYVAYLLKA MHWKCLDGLTSP HKQMGLVLAFLLA LSKITHASIPVPVSS PSKSPRSKGTSIKH AAMQLKTLSLATP PRETAVINL	52	3	21.5925048	0.008743169399	TLSLATPPR	954.5498034	1098.655792	-3.548999013	Yes
SB_0004	SB_0004	MKLKLTWVVKNS SWHKMDYESGFN MSEHTMAKTQTG IRYPTMLSPKPQQ LNQQNCSPEHYEP QLKTQRTWRFCFM SL	23	1	12.50724417	0.000999000999	MDYESGFNMSEH TMAKTQTGIR	2533.103525	2869.269804	7.891735204	No
SB_0005	SB_0005	MPPSSANPDEGY KVSASTHVKTGQ GVAHEVARNGLH FLPQKTTMALMK LKRRWI	111	8	31.85069597	0.001998001998	TLGQGVAHEVAR	1236.657452	1380.764859	-3.848103668	Yes
SB_0006	SB_0006	IYMEETSRNMVSV LESALGRTRV	55	4	22.97199697	0.000999000999	IYMEETSRNMVSV LESALGR	2284.119098	2572.347638	-9.47025908	No
SB_0007	SB_0007	MVGRFMGRGDK PTEPGDSWLSKME S	54	12	27.52984605	0.001998001998	MVGRFMGRGDKP TEPGDSWLSK	2450.183426	2914.478541	0.3292530069	Yes
SB_0008	SB_0008	ILLRMSLRQIKTLN WQLTAQYLQSTN KSLLPSLSTQHRHA HKERLKKVKGTRQ ILPRLFTKNITSSITS IRGTACPVTHV	43	1	23.99545008	0.000999000999	LFTKNITSSITSIR	1579.893313	1868.103082	-2.998073063	Yes

SB_0009	SB_0009	MGTCMNGSTRV QLSLTFNQWNWP AREEAGMTQQDE KTLWSFNLLMQT VPNKPTGPKLPNL H	20	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0010	SB_0010	IDPMTWPTEQVT LGMTAQSYSRVHI NNRVYDLVGS HPDGAAAIKGSFV QRLKSYVIWVQTG VIQVGFYXSNSSL YERTREMRPTSQS AFPRKWWHLNLVL YHPPKNRVC	53	1	20.4256365	0.000999000999	EMRPTSQSAFPR	1405.6772	1565.764758	6.03635534	No
SB_0011	SB_0011	IPLLNNMPMANLL LLIVPILIAMAF LTERKILGYMQLR KGPNNVGPYGLL QPFADAMKLF PLKPATSTITLY PTLALTIALLWTP LPMNPPLVNLNL GLLFILATSSLAVYS ILWSGWASNSNY ALIGALRAVAQTIS YEVTLAIILLSTLLM SGSFNLSTLITTE HLWLLPSWPLA MMWFISTLAETN RTPFDLAEGESELV SGFNIEYAAGPFAL FFMAEYTNIIIMM NTLTTTIFLGT ALSPELYTTYFVTK TLLTSLFLWIRTA YPRFRYDQLMHLL WKNFLPLTLALLM WYVSMPISSIPP QT	97	75	90.43099497	0.000999000999	KGPNNVGPYGLLQ PFADAMK	2101.102977	2533.421241	-4.745353471	No
SB_0012	SB_0012	MQNPPHSSPHSS PLPRYSYLSPLLY	8	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0013	SB_0013	IYSPMLHSAILPH HWCSPTVDYSLQ TTKTLEHYTYS AH ELES	1	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0014	SB_0014	MTLSKLNRYLNP MYLNGTCSASRST RRYFPYHRRAYHL SWSRPHNHFPYLL PSPVCPFPNTHNK TN	38	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0015	SB_0015	MVLNLRVHRLRRT NLQLLHTSPIIPRT RRPATPWWRWQSS STPDWSPHSYNN YITRRALMSPHI RLKNRCNSRSTSKP NHFHRYTTGGML RSMLWNLWSKP QFHAHRPRINSPK NLWNRARIYPMA PPLPPEPTVKLT	49	2	19.26859725	0.008849557522	NLWNRAR	928.4991084	1072.603732	-2.365401482	Yes
SB_0016	SB_0016	IKRTNTSLQWNP TKYYRMAHNNYP HTPYTIPHPTKNI KHKLPPSTLTKAH KNKKL	29	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0017	SB_0017	MNENLFASFIPTI LGLPAAVLILFPPL LIPTSKYLINNRLIT TQQWLIKLSKQ MMTMHNNTKGR WSLMLVSLIIFIAT TNLLGLLPHSFTPT TQLSMNLAMAIP LWAGTVIMGFRSKI KNALAHFLPQGT P T PLIPMLVIEITISLL IQPMALAVRLTAN ITAGHLLMHLIGS ATLAMSTINLPSTL IIFTILILLTILEI AVA LIQAYVFLLVSLYL HDNT	130	89	51.46183017	0.001020408163	LITTQQWLIK	1242.733572	1530.93992	-1.428598097	Yes
SB_0018	SB_0018	ILPSYKQSTSSLPS PFPTASTAQHFL	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0019	SB_0019	MWFDYFCMSPSI DEGLTLLV	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0020	SB_0020	ITTTDMTFQKTHN LNQHNPQPNY	9	1	21.06164271	0.00515995872	ITTTDMTFQK	1184.574696	1488.764668	6.110094777	No

SB_0021	SB_0021	IHSRTHNHLVLLR NHTYPHLGYHHP MRQPARTPERRH MLPILHPSRLPSPT HRTNLHSQHPRLT KHSTTHSHCPRTIK LLSQQLNMTSLHN SFYSKDTSLRTPLM TP	37	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0022	SB_0022	MRHNYNKLHLPT TNRPKIAHCMLFN QPHSPRSNSHSHP NPLKLHRRSHSHN RPRAYILITILPSKLK LRTHSQSHHNPLS RTSNSTPTNSFLM TSSKPR	62	11	19.17011656	0.002018163471	LHLPTTNRPK	1175.677457	1463.877578	2.758316996	Yes
SB_0023	SB_0023	IHTRKHPHVHTPIP HSPPIPQPRHHYR VFLL	42	1	14.03856455	0.002421307506	HHYRVFLL	1083.597754	1227.709476	-7.84342575	No

SB_0024	SB_0024	ILVQLQMKVMTM HTTMTTLTSLIP PILTTLVNPNKNS YPHYVKSIVASTFII SLFPTTMFMCLD QEVIIISNHWAT TQTTQLSLSFKLDY FSMMFIPVALFVT WSIMEFSLWYMN SDPNINQFFKYLLI FLITMLILVTANNL FQLFIGWEGVGI MSFLLISWWYAR ADANTAAIQAILY NRIGDIGFILALAW FILHSNSWDPQQ MALLNANPSLTP LGLLLAAAGKSAQ LGLHPWLPSAME GPTPVSALLHSST MVGAGIFLLIRFHP LAENSPLIQTLLC LGAITTLFAAVCAL TQNDIKKIVAFSTS SQLGLMMVTIGIN QPHLAFHICHTHA FFKAMLFMCSGSII HNLNNEQDIRKM GLLKTMPLTSTS LTIGSLALAGMPFL TGFYSKDHIETAN MSYTNAWALSITL IATSLTSAYSTRMI LLTLTGQPRFPPTL NINENNPTLLNPIK RLAAGSLFAGFLIT NNISPASPFQTTIP LYLKLTAVALTFLG LLTALDLNYLTNKL KMKSPLECTFYFSN MLGFYPSITHRTIP YLGLLTSQNLPLLL LDLTWLEKLLPKTI SQHQISTSITSTQ KGMIKLYFLSFFFP LILTLILLIT	113	63	76.17222727	0.000999000999	FPTLTNINENNPTL LNPIKR	2308.253868	2596.46168	-1.396327951	Yes
---------	---------	--	-----	----	-------------	----------------	--------------------------	-------------	------------	--------------	-----

SB_0025	SB_0025	IPPSNLNYNMYTN KQCSTSNYY	41	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0026	SB_0026	MIMQSPRTNRILP NQPWPLSFMNYS ASYTIKVYHNHHPI MLFHPQHQSYLH R	34	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0027	SB_0027	MTSPKIQNNNTP DHTANNQY	6	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0028	SB_0028	MHHYSRTDYNHD QWYEKPSLYFNK NTNDPNTQN	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0029	SB_0029	IHRPPHIQHLMR MKLRLTPWRLPD PPNHHRTIPSHAL LTRRLNRLFNRPH HSRRKLWLNHPLP SRQWRLNILYLP PTHRARPMLRIISL LRNLKRRHYPPAC NYSNSLHRLCPPV RPNIIIRGHSNYKL TIRHPMHWRDRPS SMNLRRLSRQSH PHTILYLSLHLALH YCSPSNTPPPILAR NGIKQPPRNHLPF R	121	2	13.82029738	0.004424778761	LRLTPWR	940.5606422	1084.660857	1.728571636	Yes
SB_0030	SB_0030	INTILTRPPRRPRQ LYPSQPLKHPSPH QARMMPFIRLHN SPIRP	30	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0031	SB_0031	ITIHPHPSNNPHPP YIQTTKHNISPTKPI TLLTPSRRPPHSNL NRRRTSKLPFYHH WTSSIRTMLHNN PNPNTNYLPNWK QNTQMGLSL	49	1	16.8933154	0.003144654088	RPPHSNLNR	1089.579146	1233.689229	-6.478610557	No
SB_0032	SB_0032	MHQSCCKPEMKT FQQQIREKVFNSTI STQS	20	2	13.83337106	0.003003003003	MHQSCCKPEMK	1217.535493	1722.853313	2.770027249	Yes

SB_0033	SB_0033	MYFVHYCQPPW MLYGMTNTWPP VVHKNPIHIKTPSP CLQASTAINPQLS HINCNSKATPHPL GYQQTYPPLTVHS T	22	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0034	SB_0034	MAHYSQIPSRPH GWPPSDRGPLTTI LREINIPHKSATLL APGP	10	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0035	SB_0035	MSQYLSLIPASSYY LSHLRSMQLQANM LTKVC	14	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0036	SB_0036	MTKNFHQTTPPP LLATALKHISAKPQ KQRTLTPA	3	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0037	SB_0037	MLLISSMQPPILP STHTPLLPYPEPT KPQRHPPQFM	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0038	SB_0038	MQASPFQWVHP LNHHDQKEQASS TQQCSSKRLA	12	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0039	SB_0039	MLTPGLVNFVPAT AVTRLTQVNRSRR KECFRSPPPQ	10	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0040	SB_0040	MEIETWRNRYST ARERWKIMTKHN MARTNPYTFM MN	33	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0041	SB_0041	MTLQGEPKLRPPK PDELPKNS	10	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0042	SB_0042	IKKAFKLNTHYLKN PKHMTTELLTPNW TNLSPYRRTNVSM SNMKTFFSA	71	8	16.0339017	0.008743169399	TNVSMSNMK	1010.452479	1330.657952	-8.639537899	No
SB_0043	SB_0043	MAPRGFSCLLLLT SEIDLVPKRRRA	4	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0044	SB_0044	IKNFGWGDLAGAE PNLRAVHAKTSPV KANYTQLIQ	20	1	23.17994279	0.000999000999	NFGWGDLAGAEPN LR	1544.737155	1688.850186	-6.473704492	No
SB_0045	SB_0045	ILESMTMGFTTS MLDQDIPMVQPL LKVRLFND	1	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0046	SB_0046	MAEPGNRMKLKT LQSEVQFLFTTYP WPTSYSYSLYPF	18	3	13.93237678	0.008869179601	MAEPGNR	773.3490056	917.4533302	-2.447302268	Yes
SB_0047	SB_0047	MSPYPLQSPAFL KPKKYVW	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0048	SB_0048	ISRTMRIPIPENP KFSVPPITPHPKVR SAK	10	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0049	SB_0049	IFYLSRPRNKHASF YSSSNQKNKPSFH RSCHQVFPHASN RIHNPSNSYPLQQ YTLRTMNHNNQYY QSMLIINNHSYS NKTRNSPLSLLSPR GYPRHPSDIRPASS HMTKTSPHLNHN PNLSLTKRKPSPHS LNLIHHSRQLRWI KPNPATQNLMLL NYPHRMNNSST VQP	122	4	12.05265959	0.007658643326	IFYLSRPR	1050.59742	1338.79671	3.632896529	Yes
SB_0050	SB_0050	IPTTQLKLQHDP TTSHLKQANMTN TLNSIHPPLPRRPA PANRLFAQMGHY RRIHKKQ	41	1	20.32219288	0.000999000999	QANMTNTLNSIHP PLPR	1902.973367	2047.067215	4.036097747	Yes
SB_0051	SB_0051	MTVWYKTHPIP PHTHRPHYATPTY LPFYTNLMMEI	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0052	SB_0052	ISVTAKDCKTPLCI NWTQISHFN	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0053	SB_0053	MKITSELVKRGLTP VFRFTVQCFTQPF YLTPDVRRLTIL YKQQRHWNTMPI IRRMWSWSPRHSSK PPYSSRAGPARQP SR	31	2	17.19087616	0.001001001001	HWNTMPIIRR	1322.702959	1466.810724	-3.863706822	Yes

SB_0054	SB_0054	ICNNLLHSNTHHN RRLWQLTSSPNN RCPRYGVSPHKQ HKLLTSLSPTPA RICYSGGRSRNRL NSLPSLSRELLPPW SLRRPNHLLTPSR CLLYLRGHQFHNN NYQYKTPCHNPM PNAPLRIRPNHSS PTSPISPSPCWH HYTTNRPQPQHH LLRPRRRRPHSM PTPILIFRSPWSLYS YPTRLRNNLPYCN LLLKKRTIWMHR YGLSYDINWLPRV YRVSTPYIYSRNR RHTSMFHRLRYHN HRYPHRRQSI	156	8	27.54786054	0.001154734411	ELLPPWSLR	1109.623299	1253.713095	9.8081714	No
SB_0055	SB_0055	MICCSALSPRIHLS FHRRWPDWHCIS KLITRHRTRHVLR CSPLPCPINRSCIC HRRRLHSLISILRL HPRPNLRQNPFH YHIHRRKSNFLPTT LSRPIRNAPTLGL PRCMHHMKHPHII CRLIHFSNSSNINN FHDLSLRFEAKSP NSRRTLHKPGVT MWWMPPTLPHIRR TRMHKI	99	8	26.39510071	0.000999000999	TTRHVLRCSPLPLC PINR	2075.124397	2333.278658	-3.949405601	Yes

SB_0056	SB_0056	MAHAAQVGLQD ATSPIMEELITFHD HALMIIFLICFLVLY ALFLTLTTKLTNTN ISDAQEMETVWTI LPAILLVIALPSLRI LYMTDEVNDPSLT IKSIGHQWYWTYE YTDYGGIFNSYM LPPFLFEPGDLRLL DVDNRVVLPIEAPI RMMITSQDVLHS WAVPTLGLKTDI PGRNQTTFTATR PGVYYGQCSEICG ANHSFMPIVLELIP LKIFEMGPVFTL	779	656	76.27685888	0.000999000999	MIMITSQDVLHSW AVPTLGLK	2226.154024	2514.353794	1.755039781	Yes
SB_0057	SB_0057	MPQLNTTVWPT MITPMLLTLFLITQ LKMLNTNYHLPPS PKPMKMKNYNKP WEPKWTKICSLHS LPPQS	328	277	64.10598268	0.000999000999	MLNTNYHLPPSPK PMK	1866.948398	2299.260345	-2.484358335	Yes
SB_0058	SB_0058	MTPNRGPLSPPN DLRPSHVISLPLHN APHTRPTNQHTN HMPMMARCNTN KHMPRPPTTTC KRPSMRDNPIYYL RSFFLRRIFLSLLPL QPSPPPIRRALAP NRHHPAKSPRSPT PKHIRITRIRSINHL SSP	35	2	24.59374409	0.000999000999	RPSMRDNPIYYLR	1679.856552	1968.066962	-3.172199407	Yes
SB_0059	SB_0059	IFCSHRLPRTSRHY WLNFPHYLLHPPT NISLYIQTSLWLR RRLMLAFRCGLT ISVCLHLLMRVLLF	10	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0060	SB_0060	ILTTTTQRLHRKIH PLRVRLRPYIPRPR PFLHKILLSYLLII WSRNCPPFTPTM SPTNN	18	2	21.63645375	0.001005025126	ILTTTTQRLHR	1338.773144	1482.86694	5.600029957	No
SB_0061	SB_0061	INHHPSPKSLWV TKRIRLNRIGM	7	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0062	SB_0062	IYHLTSRNTSM SLT PHILPTMPRRNNT IAVHYSYSHNPQH PLPLSQYAYCHTS LCRLRSSGGPSPTS LNLQHMWPRLRT	64	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0063	SB_0063	MLKLIVPTIMLLPL TWLSKHKMIWIN TTTHSLIISIIPLFF NQINNNLFCSPST FSSDPLTTPLLMLT TWLLPLTIMASQR HLSSEPLSRKKLYL SMLISLQISLIMTF TATELIMFYIFFETT LIPTLAIITRWGNQ PERLNAGTYFLFYT LVGSLPLLIALIYTH NTLGS LNILLTLTA QELSN SWANNLM WLAYTM AFMVK MPLYGLHLWLPK AHVEAPIAGSMVL AAVLLKLG GYG M MRLTLILNPLTKH MAYPFLVLSLWG MIIMTSSICLRQTD LKSLIAYSSISHMA LVVTAIIQTPWSF TGAVILMIAHGLT SSLLFCLANSNYER THSRIMILSQGLQ TLLPLMAFWWLL ASLANLALPPTINL LGELSVLVTTFSW SNITLLL TGLNMLV TALYSLYMFTTTQ WGS LTHHINNMK PSFTRENTLMFM HLSPIILLSLNPDIIIT GFSSCKYSLTKTSD CESDNRGLRPLIYR ESSQELLTHAPMS NNMAFSTFKG	87	8	84.24539063	0.000999000999	AHVEAPIAGSMVL AAVLLK	1889.080786	2177.288583	-1.661575297	Yes
SB_0064	SB_0064	MPTQQPFKQSYT TVSAMSVS SSP	8	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0065	SB_0065	ITMYTPTNNVQPV TTTNQRP	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0066	SB_0066	IIQLPTLLKFTTTTT PSYSFTHSTNPTSI ANPTKTLTKTSTP DPHASGYSSMAIA VVYPKTTIIPPK	16	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0067	SB_0067	MGEGLEENPTNPI TKPTLNRNKAYIIL ARTTTTTNDMKN HRCISTTRTPMTP MRKTNPLMKLIN HSFIDLPTPSNISA WWNFGSLLGACL ILQITGLFLAMHY SPDASTAFSSIAHI TRDVNYGWIIRYL HANGASMFFICLF LHIGRGLYGSFLY SETWNIGIILLAT MATAFMGYVLP WGQMSFWGATV ITNLLSAIPYIGTDL VQWIWGGYSVDS PTLTRFFTFHFILPF IIAALATLHLLFLHE TGSNNPLGITSHS DKITFHPYYTIKDA LGLLLFLLSLMTLTL FSPDLLGDPDNYT LANPLNTPPHIKPE WYFLFAYTILRSVP NKLGGVLALLLSILI LAMIPILHMSKQQ SMMFRPLSQSLY WLLAADLLILTWI GGQPVSYPFTIIG QVASVLYFTTILIL MPTISLIENKMLK WACPCSMN	63	7	21.38188719	0.002212389381	KTNPLMK	830.4683856	1262.772811	1.41407854	Yes
SB_0068	SB_0068	ILCSFMGKQIWVP PKYWLTQQPLCI SYITASHHEYCTVP	13	4	16.55336317	0.001092896175	ILCSFMGK	897.4452082	1242.673333	-2.022001366	Yes
SB_0069	SB_0069	MKTQSTSKPPPH AYKQVQQSTLNY HTSTATPKPPLTH	5	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0070	SB_0070	ITVKSLVPMDDP PQMGVPPSSV KSMSRTRVLLSLR AHNTWG	3	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0071	SB_0071	INHSRELSMHLVF SSGGYARDSIARR WSRSTLCRSICLW FLPHPIIYRTYVQY YRRTYLLKCVN	60	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0072	SB_0072	MLVGHNNNNW MSAQLSTQTS	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0073	SB_0073	ISTKPPLPRFWPQ HLNTSLPNPKNE P	4	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0074	SB_0074	ISNFIFWRYALLTV TPQLTHYFPLPLPY Y	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0075	SB_0075	MEAGVKSVDHP LPNKAETHLSCKKL QLTQNRLRKWL	5	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0076	SB_0076	MNPDQPHHLLLS LYTAIFSKPWWRL QSKRKYPRKDVR RCSPWGGKKWAT FSTPENYDSPYET	18	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0077	SB_0077	ICPQNPLNPLVNL TVSPKRNSLDTR KKPCRESKKFNTH SRPKSSHQLRKRSS STPTT	17	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0078	SB_0078	INSPMSTINQQVII TLTVNPTQACS	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0079	SB_0079	MITCSLNRDLYEW LHEGSAVSFY	10	1	12.12041551	0.009836065574	MITCSLNR	936.4520848	1153.56304	6.508987843	No
SB_0080	SB_0080	ISVGATSEQNPTSE QYMLRLHQSKRTT MLNWSNNLTNG TSYPRDNSAILF	47	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0081	SB_0081	IKVLRDLSSDRSNP GRFLSXFKFLPVRK DKRNKAYFTKRLP P	33	1	22.79775127	0.003058103976	RNKAYFTKR	#N/A	1614.978081	-6.018171773	No
SB_0082	SB_0082	MMSSQLSIMPTP TQEQGLLRWQSP VIA	19	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0083	SB_0083	IPNAYRTKNSRLYT TTQRPQRCRPLRA TTTLRWRHKTLHQ RAPKTRHIYHHPL HHRPDLSSHHRSS TMNPPPHQTQPPG QPQPRPIYSSHL	34	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0084	SB_0084	ITPAIMTLGHNMI YLHTRSDDQNPLR PCRRGVRTSLRLQ HRMRRRRLRPILH SRMHKHYYNKHP HHYNLPRNNMW RTLPLWTLHNMFC HQDPTSNLPVLM NSNSMPPILRPT HTPPMKKLPTTHP SITYMMCLHthy NLQHSPSNLRNM SDKRVTLME	101	4	24.34214933	0.000999000999	NMSDKRVTLME	1322.632225	1626.841966	-6.55855142	No
SB_0085	SB_0085	INPLAQPIYSTIFA GTLITALSSHWFFT WVGLEMNMLAFI PVLTKKMNRSTE AAIKYFLTQATAS MILLMAILFNNML SGQWTMTNTTN QYSSLMIMMAM AMKLGMAPFHF WVPEVTQGTPLTS GLLLTWQKLAPIS IMYQISPSLNVLL LTLNILSILMAGSW GGLNQTLRQKILA YSSITHMGWMM AVLPYNPNMTILN LTIYIILTTAFLLN LNSSTTTLLSRTW NKLTWLTPLIPSTL LSLGGPLPTGFLP KWAIEEFTKNNSL IIPATITLLNLY FYLRLIYSTSITLLP MSNNVMMKWQF EHTKPTPFLPTLIAL TTTTLPIPFMLMI L	68	17	30.98510606	0.001048218029	WAIIEEFTK	1135.591332	1423.796681	-0.8391981488	Yes

SB_0086	SB_0086	MFADRWLFSTNH KDIGTLYLLFGAW AGVLGTALLLIRA ELGQPGNLLGND HIYNVIVTAHAFV MIFFMVMPIMIG GFGNWLVPLMIG APDMAFPRMNN MSFWLLPPSLLLL ASAMVEAGAGTG WTVYPPLAGNYS HPGASVDLTIFSLH LAGVSSILGAINFIT TIINMKPPAMTQY QTPLFVWSVLITA VLLLLSLPVAAGIT MLLTDRLNNTTFF DPAGGGDPILYQH LFWFFGHPEVYILI LPGFGMISHIVTYY SGKKEPFGYMGM VWAMMSIGFLGF IVWAHHMFTVG MDVDTRAYFTSA TMIIAIPGKVFVS WLATLHGSNMK WSAAVLWALGFIF LFTVGGLTGIVLAN SSLDIVLHDTYYVV AHFHVLSMGAV FAIMGGFIHWFPL FSGYTLQTYAKIH FTIMFIGVNLTFPP QHFLGLSGMPRR YSDYDAYTTWNI LSSVGSFISLTAVM LMIFMIWEAFASK RKVLMVEEPSMN LEWLYGPPPYHT FEPPVYMKSRQKR KESNPPKLVSSQP HGLHDFFKVLEK PFHNFVKVKL	151	91	90.04728257	0.000999000999	VLMVEEPSMNLE WLYGCPPPYHTFE EPVYMK	3727.711022	4088.940668	-2.215073813	Yes
---------	---------	--	-----	----	-------------	----------------	---	-------------	-------------	--------------	-----

SB_0087	SB_0087	IMTNPENQNERK SVRFIHCPHNPRP TRRSTDHSISPSID PHLQMSHQPTN HHPTMTNQTNLK TNDNHTQH	31	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0088	SB_0088	IYTNHPTIYKPSHG HPLMSGHSDYRLS L	3	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0089	SB_0089	MMTHQSHAYHM VKPSPWPLTGALS ALLMTSGLAMWF HFHSMTLMLGLL TNTLTMYQWWR DVTRESTYQGHHT PPVQKGLRYGMIL FITSEVFFFAGFFW AFYHSSLAPTQPL GGHWPPTGITPL NPLEVPLLNTSVLL ASGVSITWAHSL MENNRNQMIQA LLITILLGLYFTLLQ ASEYFESPFTISDGI YGSTFFVATGFHG LHVIIIGSTFLTICFIR QLMFHFTSKHHF GFEEAAWYWHF VDVVWFLYVSIY WWGSYSFSMNST VNFQLTSFDNIQK RVMNFALILMINT LLALLMIITFWLP QLNGYMEKSTPYE CGFDPMSPARVP FSMKFFLVAITFLL FDLEIALLLPWA LQTTNPLMVMS SLLLLIALSLAYE WLQKGLDWTELV YSLNKTNDFDSL YDNHIYQMPLIYM NIMLAFTISLLGML VYRSHLMSSLLCLE GMMLSLFIMATL MTLNTHSLLANIV PIAMLVFAACEAA VGLALLVSISNTYG LDYVHNLNLLQC	111	69	55.11127489	0.000999000999	ESTYQGHHTPPVQ K	1607.769182	1895.97657	-1.701906999	Yes
SB_0090	SB_0090	MSSSKPHLSPPWL SSPDEATSQNAW TQAHTSYSTP	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0091	SB_0091	IVNLTEAYDPLFT EKAHKNC	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0092	SB_0092	MPPCLTTWLSQLL KDNSYPLVLGPKN FGATPNKSNHHA HYYNHNPDPFNP SPHPYHPR	28	4	20.47544064	0.001998001998	DNSYPLVLGPK	1201.634257	1489.846087	-5.150590851	No
SB_0093	SB_0093	MPPLCKIHCRHLY YQSLPHNNIHVPR PRSYLELTLSHNP NNPALPKLQTRLL LHNIHPCSIVRYM VHHRILTVMYKLR PKH	33	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0094	SB_0094	IPTVHRLRGRNRYI LLAHQLMMRPSR CQHSSSSNPMQ PYRRYRFHPRLSM IYPTLQLMRPTTN SPSKR	40	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0095	SB_0095	IRSPPLTPLSHRRP HPSLSPTPLKHYSC SRNLLTHPLPPSR K	14	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0096	SB_0096	MLRRYHHSVRSSL RPYTKWHQKNRS LLHFKSTRTHNSY NRHQPTTSPSIPAH LYPRLQSHTIYVL RVHHPQP	14	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0097	SB_0097	ISRNTFPHRFLQR PHHRNRKHIMHK RLSPIYSHRYLPD KRL	24	1	17.01163818	0.004842615012	NRKHIMHK	1062.586873	1494.885741	4.915416916	No
SB_0098	SB_0098	MTYSPEQSQLQY MHQQTMFNQ	4	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0099	SB_0099	MMWKTIVVFQL QEQHWPQYAKLT P	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0100	SB_0100	IMAESSATFTPMA PQYLSASSYTSGE AYITDHFSTQKPET SALSSCLQL	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0101	SB_0101	IPMKSPSTLTQSK TPSAYFSSFSP	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0102	SB_0102	IDSPINRNVFRTL LPATMNIVRYHKY LTTCS	11	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0103	SB_0103	MLTSKYSNQPSTI THQLQLQSHPSPT RMPTNLPTLNST	125	3	11.44115982	0.002997002997	MPTNLPTLNST	1187.585595	1347.676777	4.321420165	Yes
SB_0104	SB_0104	MKPFTVHSTLQSN PFSSPWMTPLR	3	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0105	SB_0105	MYYVLLRVGRFVG ILVGEGWLWSCS WCVMEVWLLYL LVSMGRGFWCGL GFYVLQVVKYLWY RTMFMVAGSNVR NT	151	5	21.9103728	0.000999000999	TMFMVAGSNVR	1211.57907	1355.671422	7.18439158	No
SB_0106	SB_0106	MGESMLGWYPN LLPHERTENS LN	4	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0107	SB_0107	MAYEGCCYSCKQ EDNADVSGFWVE KWSVM	5	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0108	SB_0108	IGVKVADDSAMIY VSSDVGDDWVKG GWGVVWVHVG	3	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0109	SB_0109	MVFHIIIGRCSPC ENNDVCFVSVCEG FSNGVCGVFF	2	1	13.14643154	0.006053268765	MVFHIIIGR	971.5374648	1131.640437	-5.273572206	No
SB_0110	SB_0110	MYYSDGYWGVVS WGMGVRGWGL GECFSGVSDGGRI GAVGERVWWGG GCGKL	54	4	28.02515	0.000999000999	IGAVGERVWWGG GCGK	1630.803791	1976.033302	-1.964915008	Yes
SB_0111	SB_0111	ISSYWLNIWCWC MYCNWDCSGE	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0112	SB_0112	MGLRRLCMMCLR FRWCGLWSRNL WGKVFLLMLGCQ WWGRLKWEVWF WVLLFFEYLVHC	23	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0113	SB_0113	MVWLWRRRGYR CAGMLGVVWGC RL	28	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0114	SB_0114	MVLEFGLVGYFLL GGGSGWVRRFLL QL	8	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0115	SB_0115	IGLICLLLGGGLV VGWGLD	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0116	SB_0116	MLRRGWNRYRRY GCMGLEWLLCW HLLGRIINWWAR RM	32	1	19.24635622	0.008474576271	IINWWAR	957.5184452	1101.610498	9.109554067	No
SB_0117	SB_0117	MLQGWMWLRSS LVWSLGRAGLFLG WLSVSSR	21	1	12.74455536	0.003996003996	AGLFLGLWLSVSSR	1391.756098	1535.870039	-7.710232259	No
SB_0118	SB_0118	IETSRARPTAASQ AAKTSMAMGTM LAKREWVLRVMR VAMMNSDSIIPSR HSREDMRCERYTS IPREMVNASMM FM	146	2	20.03051746	0.000999000999	VAMMNSDSIIPSR HSR	1799.877029	1959.982881	-4.506118554	No
SB_0119	SB_0119	MMINKRDDMTIS GRLVVCRAHGRG KRRAISRSNNKKV MATKKNFMEKGT RAGDMGSKPHS	64	4	19.56720182	0.002424242424	DDMTISGR	893.3912624	1053.484827	3.257502508	Yes
SB_0120	SB_0120	MIISSKARRVLIKI KAKFITLFWMLSK LVNWKLTVLFMLK E	24	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0121	SB_0121	METYRNSQTTSTK CQYQAAASKPKW CLDVKWNISWRM KQMVVRKVEPMM TWSPWKPVATKN VEP	24	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0122	SB_0122	MPSEMVKGDSKY SEACRRVK	32	1	13.0533318	0.006651884701	YSEACRR	883.3970162	1084.531222	-9.830153414	No
SB_0123	SB_0123	IVMSSAWIIFRL FSIRLWVAQVIDT PDASNTDVFRSGT SRGFSGVMPVGG QCPPNWGVGARL EW	10	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0124	SB_0124	MNRIIPYRRPFWT GGVWWPWVYVLS RVTSRHHWYMVS VLVSRPSMR SVM EWKWNHMARPE VIRRAERAPVRGH GLGFTMW	113	4	19.15712093	0.000999000999	HHWYMVSVLVS PSMR	1983.992327	2128.084213	4.805330945	No

SB_0125	SB_0125	IMCCRAGRGLLEV WKRRLGLRRQRFL G	15	1	25.62950155	0.000999000999	IMCCRAGRGLLEV WK	1733.889112	2152.138023	-3.205659403	Yes
SB_0126	SB_0126	IVGWVWLVCWW DIWRWGSMEGE MEWSVLRVGLG LWGQWMKRTDF RSFWFSGFVMIFY FYGLWWGR	33	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0127	SB_0127	MEGSSTIRTFRFEA KASQIMKIINITAV REMNEPTDDRMF HVVYASG	89	4	17.13238898	0.009523809524	IINITAVR	898.5599734	1042.671773	-9.311726812	No
SB_0128	SB_0128	IPDRPRKCCGKKV RFTPMNMMVKW ILA	8	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0129	SB_0129	MSSDEFANTMPV RPPTVKRKMNPR AQSTAADHFMLL PWSVASQLNTLTP VGMAMIMVAEV KYARVSTSIPTVN MWCAHTMNPVK PIDIMAQTMPPY PNGSFFPE	14	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0130	SB_0130	MGSPPPAGSKKV VLRLRSVSSMVM PAARTGRDRRSRT AVIRTDQTKRGV WYVVMAGGFM LMIVVMKLMAPK MEETPARCKEKM VRSTEAPGWE	401	9	31.99770905	0.000999000999	TDQTKRGVWYVW MAGGFMLMIVV MK	2946.478001	3410.753217	6.120421135	No
SB_0131	SB_0131	MADASRSRREGG KSQKLMFMGRN AMSGAPIIRGTSQ LPKPPIMMGITMK KIITNAWAVTML	53	3	11.91927806	0.009900990099	MADASRSR	892.418479	1036.514518	5.830421843	No
SB_0132	SB_0132	MWSLPRRLPGWP SSARMRRLRAVPR TPAHAPNNRYSVP MSLWFVENSQRS ANISGGEVKWLSE ALDCKSKDRG	107	5	24.68961083	0.002040816327	WLSEALDCK	1063.500807	1408.724693	1.226049523	Yes

SB_0133	SB_0133	IELQIRRSFKPAG ASPAFFPGGGRSR LKPVD	4	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0134	SB_0134	IKVADLRSDAEW GFAVLSCYRN	9	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0135	SB_0135	IISMKGEMGRSSV VRAMSVGRNGV GFVCSNCHFIFLL DMGSSVIEVE	38	1	10.14925663	0.007063572149	IISMKGEMGR	1120.573256	1408.789118	-8.310300416	No
SB_0136	SB_0136	MVAMMVGMMR LLFFVNSSMMAH LGKKPVSGGRPPR ERRVDGIGVSHV SLFQVRDSSRVVV LEFKLSSRNAVVV RMM	90	2	22.19791862	0.002688172043	MVAMMVGMMR	1155.509478	1347.595171	0.8468832668	Yes
SB_0137	SB_0137	MVKLRMVMGLGY GRTAIHPMWVIE EYAKILRSWVWFN PPQLPAMMDKIE RVRRLTFSEGEI WYMIEMGASFCH VRRSRPDVRGVP WVTSGTQKWKG AIPSFIAAIMIIN DEYWLVLVLMVH CPESMLLKRMAIR RIMDAVACVRKYL MAASVERGFIFLV RTGMKASMFISRP TQVKNQCELSAV MSVPAKMVE	231	23	24.95221347	0.000999000999	SWVWFNPPQLPA MMDK	1945.921851	2250.102287	8.288127825	No
SB_0138	SB_0138	MGCDRWHGFEW ILRDGFDSHSPRN KGV	20	5	18.53962327	0.003144654088	DGFDSHSPR	1016.431152	1160.534518	-1.106494702	Yes
SB_0139	SB_0139	IISIKVTLSDMFL RFEGECWRL	13	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0140	SB_0140	MLGWVVGSGFFM GGVWVGRSGIGG MLFEFMRTGRLE VGSWWQNMLCR VQGRVRHMLFLG RL	23	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0141	SB_0141	IRLWRMGRRLR RIRCWSLRLVRTPL RQGRRGFGWSLL VWR	34	5	24.18824918	0.008869179601	TPLRQGR	826.4773112	970.5764446	3.041597953	Yes
SB_0142	SB_0142	MLMVEWWLWGW LHMRLFGLLAVR RSGRSLMLTLIR GLSKRLG	10	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0143	SB_0143	MGGLGWGWPG GWVWGGGFMV EERWWELRSGR WCRGWW	12	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0144	SB_0144	MWRVLGALWW RVLWRQRRVVVA RRGLQRWGLCVV VYSLEFFVR	24	1	13.52718761	0.003816793893	QRRVVVAR	982.6148008	1126.70881	7.174189307	No
SB_0145	SB_0145	MPLRLEWVQWG VGGWPVVCC	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0146	SB_0146	ITGLCHLNKPCSW VGVMMLSWDD IYGGRRFVK	3	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0147	SB_0147	MVVRFDWWSLS MYCSEVGCSEVA PTEIFNAGLVV	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0148	SB_0148	MGSSRLAVLCPPL HGQVNFTG	1	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0149	SB_0149	IHYAEGMGVSPCY IMLGYNFSSPCG TMSIAPGFNFYRL YFIWVNLAKVV W	6	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0150	SB_0150	MAVYRLSKRWW GWSGFIDYRTGSS RGMWSTARSFEF	30	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0151	SB_0151	MERLGNLFFVYG VMWARLNIFSVLL WGGKHLKLVGVS LGFGWFGVWG	24	1	11.74932761	0.000999000999	MERLGNLFFVYGV MWAR	2038.028042	2198.13785	-5.814048991	No
SB_0152	SB_0152	IFCYDVCVESGCA DIQLLLCPTSIN	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0153	SB_0153	MMGWGRNQRQ MLRHRVLRQLA MLSRAYPDENTK CMESSREWLMG W	61	4	16.473808	0.004197271773	MMGWGRNQR	1134.517476	1310.619018	-7.343252657	No

SB_0154	SB_0154	MSDTVHFSYPQVL WARSEESSTLVRD IDFTEDGGQGTP WGGSSMGTRRDL TVMCYVR	75	3	25.31221444	0.001998001998	DIDFTEDGGQGTP WGGSSMGTR	2383.038604	2527.126774	5.516739491	No
SB_0155	SB_0155	MYYRWSSYIGTV QYSWWLAVMYE MHSGCWVVSQY LGGTQICFPMKEQ RMV	1	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0156	SB_0156	IRILALGANGGVK DFFSDLSEKGFHL RFTRLVY	22	1	24.23521817	0.001020408163	DFFSDLSEK	1199.570992	1487.783158	-5.386567532	No
SB_0157	SB_0157	MLQGQAHSILFSI REMGIRIRIVVKY STDATCPMMVKG	99	1	16.12499284	0.003631961259	EMVGIRIR	972.5538426	1132.64155	8.208387675	No
SB_0158	SB_0158	IMLCCLDMWRM GIIARMRMDSNR ARTPPSLGTD V	81	12	20.26730827	0.001048218029	MRMDSNRAR	1135.533854	1295.622706	6.291403817	No
SB_0159	SB_0159	ILSEWEVIPRGLFD PVSCNRRWSVA RAAMMKGKMK WKVKNRVRVGLS TE	33	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0160	SB_0160	IHWTRSVPMYGM ADSKFVITVAPQN DIWPHGRT	35	10	18.13464371	0.002997002997	SVPMYGMADSK	1184.520555	1488.711045	5.760069176	No
SB_0161	SB_0161	MKAVAMVASRR MMPMFQVSE	21	2	20.80739396	0.008083140878	AVAMVASRR	959.5334428	1103.624632	9.876398944	No

SB_0162	SB_0162	MARNSPVVIWRI RQAPRSEPKFHHA EMLDGVGRSMNE WLINFIRGLVLRIG VIGLVVEMQRW FFMSLVVVVRAR MMMYALFLLSVG LVMGFVGFSSKPS PIYGGVLIVSGVV GCVIILNFGGGYM GLMVFLIYLGGM MVFVGYTTAMAI EEYPEAWGSGVE VLVSVLVGLAMEV GLVLWVKEYDGV VVVVNFNSVGSW MIYEGEGSLIRE DPIGAGALYDYGR WLVVVTGWTLFV GVYIVIEIARGNRL CD	88	2	17.30034468	0.000999000999	FHHAEMLDGVGR	1367.640422	1511.729762	8.436715496	No
SB_0163	SB_0163	MGDCAVCDARVE SEYVGEMKCA	15	1	16.17840283	0.005076142132	VESEYVGEMK	1169.527414	1617.826131	1.490421419	Yes
SB_0164	SB_0164	MFEELINVVVVV YMSQWEFYDGPC NEQCQRDEYYGEV V	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0165	SB_0165	MMGVVWFFVRV NEGGKGGN	50	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0166	SB_0166	MDSCYPLKVEKA MLLDMGAWVSS SCELSR	11	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0167	SB_0167	MRGRKPLLSDSQS DVLVKLYLQEENP VMMSGLRDRRR MGDRCMNMRVF SRVNEGFMLLM WWWSEPHCVVV NM	45	2	14.61955031	0.002997002997	MGDRCMNMRVF SR	1601.704692	1802.814079	7.862381391	No
SB_0168	SB_0168	MLSPVSRVMFD QENVVTSTESSPS RLMVGKARLAR LARSHQKAISGSR VWSPWERIMMR LWVRS	63	2	19.0610428	0.008484848485	IMMRLWVR	1103.60958	1263.71823	-9.212354895	No

SB_0169	SB_0169	IMRMTAPVKLQG VWMRMAVTTRA MWLIEEYAMSDF RSVCRRQMELV MIMPHRDSTRKG	408	14	24.68111573	0.000999000999	RQMELVMIMPHR	1539.78359	1683.892152	-3.836329152	Yes
SB_0170	SB_0170	IKLLAQEFSSWA VRVSSRMFSEPRV LWV	18	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0171	SB_0171	MISSVAVNVMIKE ICREISMER	12	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0172	SB_0172	MIVRGRSQVVSIR RGVVVRGSEEKVGE QLNRLLLIWLKNS RGMMLMIRLWV VVLIQIMCFLESHV SGSNMIVGTISFSI GVGLGYVRS LGH MCWRLRLVGLGP PLLRRLRQLVWQ	155	1	12.06194623	0.009433962264	GRSQVVSIR	1000.577749	1144.676438	2.971509154	Yes
SB_0173	SB_0173	MVLFLLGMVGR MWGVSDMLVFL EVRW	12	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0174	SB_0174	MSRNHSFCLNYM PIRFSLILFVTHRP DLGLGWWMRGM MT	22	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0175	SB_0175	MVKPHLQNASIRR RLRSQSDVWM	6	1	14.2310863	0.000999000999	MVKPHLQNASIRR R	1704.968163	2009.156981	5.11088456	No
SB_0176	SB_0176	MLSRRCRRKWW RETRSTLRLVGG	1	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0177	SB_0177	MLLMRVMRMCL GVGLLDLAGWC LLGASALLIGG	9	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0178	SB_0178	MGLSRIEGLFGQV VCGGLGMCFVL HRAIIGMWLVCW LVGLVWALWGS SEITWLGRRLGG LRGPLLGVMGWV LLYDRHVIGGSLCV VVQVEAY	48	10	26.94818027	0.008474576271	RSLGGLR	757.4558484	901.5583248	-0.4346533091	Yes
SB_0179	SB_0179	IFSWVMRNSVRS MGVIMVGHVVVF SWGISL	11	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0180	SB_0180	ISKIFRGINSRTMG MKLWFAPQISEH WP	51	3	22.00463893	0.003636363636	IFRGINSR	961.5457216	1105.650919	-2.811209079	Yes
SB_0181	SB_0181	IASVFKPNVGTAH ECKTSCDVIIMRM GASIGSTTRLSTSR SRRSPGSRNNGGS M	26	1	17.90667062	0.001002004008	RSPGSRNNGGSM	1218.552341	1362.66105	-4.858555349	No
SB_0182	SB_0182	MRRDGRAMRTR MMAGRMVQTVS ISWASEMLVLVSF VVSVRKRAYRTRK QMRKMIMRAWS WKVMSSMMGE VASCRIPTCAACAI KMYRI	73	7	17.2891006	0.008849557522	TRMMAGR	821.3999924	981.4877215	9.442820089	No
SB_0183	SB_0183	MVFLMPFWKSH GGHGVGLKPALG GSIPSFV	4	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0184	SB_0184	ILCMRVLRMCGR VGGIHMVTPGLW RVLLLLGLFASKRR LLKSWKLLMLLLE KWMSLQMMGCF MWCMHRGSPSN VGAFRMGRESVV GRKLDLRRWMW	55	1	16.68926836	0.000999000999	ESVVGKLDLR	1270.735698	1558.932275	4.864339307	No
SB_0185	SB_0185	MGEISEWSLLWW QMQLLLMGHSGS GLQRSTCRVVRCL VMSLLMQCQSGH LRWKERWILGLRA LQQIISYCFRGVW RVS	42	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0186	SB_0186	MVLFRRSSKLQYG RLFRSLVG	13	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0187	SB_0187	MGVGM EWGLLL RRGRRRCWGC GLLV	20	1	20.65512067	0.001	MGVGM EWGLLLR	1360.699513	1520.792566	2.603800845	Yes
SB_0188	SB_0188	MGEVGLLWLGRI RRRGAFGIGLWQ GVLYW	5	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0189	SB_0189	MRAGVGEREVRV RSLCCLCGETPYR GHRLLGELVSCQS LRLWWVLLWRRL LQMHGLWR	43	2	25.21398491	0.006564551422	AGVGEREVR	971.5148174	1115.609647	6.50557754	No
SB_0190	SB_0190	IFMLNCKFEEAAS NLPGLPPFFPAA GEVDWSQLRIVLS C	2	1	12.7887681	0.006651884701	IFMLNCK	867.434644	1212.650713	7.869148105	No
SB_0191	SB_0191	ICVQLMQSGVLQS LAVTEIKYCNLLRA LKALGLYLT	8	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0192	SB_0192	IWAKSRLAGAGLL GRGGWMELRVLV MLACFRCEMVVG SWCWLSWVVG MR	24	3	30.49572606	0.001004016064	SRLAGAGLLGR	1069.635595	1213.734458	2.661272284	Yes
SB_0193	SB_0193	MLRFCVAGFGLIH LNCLLWWMRLRE WGEGRLVRERF GMWLRWGLVFN MWEEAGRMSEG CLG	65	1	19.80232168	0.000999000999	LVRERFGMWLR	1461.802668	1605.90347	0.8099306964	Yes
SB_0194	SB_0194	MRLLAWGNTW WQLLWNEGLFF WLELE	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0195	SB_0195	MGPMAYLADTL GWGVMGGTENF GFSGMGSILMVLE MRGFKLLFTLSK	5	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0196	SB_0196	MWDCLGYCSQC ADQGVVWVWCS PWSEDWVNG	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0197	SB_0197	MMSFTGEGALW SRPYFSCPFVQGGI WX	1	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0198	SB_0198	MSWSNIEVVNPIV DMDSRMGLRCYP	15	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0199	SB_0199	ISLVKSKRQLNPRG AIHTGPYLRNKWL CYLCTVRVPRPLN MCHWAGGASNT GDARGDVFGKQA G	28	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0200	SB_0200	MLCLVMIFHLSLA VLYLLRQVSISIAYT LFG	1	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0201	SB_0201	MLEEGDGRVCRA SGPCKTKHSTLSLL LNPPSTLKFHKG RSFLG	28	3	12.21386031	0.004243281471	MLEEGDGR	905.3912624	1065.481706	6.149467885	No
SB_0202	SB_0202	MGWARGGEVDR GLSITEQAPLEGYE APPGPLSFKLWLV VFWRAVLLI	13	1	17.17685495	0.000999000999	MGWARGGEVDR	1232.572012	1376.671919	1.585656856	Yes
SB_0203	SB_0203	MNCGGCLWGLV GSGYGVSSGVCVL GRMGGGCIDEISS MGVGGENNVLV GGWLLKVHTAKR	7	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0204	SB_0204	MCLSAVARSGGG GVVWVFFVMMS VWKVAVQTFNCY YYVLQALIN	18	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0205	SB_0205	MGCSGSSVSQCY RVHTPQTKMPNA WRAPVSG	21	6	11.77834789	0.001998001998	MGCSGSSVSQCYR	1363.531865	1621.660993	9.797445322	No
SB_0206	SB_0206	MSYLRGTCGLFRL YDPEVGTRCRMQ FTLATPKCYGPGA RRVALLCGMLISR RMVVKGPLSEGG HPWGREGIWL	119	6	33.32383618	0.000999000999	VALLCGMLISR	1174.656587	1391.768589	4.651040481	No
SB_0207	SB_0207	MCDSWGLIAVLA CKHGEGLMWIG FLCTGGQVFMV PYNHGGWQ	13	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0208	SB_0208	MAVVDGWWNT WVVPKASPWKN RE	11	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0209	SB_0209	MVELKTFSLICPW KKVFISGLQDWCI SLYYKDRPIWVFC FQLGR	30	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0210	SB_0210	IGLVGEMLCFVV WMYGGWGLLLG WGWMVMGQGR LLVC	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0211	SB_0211	MGNIRAWCGEG CLRGWLGYNCLG RLGGLVRMV LMS LRREGREVS RGL WLCSKGGRW FYR NGRWFLGGCLIPF RARMGGVLLGL Q	182	5	19.94206225	0.000999000999	MVLM SLRREGR	1346.727458	1506.831371	-4.578775128	No
SB_0212	SB_0212	MLRRHWREGSG WFSHNLRL EWCG RLMKRRLRRLVSS AWLGMVLWWF GGSGRRQGVSR S FIMRRCWMGWG GRWMSG	89	2	12.08128351	0.008083140878	RCWMGWGGR	1107.485449	1308.600806	6.261558495	No
SB_0213	SB_0213	MLQRWLLRSILRH GGQGLRSWWVF	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0214	SB_0214	MKERGQGWFGRI LLVRGLCMIMGV D	16	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0215	SB_0215	IMPFWVEVMME VEIWCCEIVLGN SF SSQVRSRRSRGRF WLVRRPR	12	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0216	SB_0216	IVVWKG DAGEML LVMRNPANRLPA ARRLMGFSRVGLF SLMLVRV GKR GW PVRVRRRIIRVL	77	1	18.13639941	0.000999000999	GDAGEMLLVMR	1190.578736	1334.687841	-5.255125499	No
SB_0217	SB_0217	MRVMDRAQAFV YDMFAVSM MW S LE	23	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0218	SB_0218	IPANARLPMVREV EVRGMVLSSPPIF RMSCLLRLWM MDPEHMNSMAL KKAWVQMCRNA RCGWLMPIVTIM SPSWLEVEKATIFL MSFCVRAQTAAN RVVMAPKHSVRV WISGLFSARGWK RMSKKIPATTMVL EWSRAETGVGPS MAEGSQGWRPN WADLPAAARRRP SSGVRLGLAFRRAI CCGSHELECRMN HAKARMKPMSP MRLYRIAWMAAV LASARAYHQLMSK KDMIPTSPQPMN SWNRLLAVTKISM VIRKMSRYLKNWL MFGSEFMYHSEN SMMDHVTNNAT GMNIMEK	254	11	41.69602147	0.000999000999	EVEVRGMVLSSPPI FR	1814.971241	1975.083261	-7.59190579	No
SB_0219	SB_0219	MMTSWSRHMNI VVGKRLMMKVD ATMDFT	31	5	16.35315213	0.001210653753	HMNIVVGK	896.4901828	1200.683129	5.095019283	No
SB_0220	SB_0220	MGGIREVRVRV MVVCMVITFIWS CTKIFGA	10	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0221	SB_0221	MGGEWGMGVW TWGCFLVWMRV LCC	13	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0222	SB_0222	MGMVQGRGRLC VLSGGWEVVWG VLYHSRLVLRVLR QVLLTQRWGLRH GL	15	3	17.1813702	0.006928406467	MGMVQGRGR	990.485115	1150.574672	6.469627156	No
SB_0223	SB_0223	MLSCWLRSLMVL GQWEVVVECLVS LGCCECKLVRWV GEGSLGCRMGS MCLRSVLAGCLI GWW	52	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0224	SB_0224	MVVHWMMSGVGL PWLWGVGVR	9	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0225	SB_0225	MPVSGGGFEAKV MFGCKVKY	35	1	17.54348204	0.000999000999	MPVSGGGFEAK	1078.511705	1382.713954	-2.302300023	Yes
SB_0226	SB_0226	IKATAISRMVSRIRI VKMMSVEGRLM VDIARVALPIRCM SRWPAVMLAVRR TARAIGWMSRLM VSMMTSMGMRG VGVPCGKKWARA FLILERKPMITVPA HKGMMAMARFMD SWVVGVNEWGR SPRRLVAMKMIK DTSMRDQVRPLV LCMVICFEVSLISH CWVVISRLLMRYL EVGINRGGNRMIS TAAGRPRIVGAM NEANRFSFILVLRV CYNFLFLWALVRE VGGSLCLMFLVG WWGMV	515	22	34.85655731	0.000999000999	AIGWMSRLMVSM MTSMGMRGVGV PCGK	2872.353417	3233.551855	6.845155335	No
SB_0227	SB_0227	MRALFQRFLGELIL GRWAWNCGLLH RFQSIDRSMPPVV	12	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0228	SB_0228	MMGEVCRSWRL VRRSRCTRRFSTIG GQLIWW	11	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0229	SB_0229	MEAMGLAWNQL WGVRFPLPFLSRFY VYGFECVVGWG ASM	8	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0230	SB_0230	IASVECGESAKYFD AGGSDDDYGSGG EMCSCVYVYSYCK YMVCSHDKP	8	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0231	SB_0231	IANSKKQLQTCRG FSRLFSRRREK	36	12	21.06855371	0.00218579235	GFSRLFSRR	1124.62028	1268.717786	3.615045629	Yes
SB_0232	SB_0232	IGSIGYGSLSGEYIV EEDSY	1	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0233	SB_0233	MLEIVMGMEYH MSNARVSGRKFF HRRCMSWS	53	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0234	SB_0234	IVVVRVFIMMMF VYSAMKNRAKGP AAYSMLKPETSSD SPSARSKGVRLVS ASVEMNHIMAKG HDGRSNQRCSCV VMRVERLKEPLIS NVDSRMMARVTS YEIVWATARSAPI RA	165	4	21.622387	0.000999000999	LVSASVEMNHIMA K	1528.774131	1816.968939	5.149099863	No
SB_0235	SB_0235	MNRRRRLRLTRGL GMGRGVHSRRA MVRRAKVGAVM	31	2	14.75986827	0.005020080321	GVHSRRAMVR	1167.640697	1311.744847	-1.567282974	Yes
SB_0236	SB_0236	MVDVAGFRGSLV KSFMASAKGCSSP	17	1	16.69596443	0.008484848485	MVDVAGFR	893.4429008	1053.531901	7.592758242	No
SB_0237	SB_0237	IFRSVSIRNAIAIR MGTMRSRRLAM GMLLRRGIEPLTV KF	65	3	20.62049424	0.005773672055	RLAMGMLLR	1059.604495	1235.707948	-9.330817665	No
SB_0238	SB_0238	INKLKLHRVFSSCC VMPASSRAGQFH WLKVRDSWTLVE PFMQVPI	47	3	15.62436483	0.00218579235	AGQFHWLK	985.5133602	1273.722598	-3.994077948	Yes
SB_0239	SB_0239	MLVMLEVMFLVN RRGKICRVPFTFFN LSLWACLWVDS EGNNDLLVDCRY WAVNCQFSVLIW RRLMRRRMFSCY LY	47	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0240	SB_0240	MDWSNVVWGV QLYVWDFLGSGC WAWTLS	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0241	SB_0241	MGVKFFTLSTRFF PSVQRAVPLWTN S	4	1	13.41600805	0.008403361345	FFTLSTR	870.4599302	1014.555688	6.236398252	No
SB_0242	SB_0242	MSFEVYLRRTGG VYALQGPVQLSTL LLVC	5	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0243	SB_0243	MRAIVFWGRKC SPFLATSWATPW PNVFTWVLALTL	2	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0244	SB_0244	MVGYLIPVWVLAI VCSDMLKPLS	0	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0245	SB_0245	ILCQLEFFTTQVSF SFIGEGVI	3	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

SB_0246	SB_0246	IAACLMLVPFDRG DLEGELTGTGMLA CVILLRANRKARTK PICLWGDVSPSKH FQCIARR	32	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0247	SB_0247	MGLAAVCVCWV GWAGVVLMLRV VWEWEGKMMC	21	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0248	SB_0248	MPPKDKIWNLVR LVLGFFVFGVWQ RCV	22	0	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
SB_0249	SB_0249	MLNVGAMNNR MRQESKTDAT	22	1	17.63291836	0.002030456853	MLNVGAMNNR	1118.532457	1262.64688	-9.768674111	No

Filtered list based on validation (only "Yes")

Prot_Accession	Description	Prot_Seq	PepQuery_hits	Confident_hits	Best_Score	pvalue	Peptide_Sequence	Peptide_Check_Mass	Exp_Pep_Masses	Error_mass_pm	Validation	mt gene or region	LEN(prot)
SB_0086	SB_0086	MFADRWLFSTNHKDIGTLYLLFGAWAGVL GTALLLIRAEELGQPGNLLGNDHIYVIVTA HAFVMIFFMVMPIIMIGGFGNWLVPLMIG APDMAFPRMNNMSFWLLPPSLLLLASAM VEAGAGTGWTVYPPLAGNYSHPGASVDLTI FSLHLAGVSSILGAINFITTIINMKPPAMTQY QTPLFVWSVLITAVLLLLSLPVAAGITMLLT DRNLNTFFDPAGGGDPILYQHLFWFFGHP EYVILILPGFGMISHIVTYYSGKKEPFGYMG MVMAMMSIGFLGFIVWAHMFVGMMDV DTRAYFTSATMIIAIPITGVKVFSLATLHGS NMKWSAAVLWALGFIFLFTVGGTIVLAN SSLDIVLHDTYVVVAHFHYVLSMGAVFAIM GGFIHWFPFLFSGYTLDQTYAKIHFTIMFIGV NLTFPQHFLGLSGMPRRYSYDPDAYTTW NILSSVGSFISLTAVMLMIFMIWEAFASKRK VLMVEEPSMNLEWLYGCPPPYHTFEEPVY MKSQRKRKESNPKLVSSQPHGLHDFKKV LEKPFHNFVKVKL	151	91	90.04728257	0.000999000999	VLMVEEPS MNLEWLY GCPPPYHT FEEPVYMK	3727.711022	4088.940668	-2.215073813	Yes	RefProt COX1	553
SB_0063	SB_0063	MLKLIPTIMLLPLTWLSKHHMIWINTTTHS LIISIPLFFNQNNNLFSCSPTFSSDPLTPLL MLTTWLLPLTIMASQRHLSSEPLSRKKLYLS MLISLQSLIMTFTATELIMFYIFFETTLIPTLAI ITRWGNQPERLNAGTYFLFYTLVGSPLLI IYTHNTLGSNILLTLTAQELNSWANNLM WLAYTMAFMVKMPLYGLHLWLPKAHVEA PIAGSMVLAALLKGGYGMMLRLTILNPLT KHMAYPFLVLSLWGMIMTSSICLRQTDLKS LIAYSSISHMALVVTAILIQTWFSFTGAVILMI AHGLTSSLLFCLANSNYERTHSRIMILSQGL QTLPLMAFWWLLASLANLALPPTINLLGEL SVLVTTFWSNITLLTGLNMLVTALYSLYM FTTTQWGSLSLTHHINNMKPSFTRENTLMFM HLSPIILLSLNPDITGFSSCKYSLTKTSDCESD NRGLRPLIYRESSQELLTHAPMSNNMAFST FKG	87	8	84.24539063	0.000999000999	AHVEAPIA GSMVLA AVLLK	1889.080786	2177.288583	-1.661575297	Yes	RefProt ND4	506
SB_0056	SB_0056	MAHAAQVGLQDATSPIMEELITFHDHALMI IFLICFLVLYALFLTLTKLNTNISDAQEMET VWTILPAILLVIALPSLRILYMTDEVNDPSLTI KSIGHQWYWTYEYTDYGGGLIFNSYMLPPLF LEPGDLRLDNDNRVVLPIEAPIRMMITTSQD VLHSWAVPTLGLKTAIPGRNLNQTTFATR PGVYYGQCSEICGANHSFMPIVLELIPKIFE MGPVFTL	779	656	76.27685888	0.000999000999	MMITSQD VLHSWAVP TLGLK	2226.154024	2514.353794	1.755039781	Yes	RefProt COX2	227

SB_0024	SB_0024	ILVQLQMKVMTMHTTMTTLTSLIPPLTT LVNPNKKNYPHYVKSIVASTFIISLFPPTMF MCLDQEVIIISNWHWATTQTTQLSLSFKLDY FSMMFIPVALFVTSIMEFSLWYMNSDPN INQFFKYLILITMLILVTANNLFQFIGWEG VGIMSFLISWWWYARADANTAAIQAILYNRI GDIGFILALAWFILHSNSWDPQOMALLNA NPSLTPLLGLLAAAGKSAQLGLHPWLP MEGPTPVSALLHSSTMVVAGIFLIRFHPLA ENSPLIQTTLCLGAITLFAAVCALQTQNDIK KIVAFSTSSQLGLMMVTIGINQPHLAFHIC THAFFKAMLFMCSGSIHNLNNEQDIRKMG GLLKTMPLTSTSLTIGSLALAGMPFLTGFYSK DHIIETANMSYTNAWALSITLIATSLTAYST RMILLTTGQPRFPTLTNINENNPILLNPIKR LAAGSLFAGFLITNNISPAPFQTTIPLYLKL ALAVTFLGLLTALDLNLTNKLKMKSPLECTFY FSNMLGFYPSITHRTIPYLGLLTSQNLPLLLD LTWLEKLLPKTISQHQISTSIITSTQKGMIKLY FLSFFFPLITLTLIT	113	63	76.17222727	0.000999000999	FPTLTNINE NNPTLLNPI KR	2308.253868	2596.46168	-1.396327951	Yes	RefProt ND5	612
SB_0057	SB_0057	MPQLNTTVVPTMITPMLTLFLITQKMLN TNYHLPPSPKPMKMKNYNKPWEKPKWKIC SLHSLPPQS	328	277	64.10598268	0.000999000999	MLNTNYHL PPSPKPMK	1866.948398	2299.260345	-2.484358335	Yes	RefProt ATP8	68
SB_0089	SB_0089	MMTHQSHAYHMVVKPSWPLTGALSALLM TSGLAMWFHFSMTLLMLGLLTNTLTMYYQ WWRDVTRESTYQGHHTPPVQKGLRYGMI LFITSEVFFAGFFWAFYHSSLAPTQLGGH WPPTGITPLNPLEVPLLNTSVLLASGVISITW AHHSLMENNRNQMIQALLITILLGLYFTLLQ ASEYFESPFTISDGIYGSTFFVATGFHGLHVII GSTFLTICFIRQLMFHFTSKHHFGFEAAAWY WHFVDVVWLFYVSIYWWGSYSFSMNSTV NFQLTSDNIQKRVMMNFALIMINTLALLL MIITFWLPQLNGYMEKSTPYECGFDPMSPA RVPFSMKFFLVAITFLFDEIAIALLPLPWAL QTTNPLMLVMSSLLIIIIIALSLAYEWLQKGL DWTELVYSLNKTNDFDSLNDNHIIYQMPLI YMNIMLAFTISLLGMLVYRSHLMSSLLCLEG MMLSLFIMATLMTLNTHSLLANIVPIAMLV FAACEAAVGLALLVSIISNTYGLDYVHNLNLL QC	111	69	55.11127489	0.000999000999	ESTYQGH TPPVQK	1607.769182	1895.97657	-1.701906999	Yes	RefProt COX3	520
SB_0017	SB_0017	MNENLFASFIAPTILGLPAAVLIILFPPLIPTS KYLINNRITTTQQWLKLTQKMMTMHNTK GRTWSLMLVSLIIFIATTLNLLGLPHSFTPTT QLSMNLAMAIPWAGTVIMGFRSKIKNAL AHFLPQGTPTLIPMLVIIETISLLIQPMALAV RLTANITAGHLLMHIGSATLAMSTINLPSTL IIFTILILITILEIAVALIQAYVFTLLVSLYLDHNT	130	89	51.46183017	0.001020408163	LITTQQWLI K	1242.733572	1530.93992	-1.428598097	Yes	RefProt ATP6	226
SB_0005	SB_0005	MPPSSANPDEGYKVSASTHVKTGQGVAH EVARNLHFLPQKTTMALMKLKGRRWI	111	8	31.85069597	0.001998001998	TLGQGVAH EVAR	1236.657452	1380.764859	-3.848103668	Yes	12S sens (chevauche MOTS-c)	56

SB_0085	SB_0085	INPLAQPVIISTIFAGTLTALSSHWFFTWVGL LEMNMLAFIPVLTKKMNPSTEAAIKYFLTQ ATASMILLMAILFNNMLSGQWTMTNTTN QYSSLMIMMAMAMKLGMAPFHFVWVPEV TQGTPLTSGLLLLTWQKLAISIMYQISPLN VSLLLTSLISIMAGSWGGLNQTQLRKILAYS SITHMGVMMMAVLPYNNMNTLNLTIIYILT TAFLLNLSNSTTTLLLSRTWNKLTWLTPLIP STLLSLGGLPLTGFLPKWAIIEEFTKNSLIIP TIMATITLLNLYFYLRLIYSTITLLPMSNNVK MKWQFEHTKPTPFLPTLIALTLLLPISPFML MIL	68	17	30.98510606	0.001048218029	WAIIEEFTK	1135.591332	1423.796681	-0.8391981488	Yes	RefProt ND2	347
SB_0192	SB_0192	IWAKSRLAGAGLLGRGGWMELRVLMVLA FRCEMVVGSWCWLSWVVGMR	24	3	30.49572606	0.001004016064	SRLAGAGLL GR	1069.635595	1213.734458	2.661272284	Yes	ND2, antisens	50
SB_0110	SB_0110	MYSDGYWGVSWGMRVWGLGECFSG VSDGGRIGAVGERVWVWGGGCGKL	54	4	28.02515	0.000999000999	IGAVGERV WVWGGCGK	1630.803791	1976.033302	-1.964915008	Yes	ND6 sens (NDS6 est sur autre brin), frameShift	50
SB_0007	SB_0007	MVGRFMGRGDKPTEPGDSWLSKME	54	12	27.52984605	0.001998001998	MVGRFMG RGDKPTEP GDSWLSK	2450.183426	2914.478541	0.3292530069	Yes	16S sens	25
SB_0178	SB_0178	MGLSRIEGLFGQVVCGLGMCFVLHRAI GMWLVCWLVLVWGWALWSGSEITWLGR RSLGGLRPLLGMVWVLLYDRHVIGGSL CVVVQVEAY	48	10	26.94818027	0.008474576271	RSLGGLR	757.4558484	901.5583248	-0.4346533091	Yes	COX3, antisens	95
SB_0055	SB_0055	MICCSALSPRIHLSFHRRWPDWHCISKLITR HRTTRHVLRCSPPLPCPINRSCICHRRRLHSLI SPILRLHPRPNLRQNPFIYHHRKSNFLPTT LSRPIRNPATLLGLPRCMMHMKHPICRLIHF SNSSNINNFHDLRSLRFEAKSPNSRRRLHKKP GVTMWMPPPTLPHIRTRMHKI	99	8	26.39510071	0.000999000999	TTRHVLRC SPLPCPINR	2075.124397	2333.278658	-3.949405601	Yes	COX1 sens, frameShift	180
SB_0125	SB_0125	IMCCRAGRLLEVWKRRLGLRRQRFLG	15	1	25.62950155	0.000999000999	IMCCRAGR LLEVWVK	1733.889112	2152.138023	-3.205659403	Yes	ATP6, antisens	27
SB_0132	SB_0132	MWLSRRLPGWSSARMRRLRAVPTPA HAPNNRYSVPMSLWVFNVSQRSANISGG EVKWLSEALDCKSKDRG	107	5	24.68961083	0.002040816327	WLSEALDC K	1063.500807	1408.724693	1.226049523	Yes	Non-codant (antisens entre nd2 et cox1)	73
SB_0169	SB_0169	IMRMTAPVKLQGVWMMRAVTTAMWL IEEYAMSDFRSVCRRQMELVMIMPHRDT RKG	408	14	24.68111573	0.000999000999	RQMELVMI MPHR	1539.78359	1683.892152	-3.836329152	Yes	ND4, antisens	58
SB_0058	SB_0058	MTPNRPGLSPNDLRPSHVISLPHNAPHT RPTNQHTNHMPMMARCNTKRHMPPPPH TTCPRKPSMRDNPYYLRSFFLRRIFLSLLPLQ PSPYPIRRALAPNRHHPAKSPRSPTPKHIRI TRIRSINHLSSP	35	2	24.59374409	0.000999000999	RPSMRDNP IYYLR	1679.856552	1968.066962	-3.172199407	Yes	COX3 sens, frameShift	134
SB_0141	SB_0141	IRLWRMGRRLRIRRCWSLRLVVRTPLRQGR RGFGWSLLVWR	34	5	24.18824918	0.008869179601	TPLRQGR	826.4773112	970.5764446	3.041597953	Yes	ND1, antisens	41
SB_0008	SB_0008	ILLRMSLRQIKTLNWQLTAQYLQSTNKSLLP SLSTQHRHAHAKERLKKVGTQRQILPRLFTKNI TSSITSIRGTACPVTHV	43	1	23.99545008	0.000999000999	LFTKNITSSI TSIR	1579.893313	1868.103082	-2.998073063	Yes	16S sens, complémentai re à SHLP1 et SHLP4	80
SB_0136	SB_0136	MVAMMVGMMLLFFVNSSMMAHLGKK PVSGGRPPRRRRVDGIKGVSHVSLFQVRDS SRVVLEFKLSSRNAVVVRMM	90	2	22.19791862	0.002688172043	MVAMMV GMMR	1155.509478	1347.595171	0.8468832668	Yes	ND2, antisens	77

SB_0180	SB_0180	ISKIFRGINSRMTGMKLVFAPQISEHWP	51	3	22.00463893	0.003636363636	IFRGINSR	961.5457216	1105.650919	-2.811209079	Yes	COX2, antisens	28
SB_0003	SB_0003	IIFPSHSHHTNLINTPAHPTQHTHTAANPM PRTNQTPKTPPTVYVAYLLKAMHWKCLDG LTSPHKQMGVLVLAFLALSKITHASIPVPVSS PSKSPRSKGTSIKHAAMQLKTLATPPRET AVINL	52	3	21.5925048	0.008743169399	TLSLATPPR	954.5498034	1098.655792	-3.548999013	Yes	12S sens	128
SB_0067	SB_0067	MGEGLEENPTNPITKPTLNRNKAYIIILARTT TTTNDMKNHRCISTTRTPMTPMRKTNPLM KLINHSFIDLPTPSNISAWWNFGSLLGACLIL QITGFLAMHYSPDASTAFSSIAHITRDVNY GWIIRYLHANGASMMFFICFLHIGRGLYYSF LYSETWNIIGIILLATMATAFMGYVLPWQG MSFWGATVITNLLSAIPYIGTDLVQWIWGG YSVDSPTLRFTHFFHFIPIAALATLHLLFLH ETGSNNPLGITSHSDKITFHPYYTIKDALGILL FLLSLMTLTLFSPDLLGDPDNYLANPLNTPP HIKPEWYFLFAYTILRSVPNKLGGVLALLSILI LAMIPILHMSKQSQSMMFRPLSQSLYWLLA ADLLILTWIGGQPVSYPTIIGQVASVLYFTTI LILMPTISLIENKMLKWACPCSMN	63	7	21.38188719	0.002212389381	KTNPLMK	830.4683856	1262.772811	1.41407854	Yes	RefProt CYTB	436
SB_0231	SB_0231	IANSKKQLQTCRGRSRLFSRRREK	36	12	21.06855371	0.00218579235	GFSRLFSRR	1124.62028	1268.717786	3.615045629	Yes	Non-codant (antisens entre nd2 et cox1)	24
SB_0187	SB_0187	MGVGMWGLLLRRRRRWCWGCLLVV	20	1	20.65512067	0.001	MGVGMWGLLLR	1360.699513	1520.792566	2.603800845	Yes	COX1, antisens - frameshift de Gau	27
SB_0050	SB_0050	IPTTQLKQHDPTTISHLKQANMTNLNSI HPPLRRRPPANRFLAQMGHYRRIHKKQ	41	1	20.32219288	0.000999000999	QANMTNT LNSIHPPLR	1902.973367	2047.067215	4.036097747	Yes	ND2 sens, frameShift	59
SB_0193	SB_0193	MLRFCVAGFLIHLNCLLWWMRLREWGE GLRLVRERFGMWLRWGLVFMWEEAGR MSEGCLG	65	1	19.80232168	0.000999000999	LVRERFGM WLR	1461.802668	1605.90347	0.8099306964	Yes	ND2, antisens	62
SB_0119	SB_0119	MMINKRDDMTISGRLVVCRAHGRGKRRAI SRSNKKVMATKKNFMEKGTTRAGDMGSK PHS	64	4	19.56720182	0.002424242424	DDMTISGR	893.3912624	1053.484827	3.257502508	Yes	ND3, antisens	60
SB_0015	SB_0015	MVLNLRVHRLRRTNLQLLHTSPIIPRTRRPA TPWRWQSSSTPDWSPHSYNNYITRRLALM SCPHIRLKNRCSRTSKPNHFHYTTGGML RSMLWNLWSKQFHAHRPRINSKPNLWN RARIYPMAPPLPPEPTVKLT	49	2	19.26859725	0.008849557522	NLWNRAR	928.4991084	1072.603732	-2.365401482	Yes	COX2 sens, frameShift	139
SB_0022	SB_0022	MRHNYNKHLPTTNRPKIAHCMFLFNQPH SPRSNSHSHPNPLKLRHRSRSHNRPRAYILI TILPSKLRTHSQSHHNPLSRTSNSSTPTNS FLMTSSKPR	62	11	19.17011656	0.002018163471	LHLPTTNRPK	1175.677457	1463.877578	2.758316996	Yes	ND4 sens, frameShift	99
SB_0138	SB_0138	MGCDRWHGFEWILRDGFDSHSPRNKGV	20	5	18.53962327	0.003144654088	DGFDSHSP R	1016.431152	1160.534518	-1.106494702	Yes	Non-codant (sens entre ND1 et ND2)	27
SB_0001	SB_0001	IWYFRLGGMHAMALRDAGAGAPYVAVSV FDSCILLFIAPTFNITGEHTY	37	4	18.28136369	0.001020408163	LGGMHAM ALR	1055.536813	1215.632296	1.251744129	Yes	D-loop, sens	50
SB_0225	SB_0225	MPVSGGGFEAKVMFGCKVKY	35	1	17.54348204	0.000999000999	MPVSGGGF EAK	1078.511705	1382.713954	-2.302300023	Yes	COX3, antisens	20

SB_0053	SB_0053	MKITSELVKRGLTPVFRFTVQCFTQPFLTPT DVRRLPTILYKQQRHWNTMPIRRMSWSPR HSSKPPYSSRAGPARQPSR	31	2	17.19087616	0.001001001001	HWNTMPII RR	1322.702959	1466.810724	-3.863706822	Yes	COX1 sens, frameShift	81
SB_0202	SB_0202	MGWARGGEVDRGLSITEQAPLEGYEAPPG PLSFKLWLVVFWRAVLLI	13	1	17.17685495	0.000999000999	MGWARG GEVDR	1232.572012	1376.671919	1.585656856	Yes	12S, antisens	47
SB_0068	SB_0068	ILCSFMGKQWVPPKYWLTHQQPLCISYITA SHHEYCTVP	13	4	16.55336317	0.001092896175	ILCSFMGK	897.4452082	1242.673333	-2.022001366	Yes	D-loop, sens	40
SB_0163	SB_0163	MGDCAVCDARVESEYVGMKCA	15	1	16.17840283	0.005076142132	VESEYVGE MK	1169.527414	1617.826131	1.490421419	Yes	ND5, antisens	22
SB_0238	SB_0238	INKLKLHRVFSSCCVMPASSRAGQFHWLKV RDSWTLVEPFMQVPI	47	3	15.62436483	0.00218579235	AGQFHWL K	985.5133602	1273.722598	-3.994077948	Yes	16S antisens (chevauche humanin)	45
SB_0235	SB_0235	MNRRPRLRLTRGLGMGRGVHSRRAMVRA KVGAVM	31	2	14.75986827	0.005020080321	GVHSRRA MVR	1167.640697	1311.744847	-1.567282974	Yes	ND1, antisens	34
SB_0046	SB_0046	MAEPGNRMKLTQSEVQFLFTTYPWPTS YSSLYPF	18	3	13.93237678	0.008869179601	MAEPGNR	773.3490056	917.4533302	-2.447302268	Yes	Non-codant (antisens entre 16S et ND1)	37
SB_0032	SB_0032	MHQSCPEMKTFFQGGQIREKVFNSTISTQS	20	2	13.83337106	0.003003003003	MHQSCPE MK	1217.535493	1722.853313	2.770027249	Yes	Non-codant (sens entre CYTB et d- loop)	30
SB_0029	SB_0029	IHRPPPIQHLMRMMKRLRTPWRLPDPNNH HRTIPSHALLTRRLNRLFINRPHHSRRLWL NHPLPSRQWRNLNLYLPLTHRARPLMRIISL LRNLKRRHYPPACNYSNLSHLRCPVVRPNIL RGHSNYKLTIRHPMHWDSPSSMNLRRLLS RQSHPHLILYLSLHLALHYCSPSNTPPPILAR NGIKQPPRNHLPPFR	121	2	13.82029738	0.004424778761	LRLTPWR	940.5606422	1084.660857	1.728571636	Yes	RefProt CYTB	199
SB_0172	SB_0172	MIVRGRSQVVSIRRGVVRGSEKVEQELNR LLLWLKNSRGMMLMIRLWVVLQIMCFL ESHVSGSNMIVGTISFISGVLGYVRSLSGHM CWRLRLVGLGPPLRRRQRLVWQ	155	1	12.06194623	0.009433962264	GRSQVVSIR	1000.577749	1144.676438	2.971509154	Yes	ND4, antisens	114
SB_0049	SB_0049	IFYLSRPRNKHASFYSSSNQKNKPSFHRSC QVFPHASNRHNPNSYPLQYTLRTMNH NQYYQSMLIINNHNSYKTRNSPLSLLSPR GYPRHPSDIRPASSHMTKTSPLNHMPNLS LTKRKPSPHSLNLIHHSRQLRWIKPNPATQN LSMLLNYPHRMNNSSSTVQP	122	4	12.05265959	0.007658643326	IFYLSRPR	1050.59742	1338.79671	3.632896529	Yes		172
SB_0103	SB_0103	MLTSKYSNQPSTITHQLQLQSHSPTRMPT NLPTLNST	125	3	11.44115982	0.002997002997	MPTNLPTL NST	1187.585595	1347.676777	4.321420165	Yes	D-loop, sens	38

ANTIBODY PRODUCTION (antigen in red)

MGLSRIEGLFGQVVCVGLGMCFLVLRHAIIGMWLVCLVGLVWGLWALWGS**SEITWLGRRSLGGLRGP**LLGVMGWVLLYDRHVIGGSLCVVVQVEAY
MWSLPR**RLPGWPSSARMRRLRAVPRTPAHAPNNRY**SVPMSLWFVENSQRSANISGGEVKWLSEALDCKSKDRG
IMRMTAPVKLQGVWMRMAVTTTRAMWLIIEYAMSDFRSVCRRQ**MELVMIMPHRDSTRKG**
MRHNYNKLHLPTTNRPKIAHCMLFNQPHSPRSNSHSHPNPLKLRHRSRSHNRPRAVILITILPSKLLRTHSQSHHNPISRTSNSTPTNSFLMTSSKPR

Table S5. List of proteins found by mass spectrometry following immunoprecipitation of MTALTND4

PROTEIN NAME	ACCESSION NUMBER	MOLECULAR WEIGHT	UNIQUE PEPTIDES COUNT	PROBABILITY
Protein [synthetic construct, MTALTND4]	AEX63613.1	16 kDa	5	100%
Glutathione peroxidase 1	P07203	22 kDa	13	100%
CLIP-associating protein 2	O75122	141 kDa	2	99%
60S ribosomal protein L26	P61254 (+1)	17 kDa	3	100%
Uncharacterized protein ZSWIM9	Q86XI8	70 kDa	3	99%
Uncharacterized protein C20orf194	Q5TEA3	132 kDa	2	99%
Histone H4 OS=Homo sapiens	P62805	11 kDa	3	100%
Tensin-3	Q68CZ2	155 kDa	2	99%
Tudor domain-containing protein 3	Q9H7E2	73 kDa	1	99%
Glutamine amidotransferase-like class 1 domain-containing protein 1	Q8NB37	23 kDa	2	100%
60S ribosomal protein L10	P27635	25 kDa	1	100%
Proteasome subunit alpha type-2	P25787	26 kDa	1	100%
Glutaredoxin domain-containing cysteine-rich protein 2	A6NFK2	28 kDa	2	99%
40S ribosomal protein S9	P46781	23 kDa	2	100%
40S ribosomal protein S3a	P61247	30 kDa	2	100%
Putative 60S ribosomal protein L13a protein RPL13AP3	Q6NVV1	12 kDa	2	100%
40S ribosomal protein S11	P62280	18 kDa	2	100%
60S ribosomal protein L11	P62913	20 kDa	1	100%
Polyubiquitin-B	P0CG47 (+1)	26 kDa	1	99%
60S ribosomal protein L13	P26373	24 kDa	1	99%
Proteasome subunit alpha type-6	P60900	27 kDa	1	99%
TBC1 domain family member 2A	Q9BYX2	105 kDa	1	99%
Probable imidazolonepropionase	Q96NU7	47 kDa	2	99%
Transforming acidic coiled-coil-containing protein 3	Q9Y6A5	90 kDa	1	99%
60S ribosomal protein L27	P61353	16 kDa	1	99%
Toll-like receptor 7	Q9NYK1	121 kDa	1	99%
Conserved oligomeric Golgi complex subunit 2	Q14746	83 kDa	1	100%

BAG family molecular chaperone regulator 2	O95816	24 kDa	1	99%
Proteasome subunit beta type-6	P28072	25 kDa	1	99%
60S ribosomal protein L12	P30050	18 kDa	1	99%
60S ribosomal protein L14	P50914	23 kDa	1	99%
40S ribosomal protein S18	P62269	18 kDa	1	99%
60S ribosomal protein L19	P84098	23 kDa	1	99%
Protein phosphatase 1 regulatory subunit 3F	Q6ZSY5	83 kDa	1	99%
Transmembrane emp24 domain-containing protein 4	Q7Z7H5	26 kDa	1	99%
Ras-related protein Rab-18	Q9NP72	23 kDa	1	99%
MAGUK p55 subfamily member 4	Q96JB8	73 kDa	1	99%
Actin, cytoplasmic 1	P60709 (+1)	42 kDa	1	99%
Zinc finger protein 775	Q96BV0	60 kDa	1	99%