



HAL
open science

Analyse bio-informatique du protéome mitochondrial et du spectre des mutations de la protéine Opa1

Marc Ferré

► **To cite this version:**

Marc Ferré. Analyse bio-informatique du protéome mitochondrial et du spectre des mutations de la protéine Opa1. Biologie cellulaire. Université d'Angers, 2009. Français. NNT: . tel-00457327

HAL Id: tel-00457327

<https://theses.hal.science/tel-00457327>

Submitted on 17 Feb 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ANALYSE BIO-INFORMATIQUE DU PROTÉOME MITOCHONDRIAL ET DU SPECTRE DES MUTATIONS DE LA PROTÉINE OPA1

THÈSE DE DOCTORAT

Spécialité : BIOLOGIE CELLULAIRE

ÉCOLE DOCTORALE BIOLOGIE/SANTÉ

Présentée et soutenue publiquement

le 15 décembre 2009

à Angers

par Marc FERRÉ

Devant le jury ci-dessous :

Christophe BÉROUD (rapporteur), Maître de conférences (HDR), Université de Montpellier 1

Dominique LAVENIER (rapporteur), Professeur, ENS-Cachan Bretagne

Guy LENAERS (examinateur), Directeur de recherche, CNRS

Vincent PROCACCIO (examinateur), Professeur, Université d'Angers

Yves MALTHIÈRY (examinateur), Professeur, Université d'Angers

Directeur de thèse : Pascal REYNIER, Professeur, Université d'Angers

Nom et coordonnées du Laboratoire :

Laboratoire de biochimie et biologie moléculaire — INSERM U694

Centre Hospitalier Universitaire d'Angers, 4 rue Larrey, 49933 Angers Cedex 9, France

AU DIRECTEUR DE THÈSE,

Monsieur Pascal REYNIER,

Vous m'avez transmis les connaissances essentielles en *mitochondriologie*. Vous avez toujours pris le temps de m'aider et de me conseiller pendant la réalisation de cette thèse. Je vous remercie encore pour votre patience et votre bienveillance.

Vous m'avez fait l'honneur d'accepter la direction de cette thèse, veuillez trouver ici le témoignage de mon profond respect et de ma gratitude.

À Monsieur Christophe BÉROUD,
À Monsieur Dominique LAVENIER,
À Monsieur Guy LENAERS,
À Monsieur Vincent PROCACCIO,

Vous me faites l'honneur de juger ce travail et de participer au jury.

Je vous en remercie et vous témoigne de toute ma reconnaissance et de mon profond respect.

À Monsieur Yves MALTHIÈRY,

Vous m'avez accueilli dans votre laboratoire et fait bénéficier de votre expérience. Vous me faites l'honneur de juger ce travail et de participer au jury.

Veillez trouver ici l'expression de toute ma reconnaissance.

À Monsieur Yves TOURMEN,

Vos connaissances sont pour moi un exemple. Vos travaux furent à l'origine de ce travail et j'ai beaucoup appris à vos côtés.

Je vous en remercie et vous assure toute ma reconnaissance ainsi que mon plus profond respect.

Merci,

À ma femme Véronique et Simon le magnifique fils qu'elle m'a donné,

À mes parents,

À Arnaud pour sa relecture attentive et ses commentaires pertinents,

À l'ensemble de l'équipe de recherche, Christophe, Claire, Dominique B., Dominique L., Gilles, Marie-Anne, Naïg, Patrizia, Pascale, Valérie, Virginie, et les autres,

À mes amis du laboratoire, Anne, Damien, Delphine, Jean-François, Julien, Lætitia, Natacha, Olivier, et les autres,

À mes amis des autres laboratoires, David, Fabien, François,

À mes enseignants de l'ICES et de l'ESEO,

À mes amis, Alban, Claire, Jean-Jacques, Morgana, Vincent et les autres.

AVANT-PROPOS

*Les espèces qui survivent ne sont pas les espèces les plus fortes,
ni les plus intelligentes, mais celles qui s'adaptent le mieux aux changements.*

Charles Darwin

Ma formation d'ingénieur diplômé de l'ESEO d'Angers me prédisposait plus naturellement à une activité informatique dans les technologies de l'information et les télécommunications plutôt qu'à une inscription en thèse d'université auprès de l'École Doctorale d'Angers. J'ai d'abord travaillé de 2001 à 2004 dans le laboratoire INSERM E0018 (puis U694) dirigé par Yves Malthiery, sur un poste financé par les crédits du Contrat État Région postgénomique/bio-informatique. Cette première expérience, arrivant à une période déterminante pour la recherche en biologie, a influencé la suite.

Thomas S. Kuhn a rendu célèbre la théorie d'une science progressant de manière fondamentalement discontinue, c'est-à-dire non par accumulation mais par rupture. J'ai eu la chance d'intégrer mon laboratoire alors que s'opérait une de ces ruptures (une « révolution scientifique » selon Kuhn). Les ordinateurs et le *Web* changeaient aussi rapidement que définitivement la recherche en biologie. L'informatique biologique a évolué pendant des décennies à la périphérie de la biologie traditionnelle, mais, en l'espace de quelques années, la pléthore de nouvelles données biologiques produites par les grands projets génomes et, par la suite, le développement d'applications informatiques pour l'analyse des données génomiques ont commencé à influencer tous les aspects de la recherche en biologie. Les recherches qui, auparavant, commençaient en laboratoire sont à présent amorcées sur des ordinateurs, les scientifiques effectuant des recherches dans les bases de données pour y trouver des informations susceptibles de suggérer de nouvelles hypothèses. Les ordinateurs sont des équipements précieux pour la compréhension de tout système qui doit être décrit de manière formelle, ce qui a permis l'essor de disciplines telles que l'informatique biologique et plus récemment la bio-informatique.

La bio-informatique est l'application des technologies de l'information à la gestion des données biologiques. C'est une discipline en pleine évolution. En l'espace de vingt ans, le stockage de données biologiques dans les banques de données publiques est devenu une pratique courante, et la taille de ces banques augmente encore aujourd'hui de manière exponentielle. Le nombre de publications en biologie a également augmenté dans les mêmes proportions. Il serait désormais impossible de se tenir au courant de toutes les

avancées du domaine sans l'aide d'outils informatiques. Le Web a permis aux utilisateurs du monde entier d'interagir avec les programmes et les banques de données d'autres sites, sous réserve qu'ils sachent mettre au point les outils appropriés. Mais la bio-informatique est avant tout et principalement une discipline biologique. Il est souvent moins question de développer des algorithmes parfaits et élégants que de répondre à des questions pratiques. Il est essentiel que les bio-informaticiens comprennent les problèmes biologiques tout autant que les solutions informatiques à mettre en œuvre pour produire des outils utilisables à leur résolution. Toute investigation en bio-informatique nécessite de cerner des hypothèses scientifiques qui, en raison de leur complexité, peuvent considérablement compliquer la programmation et la représentation des données.

C'est donc dans ce contexte exaltant que j'ai rencontré dans mon laboratoire des personnes, étudiants, praticiens hospitaliers et enseignants-chercheurs, aussi passionnés que passionnants, qui m'ont permis d'acquérir une culture biologique. Ma formation m'avait apporté la connaissance de la gestion des masses de données et de leur partage, mais la mitochondrie était pour moi une entité abstraite, fournissant l'énergie à la cellule. Depuis, je suis entré en contact avec l'intimité de cet organite en essayant de décrypter ses codes, de découvrir ses « bugs », ou plutôt ses erreurs d'information probablement à l'origine de pathologies. Les mitochondries jouent en effet non seulement un rôle déterminant dans la production énergétique et dans la mort cellulaire programmée des cellules eucaryotes, mais elles sont en outre impliquées dans un grand nombre de processus pathologiques (cancer, diabète, obésité, maladies génétiques et neurodégénératives).

Ainsi, en 2004, j'ai été recruté par le Centre Hospitalier Universitaire d'Angers en tant qu'ingénieur bio-informaticien dans le cadre de la formation d'un réseau national coordonné par le ministère de la Santé et des Affaires sociales (étude génétique des pathologies neuromusculaires et neurosensorielles et des retards mentaux). Conjointement, j'ai eu l'opportunité de m'inscrire en thèse d'université et ainsi, durant les cinq dernières années, j'ai mené parallèlement mon projet de recherche et les développements bio-informatiques du laboratoire de diagnostic. Ce mémoire est basé sur des articles publiés dans des journaux scientifiques et est organisé autour de quatre chapitres. D'abord, le sujet est introduit en s'appuyant sur des données bibliographiques pour présenter le génome mitochondrial et ses polymorphismes, le protéome mitochondrial et les pathologies associées à une protéine mitochondriale en particulier : Opa1. Puis, après avoir rappelé les objectifs de ce travail de thèse, celui-ci est exposé autour de quatre articles publiés ainsi qu'une étude non publiée. Enfin, une conclusion générale sera suivie d'annexes, comportant des informations complémentaires pour approfondir le travail exposé, mais qui ne sont pas indispensables à la compréhension du manuscrit.

TABLE DES MATIÈRES

Avant-Propos	IX
Table des matières.....	XI
Liste des figures	XV
Liste des tableaux	XXI
Abréviations	XXIII
Chapitre 1 : Introduction.....	1
1.1 Le génome mitochondrial	1
1.1.1 L'organisation des gènes	2
1.1.2 Le code génétique mitochondrial	5
1.1.3 L'héritage maternel	6
1.1.4 Un grand nombre de copies.....	7
1.1.5 L'évolution du génome mitochondrial.....	8
1.1.6 Le rythme rapide de l'évolution	8
1.1.7 L'ADNmt et l'évolution de l'homme moderne	9
1.2 Les polymorphismes de l'ADN mitochondrial	11
1.2.1 Polymorphismes de l'ADNmt humain et évolution humaine	11
1.2.2 Influence des polymorphismes de l'ADNmt	18
1.3 Le protéome mitochondrial	21
1.3.1 Les études protéomiques.....	23
1.3.2 Les progrès et les perspectives dans l'analyse protéomique de la mitochondrie.....	31
1.3.3 Les apports complémentaires de la génétique, de la bio-informatique et des approches transcriptomiques	31
1.3.4 L'origine et l'évolution du protéome mitochondrial.....	36
1.3.5 Les approches intégratives	36
1.3.6 Dysfonctionnements de la mitochondrie dans les pathologies humaines.....	39
1.4 Les pathologies associées à Opa1 : des mitochondriopathies à part entière	43
1.4.1 Description clinique des AOAD	46
1.4.2 <i>OPA1</i> est le gène majeur associé à l'AOAD	48
1.4.3 Expressivité variable de l'AOAD	49
1.4.4 Spectre des mutations du gène <i>OPA1</i>	55
1.4.5 Fonction de la protéine Opa1	57
1.4.6 Physiopathologie.....	60
1.4.7 Modèles animaux.....	65

1.4.8	Conseil génétique	66
1.4.9	Perspectives d'avenir	66
Chapitre 2 : Objectif de la thèse		69
Chapitre 3 : Travail de thèse		71
3.1	Analyse du protéome mitochondrial humain à travers l'évolution	71
3.1.1	Premier article : « Les mitochondriopathies impliquent majoritairement des protéines d'origine procaryote »	74
3.1.2	Étude complémentaire non publiée : méthode de prédiction à haut débit de nouvelles protéines mitochondriales par génomique comparative	83
3.1.3	Synthèse	101
3.2	Étude de la protéine mitochondriale Opa1 et des pathologies associées	103
3.2.1	Second article : « eOPA1: une base de données pour les mutations du gène <i>OPA1</i> »	106
3.2.2	Troisième article : « Dépistage moléculaire de 980 cas suspectés de neuropathie optique héréditaire et publication de 77 nouvelles mutations du gène <i>OPA1</i> »	114
3.2.3	Quatrième article : « L'atrophie optique dominante liée à <i>OPA1</i> n'est pas fortement influencée par le fond génétique mitochondrial »	130
3.2.4	Synthèse	139
3.3	Autres publications	141
Chapitre 4 : Discussion générale, conclusion et perspectives		147
Références		153
Annexes		173
Annexe A : Le filtre <i>mitopid1</i>		173
	Entrée/sortie	173
	Traitement des <i>hits</i>	175
	Fusion	176
	Couverture	179
	Extrémité N-terminale	179
	Extrémité C-terminale	180
	Code source	181
Annexe B : Le filtre <i>mitopid2</i>		189
	Entrée/sortie	189
	Recherche dans la table des gènes connus	189
	Code source	192
Annexe C : Le filtre <i>mitopid3</i>		201
	Entrée/sortie	201
	Couverture	201

Extrémités N- et C-terminales.....	201
Code source.....	203
Annexe D : Le filtre <i>mitopid11</i>	209
Entrée/sortie.....	209
Code source.....	210
Index	215
Résumé.....	218

LISTE DES FIGURES

Figure 1. Représentation schématique de la mitochondrie avec les détails de la membrane interne et des complexes respiratoires (I–IV). Les mitochondries sont formées par deux membranes, une membrane externe ayant principalement des propriétés de perméabilité et une membrane interne où se produisent le transport d'électrons et la phosphorylation oxydative. L'espace interne, appelé matrice, contient des enzymes du métabolisme, l'acide désoxyribonucléique mitochondrial (ADNmt), et la machinerie moléculaire pour sa réplication et son expression. Le flux d'électrons (e^-) à travers les complexes respiratoires entraîne une diminution de l'énergie libre, qui est utilisée pour pomper les protons (H^+) de la matrice vers l'espace intermembranaire. L'énergie conservée dans ce gradient de protons est utilisée pour promouvoir la synthèse de l'adénosine 5'-triphosphate (ATP) par l'ATP synthase en inversant le flux de protons de l'espace intermembranaire vers la matrice. ADP : adénosine 5'-diphosphate ; NAD^+ : forme oxydée du nicotinamide adénine dinucléotide ; NADH : forme réduite du nicotinamide adénine dinucléotide ; QU: ubiquinone..... 1

Figure 2. Cartographie de l'ADNmt de mammifère. Le cercle extérieur représente le brin lourd (H) et le cercle interne représente le brin léger (L), en fonction de leur sédimentation isopycniqne dans un gradient de chlorure de césium. La structure à trois brins indique la boucle-D. O_H et O_L spécifient les origines de réplication. La répartition des gènes entre les brins et les principales régions régulatrices sont indiquées. Les gènes codant des protéines sont les suivants: ATP6, ATP8 : gènes codant deux sous-unités du complexe V (ATPase) ; CO I, II, III : gènes codant trois sous-unités du complexe IV (cytochrome c oxydase) ; Cytb : gène codant une sous-unité du complexe III (cytochrome b) ; ND1, 2, 3, 4, 4L, 5, 6 : gènes codant 7 sous-unités du complexe I (NADH déshydrogénase) ; ARNr 12S, 16S : ARNr des petites et grandes sous-unités ribosomales ; 22 ARN de transfert..... 4

Figure 3. Transmission maternelle de l'ADNmt. La flèche noire indique l'ancêtre commun de toute la lignée qui en est dérivée. 7

Figure 4. Arbre phylogénétique des haplogroupes de l'ADNmt. L'information obtenue avec l'ensemble des différentes analyses (RFLP, séquençage de la boucle-D et séquençage complet) a été fusionnée. La longueur des lignes droites reliant les haplogroupes entre eux est proportionnelle au nombre de mutations les différenciant. Les mutations sont indiquées soit avec la position où la mutation a lieu, soit avec la position du site de restriction qui est acquis (+) ou perdu (-). Dans ce dernier cas, une lettre indique l'enzyme de restriction qui reconnaît le polymorphisme. Le code à une lettre pour les

enzymes de restriction est le suivant : a : *AluI* ; b : *Avall* ; c : *DdeI* ; e : *HaeIII* ; f : *HhaI* ; g : *HinfI* ; h : *HpaI* ; l : *MspI* ; j : *MboI* ; k : *RsaI* ; l : *TaqI* ; m : *BamHI* ; n : *HaeII* ; o : *HincII* ; q : *NlaIII* ; r : *BfaI* ; s : *AccI* ; t : *BstOI* ; u : *MseI*. Étant donné que les haplogroupes de l'ADNmt sont spécifiques d'un continent, le continent où l'haplogroupe est présent est indiqué. Les haplogroupes désignés comme asiatiques comprennent ceux qui figurent chez les Amérindiens et les Polynésiens. L'haplogroupe X désigné comme asiatique se retrouve également en Europe et chez les Américains d'origine... 16

Figure 5. Migration et distribution géographique des haplogroupes mitochondriaux. La datation est exprimée en années avant le présent. Le taux de mutation est de 2,2–2,9 % par million d'années. +/-, +/+ et -/- : présence (+) ou absence (-) des sites de restriction *DdeI* 10394 / *AluI* 10397 ; * : présence du site de restriction *RsaI* 16329. D'après MITOMAP[42] (<http://www.mitomap.org/WorldMigrations.pdf>)..... 17

Figure 6. Vue d'ensemble des processus cellulaires dans les mitochondries humaines. Schéma de section d'une mitochondrie montrant une sélection de protéines ou de complexes protéiques impliqués dans : la translocation des protéines [complexe des translocases de la membrane externe (*translocase of outer membrane*, TOM), protéines dites β -barrel distribuées dans la membrane externe (*topogenesis of mitochondrial outer-membrane β -barrel proteins*, TOB), trois complexes de translocases de la membrane interne TIM22, TIM23 et OXA1], le repliement des protéines (Hsp60, Hsp70) et la dégradation ; la morphologie mitochondriale (Mfn1, Mfn2 et Opa1), la réplication, la transcription, la maturation des ARNr et ARNt, et la traduction (ribosomes) des gènes codés par l'ADNmt, la biosynthèse des centres Fe/S [complexe d'assemblage des centres fer-souffre (*iron-sulfur cluster assembly complex*, ISC)] et l'exportation (Atm1) ; l'échange des métabolites [transporteurs de l'ADP et l'ATP (*ADP-ATP carrier*, AAC) ou échangeurs de nucléotides adényliques (*adénine nucléotide translocator*, ANT), porines ou échangeurs d'anions en fonction du potentiel (*voltage dependant anion channel*, VDAC)] ; la phosphorylation oxydative (complexes I, II, III, IV, et F_1F_0 -ATP-synthase) ; et l'apoptose (cytochrome c). La synthèse d'ATP dans la matrice mitochondriale est indiquée à l'aide du gradient de protons à travers la membrane interne. Les processus métaboliques de base tels que le cycle de Krebs, le métabolisme de certains acides aminés et lipides, et la biosynthèse de l'hème sont omis. Les invaginations de la membrane interne, appelées crêtes, sont représentées comme elles apparaissent souvent dans les représentations de sections de mitochondries en microscopie électronique. La partie de la crête où commence l'invagination, appelée jonction de la crête, est visible uniquement lorsque celle-ci se trouve dans le même plan que le plan de section..... 22

- Figure 7. Coupe axiale de l'œil droit. *D'après l'Atlas d'anatomie 3D Primal Pictures (usage éducation seul autorisé)*.....43
- Figure 8. Fond d'œil : a) d'un disque optique droit normal ; et b) d'une atrophie diffuse du disque droit chez un patient atteint d'AOAD avec une vision de 20/200.....44
- Figure 9. Champs visuels, explorés par périmétrie 30-2 Humphrey automatisée statique. Seuls les enregistrements de l'œil droit sont montrés : a) chez un patient ayant une vision normale et un fond d'œil normal (Figure 8a) ; et b) chez un patient atteint d'AOAD (fond de l'œil Figure 8b) montrant une grande irrégularité temporale paracentrale.47
- Figure 10. Représentation schématique de la structure du gène *OPA1* humain. Les exons sont représentés par des rectangles noirs ; ceux soumis à un épissage alternatif apparaissent en gris. Les domaines de la protéine Opa1 sont représentés en arrière-plan. GED : domaine GED (*GTPase effector domain*) putatif.....55
- Figure 11. Données provenant de la base de données eOPA1 (juillet 2009) [179]. Distribution des 204 mutations pathogènes dans le gène *OPA1* en fonction ses exons impliqués, comprenant leur voisinage intronique. La fréquence des mutations d'*OPA1* chez les patients est symbolisée par des étoiles, le nombre d'étoiles étant corrélé à la fréquence des mutations dans chaque exon. GED : *GTPase effector domaine*.56
- Figure 12. Données provenant de la base de données eOPA1 (juillet 2009) [179]. Répartition des 204 mutations pathogènes du gène *OPA1* en fonction de leur conséquence au niveau de la protéine.....57
- Figure 13. Structure et fonction d'Opa1 dans les mitochondries. (a) Représentation schématique des rôles d'Opa1 dans la structuration des crêtes, la jonction des crêtes et l'interaction entre les membranes externe (OMM) et interne (IMM) mitochondriales. Les points roses représentent les isoformes Opa1_L ancrées à la membrane, les points verts représentent les isoformes Opa1_S qui n'interagissent pas avec la membrane, les deux étant nécessaires pour la structure de crêtes et les ponts entre l'OMM et l'IMM. Les points rouges représentent les isoformes d'Opa1 faisant la jonction des crêtes. Le cytochrome c, représenté par des étoiles vertes, est le plus souvent ségrégué dans le volume des crêtes. (b) Les isoformes Opa1_L et Opa1_S, pouvant inclure l'exon 4, sont nécessaires pour les structures membranaires des crêtes et interagissent avec les complexes I (CI), II (CII) et III (CIII) de la chaîne (ellipses brun clair) pour promouvoir le transfert d'électrons dans l'IMM et maintenir le potentiel de membrane ($\Delta\Psi_m$). (c) Les isoformes OPA1_S (points verts) interagissent avec des protéines mitofusines (en vert et bleu dans l'OMM) pour promouvoir la fusion du réseau mitochondrial avec les isoformes Opa1_S (points verts) et éventuellement les isoformes Opa1_L (points violets). (d) Les isoformes antiapoptotiques d'Opa1 (points rouges) structurent la jonction de crêtes pour empêcher la mobilisation cytochrome c. Elles pourraient être ciblées par des protéines

proapoptotiques (en jaune) pour démanteler la structure des jonctions des crêtes.
D'après Lenaers et coll. (Int J Biochem Cell Biol, 2009) [185]. 59

Figure 14. Représentation schématique d'un neurone de cellule ganglionnaire de la rétine (CGR) avec son réseau mitochondrial et des contraintes environnementales. Le soma de la CGR est localisé à gauche, dans la rétine, tandis que la synapse de la CGR est située dans le corps genouillé latéral. Entre les deux, l'axone est démyélinisé dans le globe oculaire, pour devenir myélinisé à partir de la lame criblée de la sclère. La région démyélinisée de l'axone et le soma subissent l'assaut de la lumière du jour (soleil jaune et étincelles violettes) et de la pression intraoculaire (PIO), alors que l'axone myélinisé est protégé de la lumière et des contraintes par la pression intracrânienne (PIC). Ce stress favoriserait la production de dérivés réactifs de l'oxygène (ROS) et l'apoptose. La surproduction de ROS augmenterait l'activité de fusion du réseau mitochondrial (double flèche rose). Pour fournir l'ATP nécessaire à la transmission des signaux visuels, les mitochondries (points rouges) ont des besoins en énergie particulièrement élevés (étincelles roses) dans la région démyélinisée de l'axone dans le globe oculaire. Le réseau mitochondrial est inégalement réparti, avec de nombreuses varicosités constituées de mitochondries agrégées (points rouges) dans la région démyélinisée de l'axone, particulièrement autour de la lame criblée de la sclère, alors que les mitochondries sont rares et allongées (ovales roses) dans la région myélinisée de l'axone dans le nerf optique. *D'après une communication orale de Guy Lenaers (Colloque MeetOchondrie, La Grande-Motte, Hérault, 4 mai 2009).* 61

Figure 15. Fragmentation complète ou partielle du réseau mitochondrial dans des cellules mutées Opa1 par rapport aux témoins. Les mitochondries ont été marquées avec la sonde fluorescente Mitotracker®. Vingt et un plans-images ont été acquis le long de l'axe Z par incréments de 0,3 µm avec un microscope Leica DMI6000B (Microsystems GmbH, Wetzlar, Allemagne) et un appareil d'acquisition Roper CoolSnap HQ2 (Roper Scientific GmbH, Allemagne). Les piles d'images ont été déconvoluées à l'aide d'un algorithme de Meinel et du logiciel MetaMorph (Molecular Devices, CA, USA). Le signal fluorescent mitochondrial a été seuillé et l'image binarisée. La longueur des mitochondries et leur nombre ont été déterminés à l'aide d'une analyse morphométrique intégrée des régions créées autour des mitochondries. 64

Figure 16. Carte PCI ReMIX comportant 64 Go de mémoire FLASH et un composant reconfigurable Xilinx Virtex-2 Pro..... 85

Figure 17. Flux de traitement d'une sortie iBLAST (ReMIX) avec les filtres *mitopid* et leurs différents paramètres (couverture : rapport en pour cent entre la taille de l'alignement et la taille de la protéine bactérienne ; gap : distance maximale entre deux HSP considérés

contigus ; Nterm : limite en position N-terminale ; Cterm : limite en position C-terminale).	88
Figure 18. Flux de traitement d'une sortie iBLAST (ReMIX) intégrant le filtre <i>mitopid11</i> avec les différents paramètres (couverture : rapport en pour cent entre la taille de l'alignement et la taille de la protéine bactérienne ; Nterm : limite en position N-terminale ; Cterm : limite en position C-terminale).	89
Figure 19. Performance du filtrage : nombre d'HSP contigus retenu en fonction de la couverture (en pour cent).....	93
Figure 20. Sensibilité du filtrage en fonction de la couverture (en pour cent)	93
Figure 21. Spécificité du filtrage en fonction de la couverture (en pour cent).	94
Figure 22. Représentation graphique des résultats obtenus dans la banque de données RefSeq [223], avec différents outils de prédiction de gènes (<i>Gene Predictions</i>) et parmi les séquences des ARNm humains de GenBank [44] sur le locus 19q13.32 humain (UCSC Genome Browser, version hg18, mars 2006).	98
Figure 23. Test de fusion d'un HSP au niveau de sa séquence requête.	177
Figure 24. Test de chevauchement d'un HSP au niveau de sa séquence requête.	178
Figure 25. Test sur la partie N-terminale de la requête.....	179
Figure 26. Recherche dans la table des gènes connus selon la position du CDS par <i>mitopid2</i>	190
Figure 27. Analyse de la correspondance des HSP avec les exons des gènes de la base de l'UCSC.	191
Figure 28. Test de la couverture des groupes de HSP fusionnés avec les exons des gènes correspondants.	202

LISTE DES TABLEAUX

Tableau 1. Le code génétique mitochondrial chez les vertébrés.....	5
Tableau 2. Haplogroupes de l'ADNmt définis par analyse de restriction à haute résolution. 13	
Tableau 3. Haplogroupes mitochondriaux et longévité.....	19
Tableau 4. Haplogroupes mitochondriaux et maladies neurodégénératives	19
Tableau 5. Termes et définitions.....	25
Tableau 6. Comparaison quantitative de différentes approches pour déterminer le protéome mitochondrial chez la levure, l'homme et la souris.	27
Tableau 7. Bases de données et outils relatifs au protéome mitochondrial.....	34
Tableau 8. Loci d'atrophies optiques héréditaires. AD : autosomique dominante ; RLX : récessive liée à l'X ; AR : autosomique récessive.	45
Tableau 9. Présentations cliniques associées aux mutations d'Opa1. CGR : cellules ganglionnaires de la rétine ; LCR : liquide céphalorachidien ; NOHL : neuropathie optique héréditaire de Leber ; OXPHOS : phosphorylation oxydative (<i>oxidative phosphorylation</i>) ; CPEO : ophtalmoplégie chronique progressive externe (<i>chronique progressive external ophthalmoplegia</i>) ; RRF : Fibres rouges déchiquetées (<i>ragged red fibers</i>) ; SNC : système nerveux central.	50
Tableau 10. Correspondance entre les identifiants <i>UCSC Genome Browser Database</i> et Swiss-Prot. <i>spID</i> : identifiant Swiss-Prot ; <i>proteinID</i> : identifiant UCSC ; <i>alias</i> : nom du gène donné par l'UCSC.....	92
Tableau 11. Liste des 511 protéines mitochondriales PH+ candidates, identifiées avec un paramètre de couverture de 23 %. Les identifiants Swiss-Prot sont séparés par des virgules, avec les résultats vrais positifs en gras et les redondances regroupées par les couleurs de police bleue, rouge et verte.....	94
Tableau 12. Comparaison des prédictions obtenues par notre approche (mitopid) en fixant le paramètre de couverture à 23 % avec celles de cinq outils bio-informatiques de référence : Predotar [85], TargetP [220], SubLoc [221], MitoProt [83] et PSORT-II [222]. Les protéines sont nommées en faisant référence à leur <i>alias</i> , c'est-à-dire le nom du gène correspondant donné par l'UCSC (voir Tableau 10). Une protéine est notée « + » si sa localisation est considérée comme mitochondriale, « - » sinon. Les outils ont été utilisés avec leurs paramètres par défaut.	97
Tableau 13. Extrait de dix résultats parmi les 350 165 HSP contigus retenu par le filtrage en fixant le paramètre de couverture à 23 %. Le contenu du fichier original est présenté sous forme de tableau en ne retenant que certaines colonnes. <i>query_name</i> : numéros	

d'accession de la protéine bactérienne requête ; *strand* : sens de l'alignement ;
query_start : position du début de l'alignement de la séquence requête ; *query_end* :
position de la fin de l'alignement de la séquence requête ; *sbjct_start* : position du début
de l'alignement de la séquence cible (chromosome 19 humain) ; *sbjct_end* : position de
la fin de l'alignement de la séquence cible. 98

Tableau 14. Extrait d'un fichier de sortie de ReMIX..... 175

Tableau 15. Exemple de fusion de deux HSP en sortie de la fonction *merge*..... 176

ABRÉVIATIONS

Dans la plupart des cas, pour la commodité du lecteur, les sigles ou abréviations de la littérature anglo-saxonne ont été conservés, les références citées dans ce document appartenant à celle-ci.

1D SDS-PAGE	Électrophorèse sur gel SDS-polyacrylamide à une dimension (<i>one dimensional SDS-polyacrylamide gel electrophoresis</i>)
2D PAGE	Électrophorèse bidimensionnelle en gel de polyacrylamide (<i>two-dimensional polyacrylamide gel electrophoresis</i>)
AAP	An(s) avant le présent
AD	Autosomique dominante
ADN	Acide désoxyribonucléique
ADNmt	Acide désoxyribonucléique mitochondrial
ADP	Adénosine 5'-diphosphate
AOAD	Atrophie optique autosomique dominante
AOADC	Atrophie optique autosomique dominante et cataracte
AOADS	Atrophie optique autosomique dominante et surdité
AOAR	Atrophie optique autosomique récessive
AR	Autosomique récessive
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messenger
ARNr	Acide ribonucléique ribosomique
ARNt	Acide ribonucléique de transfert
ATP	Adénosine 5'-triphosphate
ATP6, ATP8	Gènes codant deux sous-unités du complexe V (ATPase 6 et 8)
BLAST	<i>Basic Local Alignment Search Tool</i>
Brin H	Brin lourd (<i>heavy strand</i>)
Brin L	Brin léger (<i>light strand</i>)
CDS	Région codante (<i>coding sequence</i>)
CFNR	Couche des fibres nerveuses rétiniennes
CGR	Cellule ganglionnaire de la rétine
CO I, II, III	Gènes codant trois sous-unités du complexe IV (cytochrome c oxydase)

COX	Cytochrome <i>c</i> oxydase
CPEO	Ophthalmoplégie chronique progressive externe (<i>chronic progressive external ophthalmoplegia</i>)
CSB	<i>Conserved sequence box</i>
Cytb	Gène codant une sous-unité du complexe III (cytochrome <i>b</i>)
DDP1	<i>Deafness dystonia peptide 1</i>
e⁻	Électron
ESI-MS	Spectrométrie de masse à ionisation par électrospray (<i>electrospray ionization mass spectrometry</i>)
ETAS	<i>Extended termination associated sequences</i>
Fe/S	Fer-soufre
FISH	Hybridation <i>in situ</i> fluorescente (<i>fluorescence in situ hybridization</i>)
GED	<i>GTPase effector domain</i>
GFP	Protéine fluorescente verte (<i>green fluorescent protein</i>)
HS	Brin lourd (<i>heavy strand</i>)
HsP	Promoteur du brin lourd (<i>heavy strand promoter</i>)
HSP	Résultat BLAST élémentaire correspondant à un alignement (<i>high-scoring segment pair</i>)
IMM	Membrane interne mitochondriale (<i>inner mitochondrial membrane</i>)
IMS	Espace intermembranaire (<i>inter membrane space</i>)
IRM	Imagerie par résonance magnétique
kb	Kilobase
KSS	Syndrome de Kearns-Sayre (<i>Kearns–Sayre syndrome</i>)
LC-FTICR-MS	Chromatographie en phase liquide couplée à une spectrométrie de masse à transformée de Fourier (<i>liquid chromatography fourier transform-ion cyclotron resonance mass spectrometry</i>)
LCR	Liquide céphalorachidien
LR	Répétition longue (<i>long repeat</i>)
LS	Brin léger (<i>light strand</i>)
LSDB	Base de données spécifique d'un gène (<i>Locus Specific Database</i>)
LsP	Promoteur du brin léger (<i>light strand promoter</i>)
MDLC	Chromatographie liquide multidimensionnelle (<i>multidimensional liquid chromatography</i>)

MELAS	Myopathie mitochondriale, encéphalopathie, acidose lactique et tableaux neurologiques aigus ressemblant à des accidents ischémiques cérébraux (<i>mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes</i>)
MERRF	Épilepsie myoclonique avec fibres rouges déchiquetées (<i>myoclonic epilepsy with ragged red fibers</i>)
MET	Microscopie électronique en transmission
MLPA	<i>Multiplex ligation probe amplification</i>
n-LC-MS/MS	Nano chromatographie liquide couplée à une spectrométrie de masse en tandem (<i>Nanoscale liquid chromatography tandem mass spectrometry</i>)
NAD⁺	Forme oxydée du nicotinamide adénine dinucléotide
NADH	Forme réduite du nicotinamide adénine dinucléotide
ND1, 2, 3, 4, 4L, 5, 6	Gènes codant 7 sous-unités du complexe I (NADH déshydrogénase)
NMD	<i>Nonsense-mediated mRNA decay</i>
NOHL	Neuropathie optique héréditaire de Leber
OMM	Membrane externe mitochondriale (<i>outer mitochondrial membrane</i>)
OPA1	Atrophie optique de type 1 (<i>optic atrophy 1</i>)
OPA1	<i>Optic atrophy 1</i> (gène)
ORF	Cadre de lecture ouvert (<i>open reading frame</i> , ORF)
OXPHOS	Phosphorylation oxydative (<i>oxidative phosphorylation</i>)
pb	Paires de bases
PCR	Amplification en chaîne par polymérase (<i>polymerase chain reaction</i>)
PH-	Protéine sans homologue procaryote (<i>procaryote Homologue -</i>)
PH+	Protéine avec un homologue procaryote (<i>procaryote Homologue +</i>)
PIC	Pression intracrânienne
PIO	Pression intraoculaire
QU	Ubiquinone
ReMIX	<i>Reconfigurable Memory for Indexing</i>
RFLP	Polymorphisme de taille des fragments de restriction (<i>restriction fragment length polymorphisms</i>)
RLX	Récessive liée à l'X
ROS	Dérivés réactifs de l'oxygène (<i>reactive oxygen species</i>)
RRF	Fibres rouges déchiquetées (<i>ragged red fibers</i>)
SEP	Sclérose en plaques

SNC	Système nerveux central
SNP	Polymorphisme mononucléotidique (<i>single nucleotide polymorphism</i>)
SQL	<i>Simple Query Langage</i>
SR	Répétition courte (<i>short repeat</i>)
SR	Riche en sérine et arginine (<i>serine- and arginine-rich</i>)
SSCP	Polymorphisme de conformation des acides nucléiques simple brin (<i>single strand conformation polymorphism</i>)
TCO	Tomographie de cohérence optique
XLOA	Atrophie optique liée à l'X (<i>X-linked optic atrophy</i>)

CHAPITRE 1 : INTRODUCTION

1.1 LE GÉNOME MITOCHONDRIAL

Selon la théorie endosymbiotique avancée dans les années 1970, les mitochondries, organites cytoplasmiques dont la fonction principale est la phosphorylation oxydative (Figure 1), et les plastes proviendraient d'un évènement de symbiose entre une eubactérie et une cellule eucaryote primitive [1].

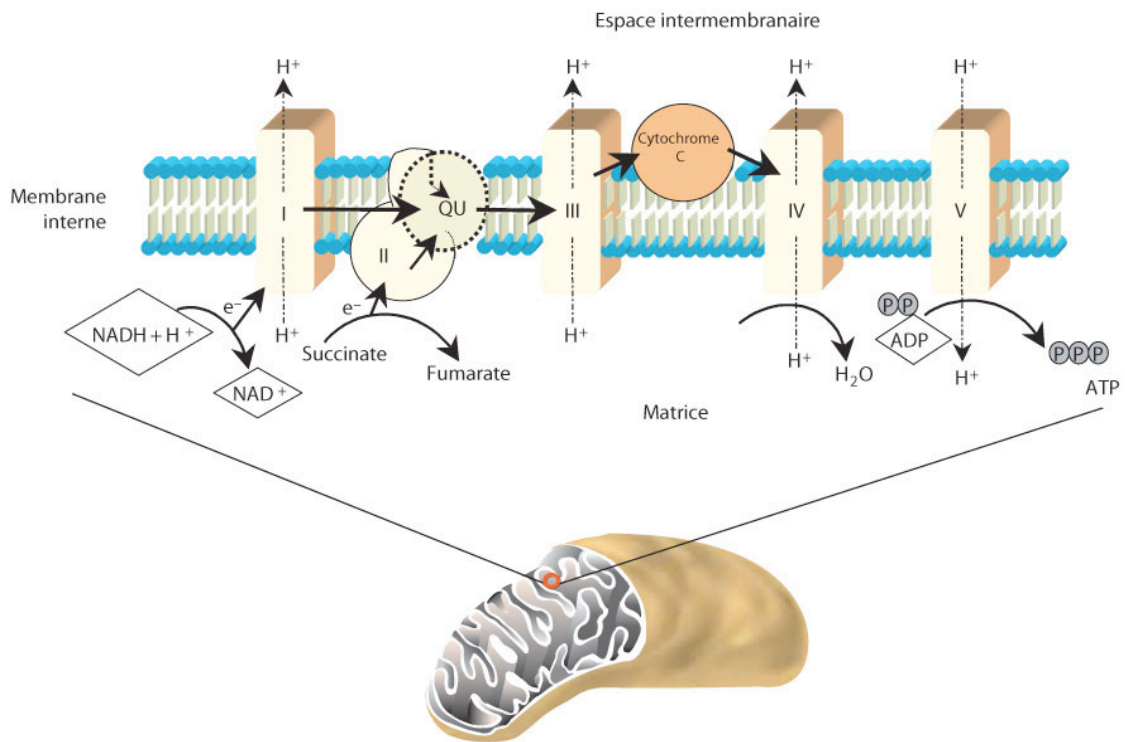


Figure 1. Représentation schématique de la mitochondrie avec les détails de la membrane interne et des complexes respiratoires (I–IV). Les mitochondries sont formées par deux membranes, une membrane externe ayant principalement des propriétés de perméabilité et une membrane interne où se produisent le transport d'électrons et la phosphorylation oxydative. L'espace interne, appelé matrice, contient des enzymes du métabolisme, l'acide désoxyribonucléique mitochondrial (ADNmt), et la machinerie moléculaire pour sa réplication et son expression. Le flux d'électrons (e^-) à travers les complexes respiratoires entraîne une diminution de l'énergie libre, qui est utilisée pour pomper les protons (H^+) de la matrice vers l'espace intermembranaire. L'énergie conservée dans ce gradient de protons est utilisée pour promouvoir la synthèse de l'adénosine 5'-triphosphate (ATP) par l'ATP synthase en inversant le flux de protons de l'espace intermembranaire vers la matrice. ADP : adénosine 5'-diphosphate ; NAD $^+$: forme oxydée du nicotinamide adénine dinucléotide ; NADH : forme réduite du nicotinamide adénine dinucléotide ; QU: ubiquinone.

Les récentes révisions de la théorie sur l'origine de la cellule eucaryote décrivent deux scénarios [2-7]. Selon l'hypothèse dite « de l'hydrogène », la cellule eucaryote dériverait d'un seul évènement au cours duquel une α -protéobactérie et une archéobactérie seraient entrées en symbiose, donnant lieu conjointement au noyau eucaryote (une chimère, car il contient du matériel génétique des deux organismes) et aux organites cytoplasmiques issus de l'eubactérie. La « théorie syncytiale », quant à elle, soutient le postulat d'un processus en deux étapes : dans une première phase, une δ -protéobactérie et une archéobactérie auraient contribué à l'origine du génome nucléaire tandis que, lors d'une symbiose postérieure, une α -protéobactérie aurait donné lieu au génome de l'organite.

Quel que soit le procédé, au cours de l'évolution, la majorité des gènes α -protéobactériens de la mitochondrie primitive a été transférée vers le noyau [8, 9]. Chez l'homme, ce processus est toujours en cours [10, 11] alors que les Archeozoa (eucaryotes sans mitochondries) ont complètement perdu leur génome mitochondrial. La persistance d'un génome mitochondrial requiert un nombre considérable de protéines, nécessaire à sa maintenance et son expression, ce qui représente un coût énergétique très important pour la cellule. La cause de la persistance de ce génome n'est pas connue, mais, selon l'hypothèse de D.C. Wallace [12], elle permettrait une évolution plus rapide des systèmes assurant la conversion énergétique face aux variations de l'environnement.

1.1.1 L'organisation des gènes

Les mitochondries modernes possèdent leur propre système génétique et de l'acide désoxyribonucléique (ADN) circulaire double brin. Dans tous les organismes, l'ADN mitochondrial (ADNmt) code pour quelques produits, essentiellement de deux types :

- les gènes codant plusieurs protéines qui, avec les produits codés par le génome nucléaire, forment les complexes de la phosphorylation oxydative ;
- les gènes impliqués dans le mécanisme de la synthèse des protéines, à savoir au moins deux types d'acide ribonucléique ribosomique (ARNr) et un ensemble plus ou moins complet de gènes codant des acides ribonucléiques de transfert (ARNt).

Tous les autres produits mitochondriaux sont codés par l'ADN nucléaire et transportés dans les organites grâce à un système complexe faisant intervenir des séquences d'adressage spécifiques et des complexes de transport transmembranaire. La taille de l'ADNmt est de 16 569 paires de bases (pb) chez l'homme [13, 14], celle-ci restant relativement constante chez les animaux (environ 13 000–19 000 pb) [15, 16] ; l'organisation de son contenu génétique est cependant très variable dans les différents embranchements taxinomiques.

L'organisation génétique de l'ADNmt humain, commune à tous les mammifères et aux vertébrés (à quelques exceptions près), est illustrée à la Figure 2. Le brun lourd (H ou HS

pour *heavy strand*) est très riche en informations génétiques ; il contient deux gènes codant des ARNr, 12 gènes codant des protéines — en particulier trois sous-unités du complexe IV (cytochrome *c* oxydase), une sous-unité du complexe III (cytochrome *b*), six sous-unités du complexe I (NADH, forme réduite du nicotinamide adénine dinucléotide) et deux sous-unités de l'adénosine triphosphatase (ATPase) — et 14 gènes codant des ARNt. Les gènes codant des protéines sont généralement ponctués par ces derniers. L'autre brin, le brin léger (L ou LS pour *light strand*), ne contient qu'un seul gène codant une protéine mitochondriale (ND6) et huit gènes codant des ARNt, dont quatre sont à proximité de l'origine de réplication du brin L. La principale région de régulation de l'ADNmt chez les mammifères est appelée boucle de déplacement (boucle-D ou *D-loop* ; environ 1 kb) parce que, lors de la réplication, le brin H naissant crée une structure triple brin avec le déplacement (D) de l'ancien brin. Cette région contient l'origine de réplication du brin H (O_H) et deux promoteurs, les promoteurs des brins H (HsP pour *heavy strand promoter*) et L (LsP pour *light strand promoter*), nécessaires à la transcription polycistronique de chacun des deux brins. Elle peut être divisée en trois domaines : un domaine central, bien conservé dans l'évolution, mais dont la fonction est encore inconnue, ainsi que deux domaines périphériques très variables, à droite et à gauche, appelés respectivement domaines CSB (pour *conserved sequence box*) et ETAS (pour *extended termination associated sequences*).

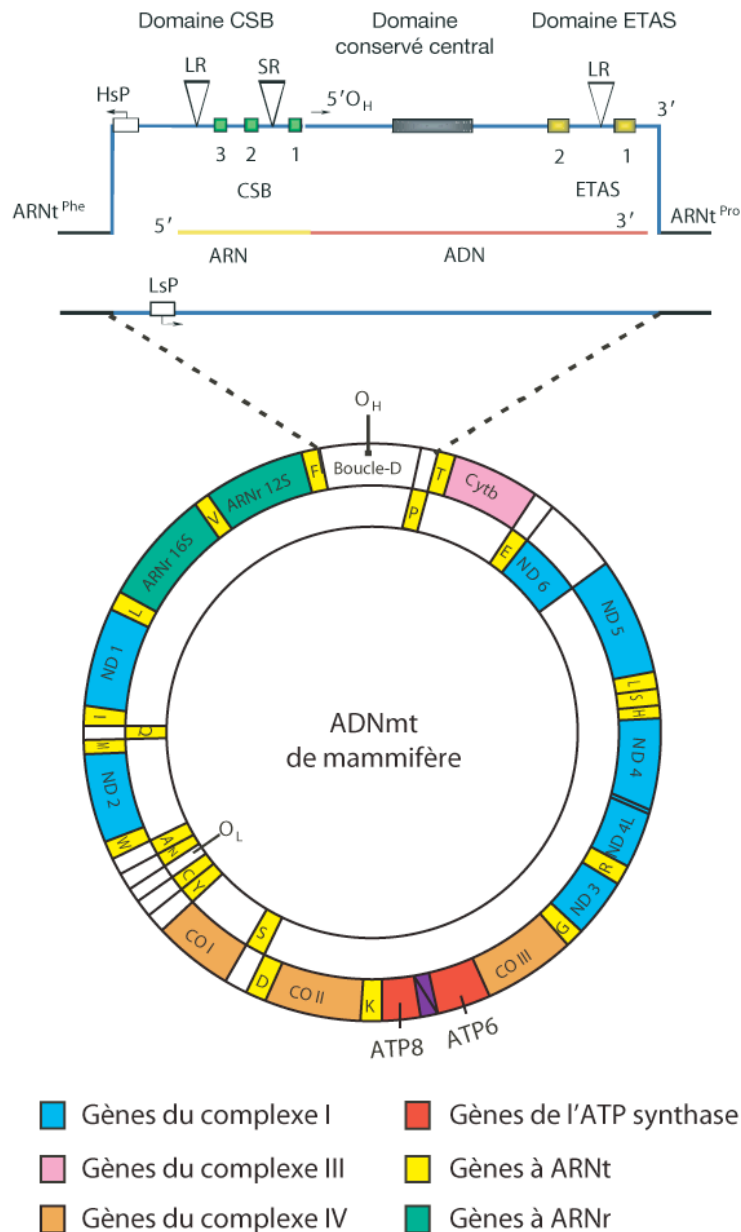


Figure 2. Cartographie de l'ADNmt de mammifère. Le cercle extérieur représente le brin lourd (H) et le cercle interne représente le brin léger (L), en fonction de leur sédimentation isopycnique dans un gradient de chlorure de césium. La structure à trois brins indique la boucle-D. O_H et O_L spécifient les origines de répliation. La répartition des gènes entre les brins et les principales régions régulatrices sont indiquées. Les gènes codant des protéines sont les suivants: ATP6, ATP8 : gènes codant deux sous-unités du complexe V (ATPase) ; CO I, II, III : gènes codant trois sous-unités du complexe IV (cytochrome c oxydase) ; Cytb : gène codant une sous-unité du complexe III (cytochrome b) ; ND1, 2, 3, 4, 4L, 5, 6 : gènes codant 7 sous-unités du complexe I (NADH déshydrogénase) ; ARNr 12S, 16S : ARNr des petites et grandes sous-unités ribosomales ; 22 ARN de transfert correspondant à leur acide aminé : A : alanine, C : cystéine, D : acide aspartique, E : acide glutamique, F : phénylalanine, G : glycine, H : histidine, I : isoleucine, K : lysine, L : leucine, M : méthionine, N : asparagine, P : proline, Q : glutamine, R : arginine, S : sérine, T : thréonine, V : valine, W : tryptophane, Y : tyrosine. Dans

l'agrandissement de la boucle-D : HsP : promoteur du brin H (*heavy strand promoter*) ; LsP : promoteur du brin L (*light strand promoter*) ; CSB et ETAS : domaines périphériques très variables (*conserved sequence box* et *extended termination associated sequences*) SR et LR: répétitions courte et longue (*short repeat* et *long repeat*) ; le domaine central conservé (barre noire), l'amorce ARN (gris) et la transition ARN/ADN sont indiqués.

1.1.2 Le code génétique mitochondrial

Le code génétique mitochondrial est légèrement modifié par rapport au code universel et il varie également d'un phylum à un autre. Les codons reconnaissant un acide aminé inhabituel sont résumés dans le Tableau 1. Le codon UGA est associé au tryptophane plutôt qu'à un codon-stop dans tous les codes génétiques des mitochondries, sauf dans les plantes. AGA et AGG codent pour l'arginine dans les champignons et les plantes, comme dans le code génétique nucléaire, mais ils codent pour la sérine chez les invertébrés (à l'exception de l'ADNmt de la drosophile, où AGG n'est pas présent). Chez les vertébrés, les codons AGA et AGG n'apparaissent à l'intérieur d'aucun gène codant des protéines mitochondriales, mais, dans certains cas, ils se retrouvent à leur extrémité 3'. Ce constat, ainsi que l'absence d'ARNt reconnaissant ces codons, appuie l'hypothèse que ce sont des codons de terminaison.

Tableau 1. Le code génétique mitochondrial chez les vertébrés.

Codon	Code mitochondrial	Code universel
UGA	Tryptophane	Stop
AUA	Méthionine	Isoleucine
AGA	Stop	Arginine
AGG	Stop	Arginine

Les gènes codant des protéines ont également des codons d'initiation de la traduction inhabituels. En plus d'AUG, tous les quatre codons AUN ont été trouvés à l'extrémité 5'-terminale des ARNm chez les mammifères et les échinodermes. Dans quelques cas, GUG semble être utilisé comme codon d'initiation dans l'ADNmt du rat, de la souris, du poulet, de l'oursin et du tournesol. Chez les nématodes, aucun des ARNm ne commence par l'AUG canonique ; AUA, AUU, UUG et GUU sont alors utilisés comme codons d'initiation. Les codons de terminaison mitochondriaux sont UAG et UAA. Chez les vertébrés, quand ils sont présents, AGA et AGG sont également terminateurs. Chez les métazoaires, certains transcrits de gènes s'achèvent avec les codons de terminaison incomplets U ou UA et les codons de terminaison complets sont créés par polyadénylation posttranscriptionnelle.

Le code génétique mitochondrial est lu par un ensemble d'ARNt plus petit que leurs homologues cytoplasmiques. La raison en est que, dans le système génétique mitochondrial, la reconnaissance entre codons et anticodons est de type « deux tiers », ce qui augmente l'instabilité.

1.1.3 L'héritage maternel

La transmission uniparentale non mendélienne de l'ADNmt est une des particularités du système génétique mitochondrial [17]. Chez les métazoaires, un seul des parents, à savoir la mère, transmet le génome à la descendance. En outre, l'héritage uniparental exclut les événements de recombinaison entre les patrimoines génétiques des parents, comme c'est le cas pour l'ADN nucléaire. Des recombinaisons entre l'ADNmt du père et de la mère ont été décrites chez les humains et les chimpanzés [18]. Toutefois, la majorité des données disponibles à l'heure actuelle confirme l'absence de la recombinaison des gènes dans l'ADNmt des animaux [19].

L'ADNmt étant strictement hérité maternellement, sans contribution paternelle ou recombinaison, cela signifie que les mutations sont l'unique source de variation entre les types d'ADNmt, et donc le nombre de mutations qui sépare deux types d'ADNmt est une mesure directe de la distance avec leur ancêtre commun. En principe, en l'absence de recombinaison, les méthodes classiques d'analyse phylogénétique peuvent être utilisées pour construire les arbres ou les généalogies illustrant les relations entre les types d'ADNmt. Dans la pratique, l'hypervariabilité (c'est-à-dire l'inégalité des taux de remaniement à différentes positions nucléiques) complique fortement cette procédure, ce que nous discuterons plus loin dans ce manuscrit.

À la lumière des règles fondamentales de la génétique des populations, le maintien de ce génome mitochondrial asexué dans les cellules eucaryotes est un paradoxe. En effet, selon le principe du cliquet de Muller [20, 21], l'ADNmt devrait avoir disparu au cours de l'évolution à cause de l'accumulation de mutations délétères qui ne peuvent pas être corrigées par la recombinaison. Ainsi, pour sa survie, l'ADNmt a besoin d'utiliser des stratégies spécifiques. Par exemple, au cours de la maturation des ovocytes chez les humains et les autres animaux, l'introduction d'un important « goulet d'étranglement » dans le nombre de mitochondries réduit, sélectionne puis amplifie le nombre de molécules d'ADNmt passant par la lignée germinale. Chez les animaux, un autre mécanisme qui contribue à la sélection des mitochondries passant à la génération suivante semble être l'atrésie des cellules germinales chez les femelles, dont le rôle supposé est la nécessité de transmettre à la génération suivante des mitochondries fonctionnelles et viables.

Ces caractéristiques de l'ADNmt, à savoir l'héritage maternel et l'absence de recombinaison, font que cette molécule est essentielle pour les études de l'évolution

moléculaire, car elle permet de remonter l'évolution d'une espèce jusqu'à l'ancêtre commun, en suivant une voie évolutive linéaire. Elle est également particulièrement adaptée pour des études phylogénétiques moléculaires, car tous les gènes mitochondriaux sont hérités comme un seul groupe de liaison et il n'y a pas de processus perturbant leur histoire évolutive, qui doit alors coïncider avec l'histoire évolutive des espèces.

La Figure 3 représente la voie de l'héritage maternel chez les humains et la fusion de toutes les lignées en un seul point qui représente le plus récent ancêtre commun, aussi appelé la « mère de tous ».

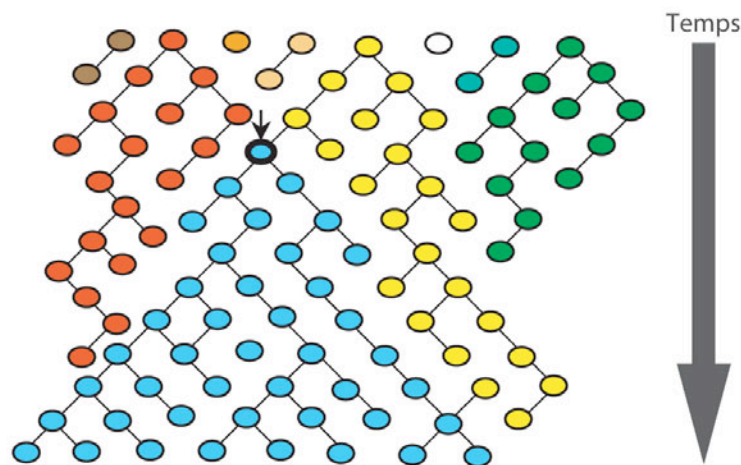


Figure 3. Transmission maternelle de l'ADNmt. La flèche noire indique l'ancêtre commun de toute la lignée qui en est dérivée.

1.1.4 Un grand nombre de copies

Un grand nombre de copies de l'ADNmt est présent dans les cellules humaines. Les cellules somatiques possèdent seulement deux copies de chaque gène nucléaire, mais en moyenne des milliers de copies d'ADNmt. Cette propriété et la localisation cytoplasmique de l'ADNmt facilitent son obtention pour l'analyse, et en font la molécule de choix pour les analyses d'ADN ancien et pour certaines applications dans le domaine des empreintes génétiques.

Cependant, cette propriété complique également la génétique des populations, car il y a plusieurs niveaux où les groupes de molécules d'ADNmt peuvent être définis : au sein d'une même mitochondrie, au sein d'une même cellule, dans un tissu particulier, au sein d'un individu et au sein d'un groupe d'individus (la définition traditionnelle d'une population). En outre, les copies multiples de l'ADNmt pour un même individu ne sont pas nécessairement toutes identiques ; l'existence de plusieurs types d'ADNmt chez un individu est connue sous le nom d'hétéroplasmie. La fréquence observée de l'hétéroplasmie chez l'homme dépend de la méthode utilisée pour la mesurer. Les premières estimations d'une fréquence relativement réduite de l'hétéroplasmie étaient dues aux méthodes peu sensibles utilisées ; plus

récemment, des études ont montré que près de 14 % de la population possède un deuxième type d'ADNmt, présent à une fréquence d'au moins 1 % [22] et, de fait, il est tout à fait probable que tout individu possède plusieurs types d'ADNmt (polyplasmie) parmi les trillions (10^{18}) et les trillions de génomes mitochondriaux de son corps.

1.1.5 L'évolution du génome mitochondrial

Depuis l'événement endosymbiotique qui aurait conduit à l'adoption de la mitochondrie originelle, l'impression générale est que le génome mitochondrial a connu une forte pression de sélection pour la réduction de sa taille [4]. Cette réduction de taille est due au transfert de matériel génétique du génome mitochondrial au génome nucléaire et à la perte des gènes redondants présents dans les deux génomes [8, 9]. Deux autres caractéristiques de l'ADNmt humain en découlent probablement. Tout d'abord, le code génétique de l'ADNmt humain a évolué, et ceci, de manière autonome vis-à-vis des autres organismes [23]. Le changement de ce code est probablement survenu lorsque le nombre de gènes restant dans le génome mitochondrial était assez petit pour qu'un changement dans le code génétique n'eût pas eu un effet délétère significatif sur la fonction des protéines. Ensuite, un certain nombre de copies partielles ou complètes du génome mitochondrial a été inséré dans le génome nucléaire humain tout au long de l'évolution [10, 11, 24] ; selon les études, entre 600 et 1 100 intégrations indépendantes ont été retrouvées réparties sur tous les chromosomes. Bien que ces inserts nucléaires possèdent des propriétés utiles pour des analyses évolutives [25], ils furent à l'origine de nombreuses conclusions erronées, des inserts nucléaires ayant été analysés par inadvertance à la place du génome mitochondrial.

1.1.6 Le rythme rapide de l'évolution

L'ADNmt évolue rapidement, 5 à 10 fois plus rapidement que la moyenne d'un segment d'ADN nucléaire [26]. Le taux d'évolution de l'ADNmt dépend de deux processus : le taux de nouvelles mutations qui se produisent et la probabilité qu'une mutation nouvellement apparue devienne fixe ; ces deux facteurs sont élevés pour l'ADNmt. Le taux de mutation intrinsèque des mutations ponctuelles de l'ADNmt chez l'homme, tel que mesuré par des études familiales, est au moins 10 fois plus élevé que pour l'ADN nucléaire [27]. Ceci pourrait résulter d'une réparation moins efficace des mutations de l'ADNmt et d'une plus grande exposition de l'ADNmt aux agents mutagènes tels que les dérivés réactifs de l'oxygène (ROS pour *reactive oxygen species*) qui sont présents à des concentrations plus élevées dans les mitochondries que dans le noyau. De plus, l'ADNmt se réplique constamment, y compris dans les cellules postmitotiques, ce qui l'expose davantage aux erreurs liées aux processus de réplication.

Ce taux d'évolution rapide implique que l'ADNmt est très utile pour les comparaisons d'espèces étroitement liées ou de populations de la même espèce, car dans ces deux cas un nombre suffisant de mutations aura eu lieu, même sur une courte période d'évolution. Cependant, la nature des mutations de l'ADNmt est inhabituelle puisque les transitions (substitutions entre deux bases puriques A↔G ou pyrimidiques C↔T) sont beaucoup plus fréquentes que les transversions (substitutions entre bases puriques et pyrimidiques) et, de plus, certains sites sont hypervariables, évoluant beaucoup plus rapidement que la moyenne [28]. Il résulte de ces deux aspects du processus mutationnel une haute fréquence de substitutions parallèles ou multiples à la même position nucléique, ce qui complique la reconstruction phylogénétique.

1.1.7 L'ADNmt et l'évolution de l'homme moderne

Partant des propriétés mutationnelles de l'ADNmt, les premières études de variations ont conduit à ce qui est devenu l'hypothèse de l'« Ève africaine » [29, 30] ; des travaux postérieurs ont largement soutenu et étendu cette hypothèse. Il y a deux aspects à cette hypothèse : d'abord, tous les types d'ADNmt dans les populations modernes remontent à un seul ancêtre qui vivait sans doute en Afrique ; deuxièmement, cet ancêtre vivait probablement il y a environ 150 000 ans.

L'étude de l'ADNmt montre donc que tous les types d'ADNmt humain existants ont une origine africaine commune récente. Ce résultat s'inscrit tout à fait avec les éléments de preuve obtenus avec des fossiles qui suggèrent l'origine de l'homme moderne en Afrique, il y a 100 000–150 000 ans [31, 32], mais ne soutiennent pas une autre théorie qui considère que les différences régionales entre les populations humaines ont une histoire ancienne qui remonte aux premières migrations des hominidés depuis l'Afrique, il y a 1,5–2 millions d'années [33]. Cependant, ces conclusions doivent être modérées par le fait que l'ADNmt représente un seul locus génétique, et l'histoire d'un seul locus peut être influencée par la sélection ou la dérive génétique. Néanmoins, l'accumulation d'informations sur les variations de l'ADN autosomique et du chromosome Y soutient l'hypothèse de l'origine africaine récente. Ainsi, l'ADNmt est une source importante de connaissances dans l'histoire évolutive des populations humaines.

1.2 LES POLYMORPHISMES DE L'ADN MITOCHONDRIAL

L'ADNmt humain possède un degré élevé de variabilité interindividuelle, chaque nucléotide de la molécule pouvant potentiellement être touché. Le génome mitochondrial se distingue de l'ADN nucléaire par un taux de mutations 5 à 10 fois plus élevé, une transmission exclusivement maternelle et une absence de recombinaison. Par conséquent, de nouveaux variants de l'ADNmt sont plus susceptibles d'apparaître, de sorte que les différentes lignées d'ADNmt d'une population évoluent indépendamment les unes des autres par l'accumulation de mutations séquentielles. Cela permet de reconstruire les relations phylogénétiques entre les différentes lignées d'ADNmt (ou entre les groupes de lignages partageant un ancêtre commun, caractérisés par des mutations communes) qui sont présentes dans une population, ou dans différentes populations. Ainsi, toutes ces caractéristiques évolutives ont conduit les généticiens à étudier les polymorphismes de l'ADNmt humain dès 1980.

La connaissance de la séquence complète de l'ADNmt en 1981 [13] a permis de renforcer les études. Initialement, des analyses de restriction étaient utilisées pour identifier les variants de l'ADNmt. Puis, le séquençage de deux portions hypervariables dans la région de contrôle de l'ADNmt (boucle-D) a également été réalisé. Plus récemment, les progrès techniques ont permis le séquençage complet de l'ADNmt pour la recherche de variations. Ainsi, de nombreuses variations apparemment neutres nommées polymorphismes ont été identifiées, permettant de comprendre la variabilité des populations humaines, de reconstruire les migrations anciennes et, en définitive, de faire la lumière sur l'origine et les migrations d'*Homo sapiens*.

Nous allons ici passer en revue les différents types de polymorphismes de l'ADNmt humain, comment ils sont liés les uns aux autres et ce qu'ils nous permettent de comprendre sur l'origine et les migrations d'*Homo sapiens* dans le monde. Dans la dernière partie, nous allons également voir les influences possibles des différentes lignées d'ADNmt sur les phénotypes complexes, tels que le vieillissement et la longévité.

1.2.1 Polymorphismes de l'ADNmt humain et évolution humaine

Avant que le séquençage d'un grand nombre de molécules d'ADNmt ne soit facilement réalisable, l'ADNmt était étudié principalement par trois approches : (1) l'analyse de restriction à faible résolution, (2) l'analyse de restriction à haute résolution et (3) le séquençage de deux régions hypervariables de la boucle-D.

1.2.1.1 L'analyse de restriction à basse résolution

Depuis 1983, la variabilité de l'ADNmt humain est étudiée par l'analyse des polymorphismes de taille des fragments de restriction (RFLP pour *restriction fragment length polymorphisms*)

reconnus par six enzymes (*MspI*, *BamHI*, *Avall*, *HaeIII*, *HpaI*, *HincII*) qui sont en mesure de cribler environ 2 % de l'ADNmt [34]. La génétique des populations humaines a bénéficié de cette approche durant plus d'une décennie. Plus de 300 haplotypes différents de l'ADNmt et des associations de variants au sein d'un même génome d'ADNmt ont été décrits dans le monde entier (voir <http://www.mitomap.org/cgi-bin/tbl3gen.pl>). Le constat le plus important est que l'ADNmt humain est, de loin, plus variable en Afrique que dans les autres populations humaines, ce qui suggère que l'homme serait originaire d'Afrique et aurait ensuite colonisé les autres continents, diffusant ainsi un sous-ensemble des variations accumulées dans la population d'origine.

Parmi tous les polymorphismes décrits par cette approche, certains semblaient spécifiques à un continent et furent extrêmement utiles pour distinguer les populations entre elles. Par exemple, le site *HpaI* à la position nucléique 3592 permet de distinguer les populations africaines des non-africaines. En fait, 90 % des Africains possèdent ce site, alors que la quasi-totalité des non-Africains ne l'a pas. Les polymorphismes trouvés avec une faible résolution d'analyse ont également permis de faire la première comparaison entre les polymorphismes spécifiques de l'ADNmt et ceux du chromosome Y. Le chromosome Y, comme l'ADNmt, ne se recombine pas, mais, contrairement à l'ADNmt, il est hérité seulement par le lignage paternel. Chromosome Y et ADNmt permettent donc d'explorer les versants paternel et maternel de l'histoire humaine. Un résultat intéressant de cette comparaison est que les femmes migrent plus que les hommes (par patrilocalité, coutume voulant qu'un nouveau couple habite sur les mêmes lieux que la famille du mari) [35].

1.2.1.2 L'analyse de restriction à haute résolution

L'analyse de restriction à haute résolution est réalisée avec 14 enzymes (*AluI*, *Avall*, *DdeI*, *HaeIII*, *HhaI*, *HinfI*, *HpaI*, *MspI*, *MboI*, *RsaI*, *TaqI*, *BamHI*, *HaeIII*, *HincII*). Ces enzymes possèdent un site de coupure fréquent d'environ 4 pb, permettant l'examen de près de 20 % de l'ADNmt humain. Cette technique a mis en évidence plusieurs centaines de polymorphismes dans des échantillons du monde entier (voir <http://www.mitomap.org/cgi-bin/tbl4gen.pl>). L'analyse de restriction à haute résolution est devenue très célèbre en 1987, lorsque Cann *et coll.* [29] ont analysé des échantillons provenant d'hommes de l'ensemble de la planète. Ils ont ainsi confirmé que la variabilité de l'ADNmt est plus élevée en Afrique que partout ailleurs dans le monde. L'analyse phylogénétique des haplotypes définis par les 14 enzymes a indiqué que l'origine de l'ADNmt de l'homme moderne pourrait être attribuée à l'Afrique. En supposant un taux d'évolution de l'ADNmt de 2–4 % par million d'années (calculé dans des études précédentes), Cann *et coll.* suggèrent que le dernier ancêtre commun de tous les humains modernes aurait vécu en Afrique il y a 150 000 à 500 000 ans.

Ce dernier ancêtre commun a été baptisé « Ève africaine ». Bien que certaines oppositions aient été exprimées à la suite de cette première étude, la plupart des conclusions ont été confirmées par plusieurs réanalyses. À partir de 1990, l'analyse à haute résolution a été largement utilisée en raison des développements de la technique d'amplification en chaîne par polymérase (*polymerase chain reaction*, PCR) alors que Cann et coll. ont utilisé la technique de Southern (*Southern blot*), une technique extrêmement génératrice d'erreurs du fait des profils donnés par les enzymes de restriction à sites de coupure fréquents. Une large gamme d'échantillons provenant de sujets du monde entier a été analysée par amplification PCR des fragments d'ADNmt suivie d'analyses de restriction. Les analyses phylogénétiques des haplotypes résultants ont permis d'identifier des groupes d'haplotypes partageant une origine commune. Ces groupes ont été nommés haplogroupes de l'ADNmt. Les haplogroupes sont spécifiques de chaque continent et représentent majoritairement la variabilité du continent. En effet, il a été constaté que 80 % des Africains sont d'haplogroupe L et 98 % des Européens sont d'un haplogroupe parmi neuf différents [36] (Tableau 2).

Tableau 2. Haplogroupes de l'ADNmt définis par analyse de restriction à haute résolution.

Continent	Haplogroupe	Variants de l'ADNmt
Afrique	L1	3592+ <i>HpaI</i> , 10806+ <i>HinfI</i>
Afrique	L2	3592+ <i>HpaI</i> , 16389+ <i>HinI</i> /16390– <i>Avall</i>
Europe	H	7025– <i>AluI</i> 39
Europe	T	13366+ <i>BamHI</i> , 15606+ <i>AluI</i>
Europe	U	12308+ <i>HinfI</i>
Europe	V	4577– <i>NlaIII</i>
Europe	W	8249+ <i>Avall</i> /8250– <i>HaeIII</i> ; 8944+ <i>RsaI</i>
Europe	X	1715– <i>DdeI</i>
Europe	I	10032+ <i>AluI</i> ; 4529– <i>HaeII</i> ; 8249+ <i>Avall</i> ; 16389+ <i>BamHI</i> / <i>MboI</i>
Europe	J	13704– <i>BstNI</i> ; 16065– <i>HinfI</i>
Europe	K	9052– <i>HaeII</i> / <i>HhaI</i>
Asie	A	663+ <i>HaeIII</i>
Asie	B	8271–8281 délétion de 9 pb ; 16517+ <i>HaeIII</i>
Asie	F	12406– <i>HpaI</i> / <i>HincII</i> ; 16517+ <i>HaeIII</i>
Asie	M	10394+ <i>DdeI</i> ; 10397+ <i>AluI</i>
Asie	M–C	13259–/13262+ <i>HincII</i> / <i>AluI</i>
Asie	M–D	5176– <i>AluI</i>

Amérique	A	663+ <i>HaeIII</i>
Amérique	A (Na-Dene) ^a	663+ <i>HaeIII</i> ; 16329– <i>RsaI</i>
Amérique	B	8271–8281 délétion de 9 pb ; 16517+ <i>HaeIII</i>
Amérique	M–C	13259–/13262+ <i>HincII/AluI</i>
Amérique	M–D	5176– <i>AluI</i>

Cette distribution des haplogroupes, spécifique aux continents, a suscité de nombreuses études sur l'origine des populations humaines, ainsi que sur l'histoire des migrations de l'homme à travers les continents. En fait, en étudiant l'évolution de chaque haplogroupe et leurs relations phylogénétiques, d'importantes conclusions ont été rendues possibles sur l'origine des êtres humains, les routes qu'ils ont suivies pour coloniser le monde et les événements qui ont conduit à la répartition actuelle de la population humaine. Ces études furent concevables grâce à l'analyse de restriction à haute résolution, puis les progrès technologiques, notamment le séquençage de la région hypervariable de la boucle-D et plus récemment de l'ensemble de l'ADNmt, ont permis d'en affiner les résultats.

1.2.1.3 Séquençage des régions hypervariables de la boucle-D

La seule séquence non codante de l'ADNmt est la boucle-D, dans laquelle deux régions extrêmement variables ont été identifiées : la région hypervariable I (entre les positions 16024 et 16383) et la région hypervariable II (entre les positions 57 et 372). De nombreux polymorphismes de ces deux régions ont été identifiés par séquençage dans un grand nombre d'échantillons (voir <http://www.mitomap.org/cgi-bin/tbl6gen.pl>). Il est rapidement devenu évident que la combinaison des données de séquence de la boucle-D avec les données obtenues par le biais d'analyses de restriction serait extrêmement utile pour clarifier les relations entre les haplogroupes. De plus, puisque la boucle-D évolue plus vite que le reste de la molécule, sa variabilité pourrait donner des informations concernant une courte durée.

Ainsi, l'analyse des séquences de la boucle-D de neuf haplogroupes européens a donné des indices importants sur l'histoire des Européens. Il a été établi que l'haplogroupe U fut présent en Europe durant environ 50 000 ans (cette date a été estimée en tenant compte de l'ensemble des variations de la boucle-D observées dans les molécules considérées comme U). Tous les autres haplogroupes seraient arrivés en Europe il y a 25 000 ans (avant la dernière glaciation) ou il y a quelques 10 000–12 000 années (immédiatement après la dernière glaciation) [37]. En outre, en étudiant la distribution des variants de la boucle-D sur

^a Na-Dene est un des trois grands groupes linguistiques des Amérindiens. Les locuteurs Na-Dene sont localisés en Amérique du Nord.

le territoire, il a été possible d'établir l'origine de certains haplogroupes. Par exemple, la variabilité des molécules appartenant à l'haplogroupe V est élevée dans la péninsule ibérique et diminue en Europe centrale et du nord. Ce constat a été interprété comme une preuve que l'haplogroupe V a pris naissance dans la péninsule ibérique (il y a 15 000 ans) et s'est ensuite déplacé vers l'est en apportant seulement une partie de la variabilité [38].

1.2.1.4 Le séquençage complet de l'ADNmt

Rendu possible par les performances atteintes par des séquenceurs d'ADN automatiques, le séquençage complet de l'ADNmt n'a commencé que récemment. Ainsi, la combinaison des données obtenues par cette approche avec les connaissances déjà accumulées permet de définir l'arbre phylogénétique de l'ADNmt humain (Figure 4). Celui-ci montre que le macrohaplogroupe M (qui est présent en Asie de l'Est, Inde et Éthiopie) possède une seule origine. Il est probablement originaire d'Afrique puis s'est propagé à l'Inde et l'Asie du Sud-Est par la première migration hors de l'Afrique d'*Homo sapiens* il y a environ 55 000–60 000 ans. En outre, la Figure 4 montre que la plupart des haplogroupes européens et nord-asiatiques ont une origine commune, comme en atteste le partage de la mutation 10873, qui définit le macrohaplogroupe N. L'haplogroupe N, comme l'haplogroupe M, est également originaire d'Afrique, mais a été apporté à l'Asie par le deuxième exode d'*Homo sapiens* venu d'Afrique, il y a environ 50 000 ans. Il convient de noter que les dates précédentes ont été estimées en tenant compte non seulement de la génétique, mais aussi des données paléoclimatiques [39].

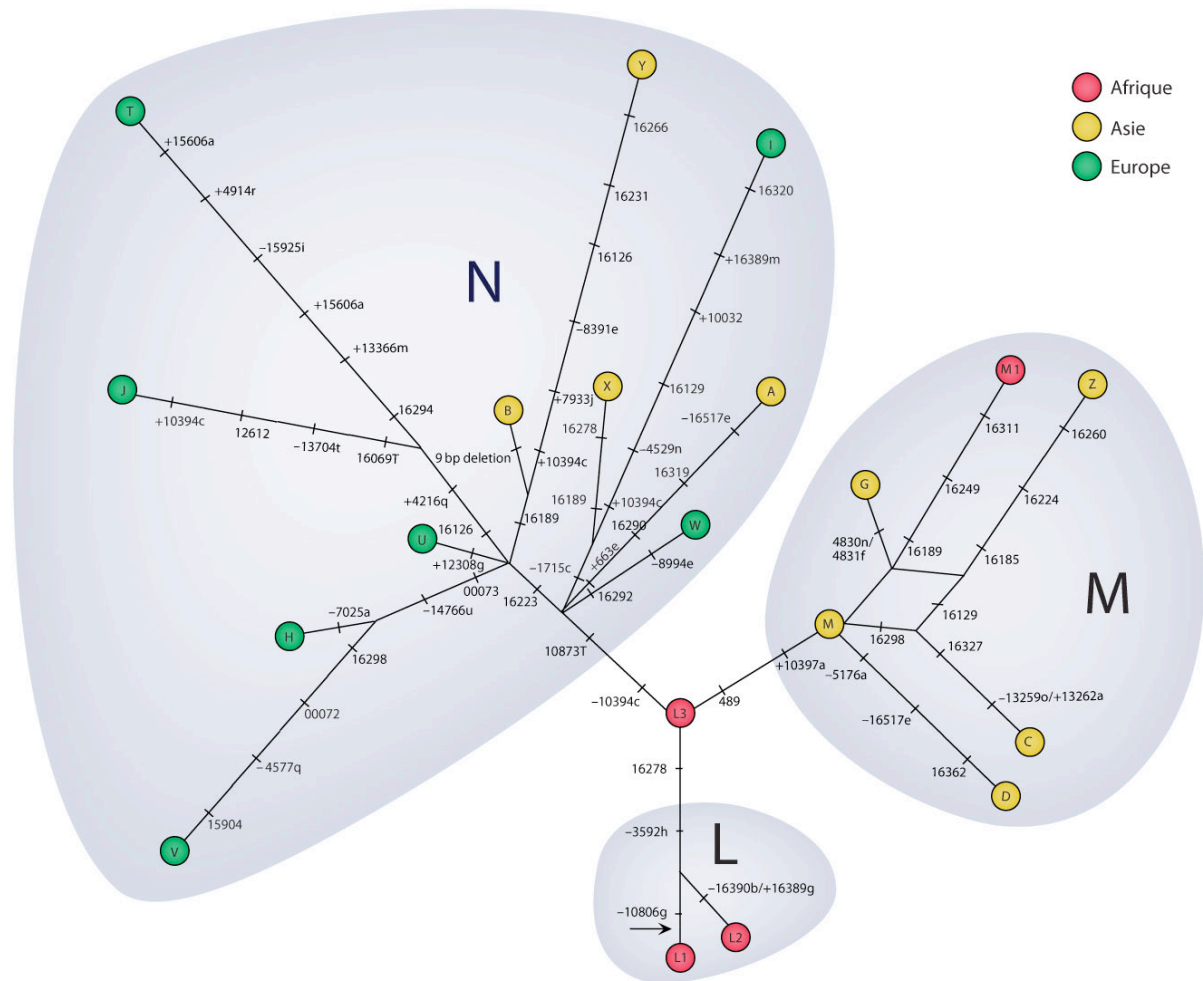


Figure 4. Arbre phylogénétique des haplogroupes de l'ADNmt. L'information obtenue avec l'ensemble des différentes analyses (RFLP, séquençage de la boucle-D et séquençage complet) a été fusionnée. La longueur des lignes droites reliant les haplogroupes entre eux est proportionnelle au nombre de mutations les différenciant. Les mutations sont indiquées soit avec la position où la mutation a lieu, soit avec la position du site de restriction qui est acquis (+) ou perdu (-). Dans ce dernier cas, une lettre indique l'enzyme de restriction qui reconnaît le polymorphisme. Le code à une lettre pour les enzymes de restriction est le suivant : a : *AluI* ; b : *Avall* ; c : *Ddel* ; e : *HaellI* ; f : *HhaI* ; g : *Hinfl* ; h : *HpaI* ; l : *MspI* ; j : *Mbol* ; k : *RsaI* ; l : *TaqI* ; m : *BamHI* ; n : *Haell* ; o : *HincII* ; q : *NlaIII* ; r : *BfaI* ; s : *Accl* ; t : *BstOI* ; u : *MseI*. Étant donné que les haplogroupes de l'ADNmt sont spécifiques d'un continent, le continent où l'haplogroupe est présent est indiqué. Les haplogroupes désignés comme asiatiques comprennent ceux qui figurent chez les Amérindiens et les Polynésiens. L'haplogroupe X désigné comme asiatique se retrouve également en Europe et chez les Américains d'origine.

Ainsi, l'analyse des variations de l'ADNmt a permis la reconstruction des migrations anciennes de l'espèce humaine (Figure 5). Cela a fourni la preuve que l'espèce humaine est apparue en Afrique il y a environ 130 000 à 170 000 ans. Elle a ensuite migré hors d'Afrique vers l'Asie il y a environ 60 000 à 70 000 ans et vers l'Europe il y a environ 40 000 à 50 000

ans, puis a migré d'Asie et peut-être d'Europe vers les Amériques il y a environ 26 000 à 34 000 ans [40, 41]. Bien que le calendrier de ces événements soit encore actuellement discuté, la contribution cruciale de l'étude des polymorphismes de l'ADNmt dans la reconstruction de l'histoire de l'humanité est largement reconnue.

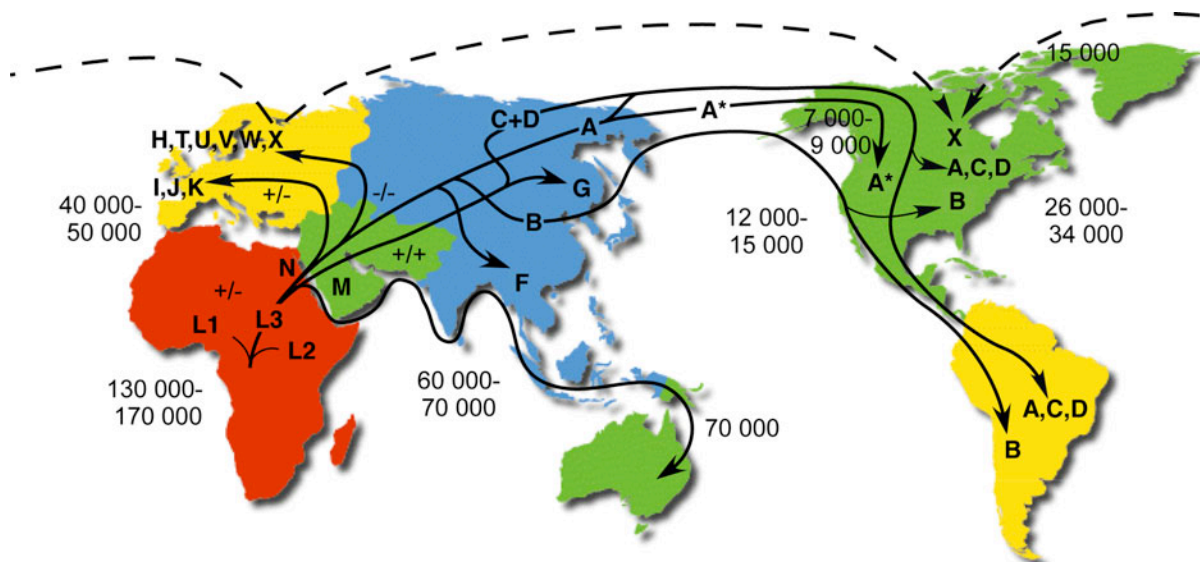


Figure 5. Migration et distribution géographique des haplogroupes mitochondriaux. La datation est exprimée en années avant le présent. Le taux de mutation est de 2,2–2,9 % par million d'années. +/-, +/+ et -/- : présence (+) ou absence (-) des sites de restriction *Ddel* 10394 / *AluI* 10397 ; * : présence du site de restriction *RsaI* 16329. D'après MITOMAP[42] (<http://www.mitomap.org/WorldMigrations.pdf>).

En outre, Pereira *et coll.* [43] ont analysé récemment tous les génomes mitochondriaux humains séquencés disponibles dans GenBank [44] (5 140 séquences à la fin du mois d'août 2008) afin de présenter une image globale de la diversité de l'ADNmt de la population mondiale. Pour effectuer cette tâche, les auteurs ont développé un outil informatique pour comparer et classer exhaustivement de grands ensembles de données génétiques. La diversité observée dans les 5 140 génomes mitochondriaux humains a été comparée à toutes les mutations synonymes possibles de la séquence de référence du génome mitochondrial humain. Cette approche a montré que les séquences des structures secondaires, ARNr et ARNt, sont fortement conservées, tandis que pour les gènes codant des protéines il y a un biais en faveur d'une moindre variation de la seconde position des codons. L'analyse au niveau des acides aminés a montré une tolérance des variations substituant les acides aminés V, I, A, S et T entre eux, définissant un groupe d'acides aminés liés par une transition (substitution entre deux bases puriques $A \leftrightarrow G$ ou pyrimidiques $C \leftrightarrow T$) unique qui possèdent des propriétés chimiques similaires.

1.2.2 Influence des polymorphismes de l'ADNmt

La variabilité de l'ADNmt chez l'homme a longtemps été considérée comme neutre et seulement utile pour la reconstruction de l'histoire des populations humaines. Toutefois, en 1988, deux maladies ont été associées à des mutations de l'ADNmt : la neuropathie optique héréditaire de Leber (NOHL ; MIM #535000) [45] et le syndrome de Kearns-Sayre (KSS pour *Kearns–Sayre syndrome* ; MIM #530000) [46]. À ce jour, plus de 200 mutations pathogènes ont pu être identifiées (voir <http://www.mitomap.org/>). De plus, l'analyse statistique des changements d'acides aminés causés par les polymorphismes de l'ADNmt a montré que les changements de nucléotides du génome mitochondrial subissent des pressions évolutives qui sont spécifiques des différentes régions de la molécule [47, 48]. Ainsi, l'étude de maladies mitochondriales où des mutations ponctuelles de l'ADNmt déclenchent des phénotypes cliniques complexes et la découverte de variants sélectionnés par l'évolution ont remis en cause l'hypothèse de la neutralité des variations de l'ADNmt.

Cette hypothèse a été définitivement abandonnée quand il a été démontré que les différentes lignées d'ADNmt qui sont présentes dans les populations humaines sont qualitativement différentes les unes des autres. Ruiz-Pesini *et coll.* [49] ont proposé que l'haplogroupe T par rapport à l'haplogroupe H (deux des neuf haplogroupes européens) afficherait une baisse significative de l'activité des complexes I et IV de la chaîne respiratoire. Des variations d'activité enzymatique agiraient sur la mobilité des spermatozoïdes. En conséquence, les haplogroupes H et T seraient respectivement plus fréquents chez les personnes non asthénozoospermes et asthénozoospermes. Par ailleurs, l'étude de cohortes d'individus japonais centenaires montre une fréquence de l'haplogroupe asiatique D significativement plus importante que des individus plus jeunes, de sexe identique et ayant la même origine ethnique et géographique, suggérant ainsi des effets spécifiques des haplogroupes sur les phénomènes de vieillissement [50]. Par la suite, des données similaires, résumées dans le Tableau 3, ont été obtenues dans des populations centenaires européennes, avec en particulier une surreprésentation des haplogroupes européens J et U(Uk) [51-54].

Tableau 3. Haplogroupes mitochondriaux et longévité.

Population	Haplogroupe mitochondrial	Référence
Japonaise	D	Tanaka <i>et coll.</i> , 1998 [50]
Française	K	Ivanova <i>et coll.</i> , 1998 [52]
Italienne	J	De Benedictis <i>et coll.</i> , 1999 [51]
Irlandaise	J / Uk	Ross <i>et coll.</i> , 2001 [54]
Finlandaise	J/U(Uk)	Niemi <i>et coll.</i> , 2003 [53]

De la même manière, plusieurs analyses phylogénétiques du génome mitochondrial, résumées dans le Tableau 4, indiquent des différences significatives chez des patients atteints de maladies neurodégénératives par rapport à des populations contrôles, avec un effet protecteur des haplogroupes J et U ainsi qu'un effet délétère de l'haplogroupe H dans ces pathologies.

Tableau 4. Haplogroupes mitochondriaux et maladies neurodégénératives

Population	Haplogroupe mitochondrial	Effet	Maladie neurodégénérative	Référence
Canadienne-française	T	Protecteur	Maladie d'Alzheimer	Chagnon <i>et coll.</i> , 1999 [55]
Ascendance européenne	J / U	Protecteur	Maladie de Parkinson	Van der Walt <i>et coll.</i> , 2003 [56]
Britannique	H	Susceptibilité	Démence à corps de Lewy	Chinnery <i>et coll.</i> , 2000 [57]

Dans tous les cas, toutes les études sur les différences qualitatives liées aux polymorphismes de l'ADNmt et les effets consécutifs de ces polymorphismes sur des phénotypes complexes doivent tenir compte de la synergie entre le génome nucléaire et le génome mitochondrial.

L'influence des haplotypes de l'ADNmt sur la respiration cellulaire n'a été démontrée qu'en 2006 par Moreno-Loshuertos *et coll.* [58]. En utilisant des lignées de cellules comportant quatre haplotypes différents d'ADNmt murin dans un contexte nucléaire identique, les auteurs ont montré que le même niveau de respiration des lignées cellulaires n'est qu'apparent et serait une conséquence des mécanismes compensatoires déclenchés par une production différente de ROS. Certains types de patrimoines génétiques mitochondriaux sont associés à une capacité de respiration réduite. Ces différences fonctionnelles s'expliqueraient par la présence d'un polymorphisme dans l'ARNt de l'arginine. Les lignées cellulaires mutées présenteraient une production de ROS plus élevée et une quantité accrue d'ADNmt. Le mécanisme moléculaire précis n'est pas connu, mais ce

polymorphisme influencerait la fidélité de la traduction plutôt que la synthèse des protéines. Si de faibles quantités de polypeptides présentant des mutations faux-sens passent le système de contrôle de la qualité des protéines et sont intégrées dans des complexes de la chaîne respiratoire, un taux plus élevé de ROS serait produit, ce qui déclencherait les conséquences phénotypiques observées. Le complexe I serait alors plus susceptible d'être affecté que n'importe quel autre complexe, car il contient 7 des 13 polypeptides codés par l'ADNmt. De nombreux exemples de mutations des ARNt sont associés à des maladies humaines pour lesquelles un phénotype biochimique est difficile à évaluer [59, 60]. L'existence de tels mécanismes compensatoires inductibles et la disponibilité des drogues pour les inhiber pourraient ainsi contribuer à leur compréhension.

1.3 LE PROTÉOME MITOCHONDRIAL

Les mitochondries sont le siège d'un grand nombre de métabolismes dans les cellules eucaryotes. Le plus important est sans doute la phosphorylation oxydative, dans laquelle cinq complexes majeurs coopèrent pour transformer l'énergie apportée par les nutriments en énergie facilement utilisable par la cellule, sous forme d'adénosine 5'-triphosphate (ATP). L'ATP comporte en effet deux liaisons pyrophosphates dont l'hydrolyse est fortement exergonique. Le métabolisme mitochondrial aérobie fournit la majorité de cet ATP cellulaire tandis que dans les conditions anaérobies, la glycolyse cytoplasmique est le principal pourvoyeur d'ATP. Les autres réactions qui se produisent dans les mitochondries sont le cycle de Krebs, la biosynthèse de l'hème, la β -oxydation des acides gras et le métabolisme de certains acides aminés, ainsi que les voies de formation et d'exportation des centres fer-souffre (Fe/S), plus récemment découvertes (Figure 6).

Toutes les protéines catalysant ces réactions sont localisées dans un organite qui est délimité par deux membranes le séparant du reste de la cellule, même s'il n'est pas fonctionnellement isolé. Les interactions avec le cytosol et le réticulum endoplasmique nécessitent un grand nombre d'échangeurs, de transporteurs et de canaux dans les deux membranes. Par exemple, chez l'homme, les transporteurs de l'ADP-ATP transfèrent près de 50 à 60 kg d'ATP par jour de part et d'autre de la membrane mitochondriale interne.

La majorité des protéines mitochondriales est codée par le génome nucléaire puis synthétisée dans le cytoplasme et importée posttraductionnellement dans les mitochondries ; le génome mitochondrial code quant à lui seulement quelques protéines (par exemple, huit chez *Saccharomyces cerevisiae*, 13 chez *Homo sapiens*, 25 chez *Arabidopsis thaliana* et 64 chez *Reclinomonas americana* ; voir <http://gobase.bcm.umontreal.ca/>). Un nombre considérable de protéines mitochondriales est impliqué dans des processus d'importation et d'assemblage des protéines codées par le noyau. Un groupe de protéines chaperonnes veille au bon repliement *de novo* des protéines mitochondriales et participe aux décisions de repliement ou de dégradation. L'élimination protéolytique des protéines mal repliées et la régulation de la rotation des protéines se font par l'intermédiaire d'une série de protéases associées aux différents sous-compartiments des mitochondries. Certaines protéines sont impliquées dans la fusion et la fission de ces organites (l'une d'elles, Opa1, sera détaillée plus loin dans ce document) qui forment des réseaux interconnectés hautement dynamiques [61]. Ces protéines sont généralement indispensables à la bonne transmission de l'ADNmt et certaines de leurs mutations sont associées à des maladies humaines (voir Dysfonctionnements de la mitochondrie dans les pathologies humaines, p. 39).

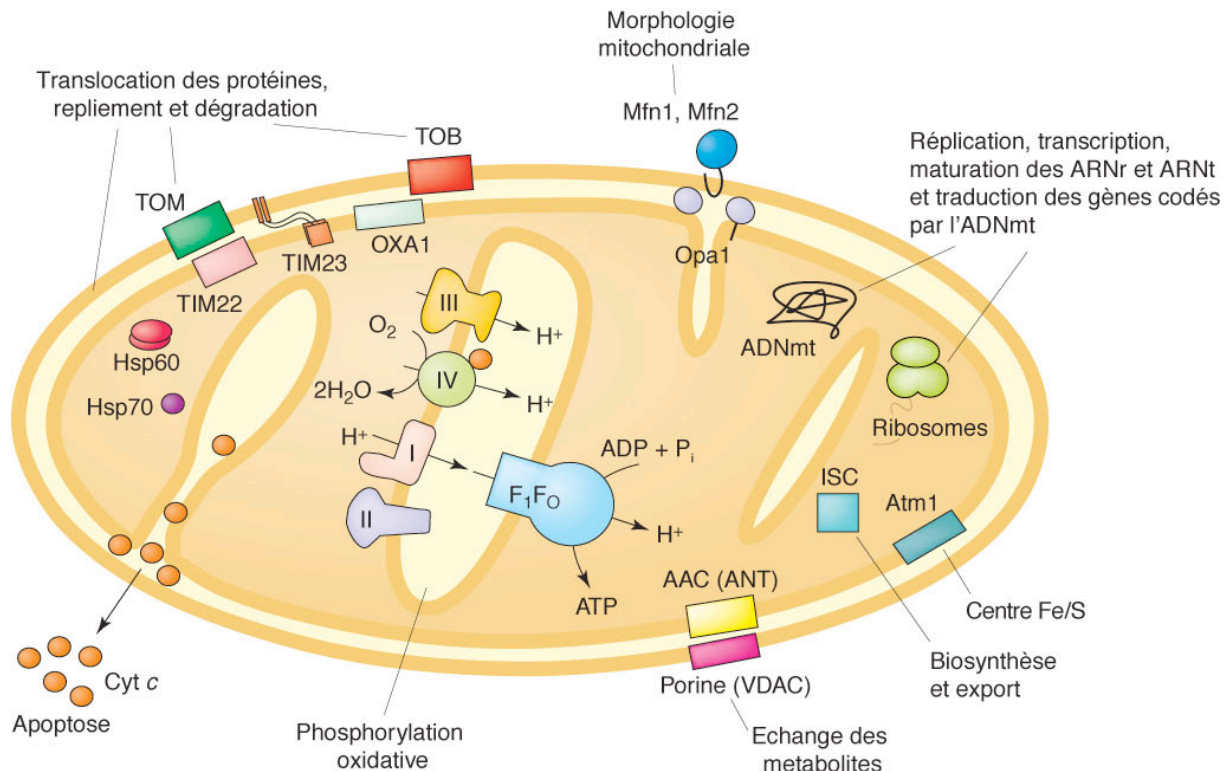


Figure 6. Vue d'ensemble des processus cellulaires dans les mitochondries humaines. Schéma de section d'une mitochondrie montrant une sélection de protéines ou de complexes protéiques impliqués dans : la translocation des protéines [complexe des translocases de la membrane externe (*translocase of outer membrane*, TOM), protéines dites β-barrel distribuées dans la membrane externe (*topogenesis of mitochondrial outer-membrane β-barrel proteins*, TOB), trois complexes de translocases de la membrane interne TIM22, TIM23 et OXA1], le repliement des protéines (Hsp60, Hsp70) et la dégradation ; la morphologie mitochondriale (Mfn1, Mfn2 et Opa1), la réplication, la transcription, la maturation des ARNr et ARNt, et la traduction (ribosomes) des gènes codés par l'ADNmt, la biosynthèse des centres Fe/S [complexe d'assemblage des centres fer-souffre (*iron-sulfur cluster assembly complex*, ISC)] et l'exportation (Atm1) ; l'échange des métabolites [transporteurs de l'ADP et l'ATP (*ADP-ATP carrier*, AAC) ou échangeurs de nucléotides adényliques (*adénine nucléotide translocator*, ANT), porines ou échangeurs d'anions en fonction du potentiel (*voltage dependant anion channel*, VDAC)] ; la phosphorylation oxydative (complexes I, II, III, IV, et F₁F₀-ATP-synthase) ; et l'apoptose (cytochrome c). La synthèse d'ATP dans la matrice mitochondriale est indiquée à l'aide du gradient de protons à travers la membrane interne. Les processus métaboliques de base tels que le cycle de Krebs, le métabolisme de certains acides aminés et lipides, et la biosynthèse de l'hème sont omis. Les invaginations de la membrane interne, appelées crêtes, sont représentées comme elles apparaissent souvent dans les représentations de sections de mitochondries en microscopie électronique. La partie de la crête où commence l'invagination, appelée jonction de la crête, est visible uniquement lorsque celle-ci se trouve dans le même plan que le plan de section. Si ce n'est pas le cas, elle conduit à un aspect isolé de la

membrane de la crête, même si la continuité sur l'ensemble de la membrane interne est maintenue.

En plus de ces fonctions souvent complexes, les mitochondries jouent un rôle essentiel dans l'apoptose. La libération par la mitochondrie du cytochrome c, de facteurs inducteurs de l'apoptose et de plusieurs autres protéines favorise la mort cellulaire programmée [62]. En outre, le vieillissement est fortement corrélé avec l'altération de la fonction mitochondriale. Par exemple, l'accumulation de mutations de l'ADNmt conduit à un vieillissement prématuré dans un modèle murin invalidant l'activité exonucléase de l'ADN polymérase γ mitochondriale [63].

Pour mieux comprendre les réseaux fonctionnels des protéines mitochondriales, celles-ci doivent être connues dans leur ensemble. Plusieurs études réalisées ces dernières années ont abordé cette question sous différents angles. Dans les sections suivantes, nous présentons un aperçu de ces études et de leurs différentes approches. L'accent sera mis sur les mitochondries de *Saccharomyces cerevisiae* (levure de boulanger), grâce à laquelle la plupart des grands ensembles de données ont été obtenus. En effet, cet organisme s'avère être un excellent modèle d'étude de la mitochondrie, ce qui n'est pas surprenant compte tenu de la conservation des fonctions fondamentales de cet organelle avec de nombreux organismes. Les progrès récents dans la détermination des protéomes mitochondriaux des mammifères et des plantes sont également discutés. En outre, nous soulignons l'importance de l'intégration de la bio-informatique, des approches génétiques et biochimiques, ainsi que de l'analyse fonctionnelle détaillée de plusieurs machines moléculaires mitochondriales.

1.3.1 Les études protéomiques

L'approche la plus directe pour obtenir le protéome mitochondrial complet est d'isoler les mitochondries hautement purifiées et de déterminer ce sous-ensemble de protéines mitochondriales par spectrométrie de masse. Plusieurs groupes ont purifié des mitochondries de levures, de souris et d'homme par centrifugation différentielle classique, gradient d'ultracentrifugation ou électrophorèse *free-flow* (*free-flow electrophoresis*, FFE) [64-69]. Des études similaires sur le protéome mitochondrial de plantes ont été réalisées [70-75]. En général, les mitochondries sont solubilisées et les protéines sont séparées de diverses façons. L'électrophorèse bidimensionnelle en gel de polyacrylamide (2D PAGE, pour *two-dimensional polyacrylamide gel electrophoresis*), qui consiste en une séparation basée sur la charge suivie d'une séparation basée sur la masse, est une méthode ayant nombreux avantages, car elle offre la possibilité, par exemple, de comparer un grand nombre de protéines sous l'influence de différentes conditions de croissance ou en réponse

à l'administration de médicament. Cela permet d'identifier rapidement les protéines dont les niveaux d'expression sont modifiés.

Ohlmeier *et coll.* [65] ont comparé le protéome de la levure de boulanger en croissance fermentaire vs oxydative. Dans l'ensemble, les différences étaient assez mineures, seulement 18 des 252 protéines identifiées sont devenues significativement plus abondantes. Toutefois, d'autres études utilisant l'électrophorèse bidimensionnelle en gel de polyacrylamide (2D PAGE pour *two-dimensional polyacrylamide gel electrophoresis*) ont mis en évidence un inconvénient majeur de cette méthode : l'ensemble des protéines identifiées est largement orienté vers les protéines abondantes et faciles à solubiliser [76].

Une étude vaste et approfondie réalisée par Sickmann *et col.* [68] propose une façon de surmonter ce problème. Des mitochondries hautement purifiées ont été préparées à partir de la levure de boulanger et différentes approches protéomiques ont été appliquées. En plus du 2D PAGE, ces approches incluent une digestion avec différentes protéases suivie d'une chromatographie liquide multidimensionnelle (*multidimensional liquid chromatography*, MDLC) puis d'une spectrométrie de masse à ionisation par électrospray (*electrospray ionization mass spectrometry*, ESI-MS), ou bien une électrophorèse sur gel SDS-polyacrylamide à une dimension (1D SDS-PAGE pour *one dimensional SDS-polyacrylamide gel electrophoresis*) suivie d'une nano chromatographie liquide couplée à une spectrométrie de masse en tandem (*nanoscale liquid chromatography tandem mass spectrometry*, n-LC-MS/MS). Au total, plus de 20 millions de spectres ont été enregistrés, ce qui donne une liste de 749 protéines mitochondriales non redondantes. Dans cette approche, la quasi-saturation a été atteinte, car, dans les spectres analysés à la fin, très peu de nouvelles protéines ont été détectées par rapport aux spectres analysés au début [68]. Ces protéines mitochondriales prédites couvrent environ 92 % des protéines mitochondriales déjà connues figurant dans la base de données MITOP [77]. Cette sensibilité (Tableau 5) pourrait cependant être inférieure à 92 %. En effet, MITOP surestime la véritable couverture du protéome mitochondrial, car cette base de données est elle-même orientée vers les protéines les plus solubles et abondantes ainsi que vers certaines catégories fonctionnelles telles que les complexes de la chaîne respiratoire. Actuellement, la base de données MITOP2 (<http://www.mitop2.de>) apparaît comme le meilleur ensemble de référence disponible (Tableau 5) [67, 77, 78]. Celle-ci représente une liste mise à jour des protéines qui ont été clairement localisées à la mitochondrie dans des expériences avec des protéines individualisées. Cette référence contient 477 protéines, dont 76 % ont été identifiés par Sickmann *et coll.* [68] (Tableau 6). Il s'agit de la sensibilité la plus élevée à ce jour pour une étude visant à évaluer le protéome mitochondrial complet d'un organisme unique (Tableau 6). En utilisant une combinaison d'approches protéomiques, Sickmann *et coll.* [68] ont minimisé les biais et les problèmes associés aux différentes méthodes prises

individuellement. Par exemple, l'analyse 2D PAGE tend à surreprésenter les protéines hydrophobes telles que les protéines membranaires périphériques ou intégrées [76]. Cela a été, au moins partiellement, surmonté par la combinaison de 1D SDS-PAGE et n-LC-MS/MS. Un autre problème souvent rencontré est que certaines protéines (par exemple, les petites protéines ou les protéines dont la composition en acides aminés est inhabituelle) ne donnent pas de fragments peptidiques de taille appropriée en raison de sites de clivage incorrectement espacés. Par conséquent, Sickmann *et coll.* ont utilisé quatre protéases différentes. Ainsi, ces différentes approches ont finalement conduit à la grande spécificité (48,5 % ; Tableau 5) de cette étude.

Tableau 5. Termes et définitions.

Sensibilité

La sensibilité d'une méthode pour déterminer le protéome mitochondrial est la mesure de la probabilité de détecter une protéine mitochondriale par cette approche, mais ne dit pas si un candidat est vraiment une protéine mitochondriale. Il s'agit d'une mesure de la couverture d'une approche, car celle-ci donne le pourcentage de protéines mitochondriales connues qui ont été identifiées par cette approche. La valeur absolue de la sensibilité dépend de l'ensemble de référence utilisé et ne doit donc être considérée que comme une estimation approximative. Toutefois, lorsque le même ensemble de référence est toujours utilisé, il est possible de comparer les différentes approches entre elles (Tableau 6).

Spécificité

La spécificité d'une approche pour déterminer le protéome mitochondrial est la mesure de la précision d'une approche donnée. Celle-ci peut être estimée à partir du pourcentage de protéines mitochondriales prédites par cette approche qui sont présentes dans l'ensemble de référence. Cette valeur est également tributaire de la qualité de l'ensemble de référence et ne fournit qu'une estimation approximative de la précision de l'approche. Pour un ensemble de référence donné, différentes approches peuvent être comparées les unes avec les autres.

Ensemble de référence

L'ensemble de référence donné par les administrateurs de la base de données MITOP2 pour *Saccharomyces cerevisiae* est basé sur des études de gènes isolés et ne contient aucune information provenant des approches comparées dans le Tableau 6 [67]. Ainsi, aucun biais n'est généré en faveur des approches à haut débit discutées. En raison du caractère incomplet de l'ensemble de référence lui-même, les calculs des spécificités et

des sensibilités ne sont que des estimations.

Sensibilité vs spécificité

Une approche qui est uniquement capable de prédire un nombre limité de protéines aura donc toujours une faible sensibilité (étant donné que l'ensemble de référence est beaucoup plus important). Toutefois, lorsque beaucoup de ces quelques protéines prédites sont présentes dans l'ensemble de référence (ce qui signifie qu'il y a probablement quelques faux positifs), la spécificité est élevée. Inversement, une approche qui prévoit que la quasi-totalité des protéines connues dans une cellule est localisée dans la mitochondrie aura une grande sensibilité, mais est plus susceptible également d'avoir une faible spécificité en raison du grand nombre de faux positifs.

Tableau 6. Comparaison quantitative de différentes approches pour déterminer le protéome mitochondrial chez la levure, l'homme et la souris^a.

Espèce	Approche	Nombre de protéines mitochondriales prédites ^b	Sensibilité (%) ^b	Spécificité (%) ^b
<i>Saccharomyces cerevisiae</i>				
Protéomique	Mitochondries hautement purifiées analysées par 2D PAGE, 1D SDS-PAGE et n-LC-MS/MS [68]	749	76,1	48,5
	Mitochondries hautement purifiées (par gradient de densité et électrophorèse <i>free-flow</i>) analysées par LC-MS/MS et LC-FTICR-MS [67]	546	53,7	46,9
	Protéines présentes dans les deux études [67, 68]	377	49,9	63,1
	Mitochondries hautement purifiées analysées par SDS-PAGE et LC-MS/MS [66]	177	28,3	76,3
	Mitochondries hautement purifiées analysées par 2D PAGE (croissance fermentative vs non fermentative) [65]	252	30,4	57,5

^a Abréviations : 1D SDS-PAGE, électrophorèse sur gel SDS-polyacrylamide à une dimension (*one dimensional SDS-polyacrylamide gel electrophoresis*) ; LC-FTICR-MS : chromatographie en phase liquide couplée à une spectrométrie de masse à transformée de Fourier (*liquid chromatography fourier transform-ion cyclotron resonance mass spectrometry*) ; n-LC-MS/MS : nano chromatographie liquide couplée à une spectrométrie de masse en tandem (*Nanoscale liquid chromatography tandem mass spectrometry*).

^b La sensibilité est la fraction de l'ensemble de référence couverte par l'ensemble des données mitochondriales prédites ; la spécificité est la fraction des protéines mitochondriales prédites confirmée par l'ensemble de référence (voir Tableau 5, p. 25). Les ensembles de données des études de *Saccharomyces cerevisiae* et *Homo sapiens*, et les ensembles de référence mitochondriaux contenant respectivement 477 et 656 protéines ont été récupérés à partir de la base de données MITOP2 (<http://www.mitop2.de>).

Espèce	Approche	Nombre de protéines mitochondriales prédites ^b	Sensibilité (%) ^b	Spécificité (%) ^b
Génétique	Étude à grande échelle de suppression de phénotype pour des souches ayant un déficit respiratoire (<i>petites</i>) [79]	381	39,2	49,1
	Étude à grande échelle de suppression de phénotype pour une croissance lente sur source de carbone non fermentescible [80]	466	47,6	48,7
	Analyse par immunofluorescence de protéines marquées par un déterminant antigénique [81]	364	46,5	61,0
	Analyse par fluorescence de protéines marquées par une GFP C-terminale [82]	527	69,2	62,6
Bio-informatique	Homologie avec <i>Rickettsia</i> (base de données MITOP2 [67, 78])	931	42,8	21,9
	Prédiction du signal d'adressage N-terminal avec un score MITOPROT >0.8 [83]	789	54,7	33,1
	Prédiction du signal d'adressage N-terminal avec un score MITOPROT >0.9 [83]	574	46,3	38,5
	Prédiction du signal d'adressage N-terminal par PSORT [84]	981	52,8	25,7
	Prédiction du signal d'adressage N-terminal par Predotar [85]	397	43,4	52,1
	Prédiction de la localisation subcellulaire par l'algorithme de Bayes [86]	500	43,4	41,4
Transcriptomique	Synexpression après surexpression de Hap4 [87]	514	45,1	41,8
	Synexpression après changement diauxique [88]	416	15,9	18,3
Intégration de la protéomique, transcriptomique, génétique et bio-informatique	Analyse intégrative de 22 jeux de données et un score MITOP2 >95 [67]	557	81,1	69,5
	Jeu de données d'interactions protéine-protéine à haut degré de confiance [89] combiné avec les protéines mitochondriales connues [67, 78]	154	17,4	53,9
	Jeu de données d'interactions protéine-protéine à degré de confiance moyen [89] combiné avec les protéines mitochondriales connues [67, 78]	815	49,7	29,1

Espèce	Approche	Nombre de protéines mitochondriales prédites ^b	Sensibilité (%) ^b	Spécificité (%) ^b
<i>Homo sapiens et Mus musculus</i>				
Protéomique	Mitochondries hautement purifiées de cœur humain analysées par 1D SDS-PAGE et LC-MS/MS [69]	544	32,9	39,7
	Mitochondries hautement purifiées de cœur de souris analysées par LC-MS/MS [64]	181	32,5	71,9
	Mitochondries hautement purifiées de cerveau de souris analysées par LC-MS/MS [64]	190	31,5	67,9
	Mitochondries hautement purifiées de foie de souris analysées par LC-MS/MS [64]	279	42,8	62,7
	Mitochondries hautement purifiées de rein de souris analysées par LC-MS/MS [64]	265	37,9	58,5
	Mitochondries hautement purifiées de l'ensemble des quatre tissus (cœur, cerveau, foie et reins de souris) analysés par LC-MS/MS [64]	409	53,1	53,1

Dans une étude récente de protéomique sur les mitochondries de *S. cerevisiae*, des cellules ont été cultivées sur des sources de carbone fermentescibles et non-fermentescibles [67]. Les mitochondries ont ensuite été purifiées par gradient de centrifugation classique ou par FFE. La matrice et les fractions associées à la membrane ont été produites et analysées séparément par LC-MS/MS ou par chromatographie en phase liquide couplée à une spectrométrie de masse à transformée de Fourier (*liquid chromatography fourier transform-ion cyclotron resonance mass spectrometry*, LC-FTICR-MS). Ainsi, 28 jeux de données ont été produits, conduisant à l'identification de 546 protéines non redondantes. Curieusement, cette étude a révélé de nombreux (209) candidats mitochondriaux qui n'avaient pas été découverts dans l'étude de Sickmann *et coll.* [68]. Inversement, 412 protéines prédites par cette dernière étude n'ont pas été détectées par Prokisch *et coll.* [67]. Ces auteurs suggèrent que de nombreuses protéines, mais pas toutes, détectées par une seule des deux approches sont des contaminants non mitochondriaux. Cela pourrait s'expliquer par les différences de conditions de croissance et de méthodes de purification. Néanmoins, 337 protéines ont été détectées par les deux études protéomiques, représentant 62 % de toutes les protéines détectées par Prokisch *et coll.* [67]. En intégrant ces deux études, la sensibilité diminue à 49,9 % et la spécificité augmente à 63,1 % (Tableau 6).

Les premières études sur le protéome des mitochondries de la plante *Arabidopsis thaliana* basé sur le 2D PAGE et la spectrométrie de masse ont conduit à l'identification de près de 100 protéines non redondantes [74, 75]. Les mêmes limitations à cause des protéines peu abondantes et hydrophobes comme discuté ci-dessus semblent avoir été résolues dans une étude de Heazlewood *et coll.* [71]. Des mitochondries hautement purifiées de *A. thaliana* par deux gradients de Percoll successifs ont été analysées par LC-MS/MS. Quatre cent seize protéines mitochondriales supposées ont ainsi été détectées. Une quantité considérable de protéines impliquées dans la signalisation cellulaire comme une kinase Sérine/Thréonine, une protéine kinase activée par des agents mitogènes, une sous-unité régulatrice 2A de protéine phosphatase, une protéine de liaison à la GTP et plusieurs protéines de liaison calcium-dépendantes ont été détectées. Fait intéressant, plusieurs candidats pour des protéines localisées à la mitochondrie, dont on ne connaît pas la fonction chez les plantes, semblent être absents de chez l'homme, la levure et *Rickettsia prowazekii* (qui est le plus proche descendant actuel de l'endosymbionte bactérien) [71]. Cela souligne la spécificité des fonctions de certaines protéines mitochondriales chez les plantes.

Le protéome mitochondrial humain a été largement étudié à l'aide de mitochondries de cœur hautement purifiées [69]. Les mitochondries ont été solubilisées avec un détergent non ionique et les complexes de protéines ont été séparés par centrifugation sur gradient de saccharose puis 1D SDS-PAGE. L'analyse par spectrométrie de masse a révélé plus de 540

protéines mitochondriales supposées (Tableau 6). Comme pour les plantes, des kinases orthologues, qui sont absentes des mitochondries de levure, ont également été détectées, ce qui est cohérent avec la présence de circuits de régulation avancés dans les organismes multicellulaires.

1.3.2 Les progrès et les perspectives dans l'analyse protéomique de la mitochondrie

Le nombre de faux positifs dans les études de protéomique dépend bien sûr de la pureté des mitochondries isolées. Un développement technique récent pour améliorer ceci est l'utilisation de la FFE pour purifier les mitochondries [67, 90]. Bien que la solubilisation et la détection seront certainement améliorées dans les études futures, toute approche protéomique est susceptible de ne pas mettre en évidence les protéines peu abondantes et également les protéines qui sont difficiles à extraire des échantillons biologiques. Fait intéressant, l'étude de Prokisch *et coll.* [67] prend en compte les protéines associées à la membrane. Cela est probablement dû à des avancées technologiques en matière d'épuration (c'est-à-dire la FFE) et en spectrométrie de masse (LC-MS/MS et LC-FTICR-MS). Toutefois, dans une autre étude, il a été observé un biais dans la détection des protéines associées à la membrane [64]. Les auteurs de cette étude attribuent cela à la plus faible quantité de protéines membranaires, en moyenne, par rapport aux protéines solubles et non pas à la limite de sensibilité de la spectrométrie de masse pour les protéines hydrophobes. Néanmoins, les protéines hydrophobes semblent toujours être un problème pour les approches fondées sur le 2D PAGE [76]. En outre, de nombreuses protéines sont exprimées d'une manière conditionnelle (par exemple, dans des conditions de stress) ou à certains stades de croissance ou de développement d'un organisme. Par conséquent, il sera difficile d'identifier ces protéines potentiellement intéressantes par une approche protéomique.

1.3.3 Les apports complémentaires de la génétique, de la bio-informatique et des approches transcriptomiques

1.3.3.1 Génétique

Deux études portent sur le marquage d'un grand nombre de gènes de la levure et la localisation des protéines de fusion exprimées par microscopie à fluorescence [81, 82]. Le marquage a été effectué de manière systématique par la fusion de la protéine fluorescente verte (*green fluorescent protein*, GFP) à l'extrémité C-terminale par mutagenèse aléatoire par transposition (*random transposon-mediated mutagenesis*) [82], ou par clonage de cadres de lecture ouverts (*open reading frames*, ORF) par amplification PCR dans un vecteur de

surexpression/marquage [81]. Ces études ont permis de mettre en évidence de nombreuses protéines essentielles et non essentielles attribuées à la mitochondrie, mais aussi à d'autres compartiments cellulaires. Huh *et coll.* [82] montrent en particulier que de hautes sensibilités et spécificités sont obtenues en prenant comme référence l'ensemble MITOP2 (Tableau 6). Dans une autre approche de marquage, la moitié N-terminale d'une GFP a été fusionnée au hasard avec les séquences d'une banque d'ADNc et co-exprimée avec un rapporteur ciblé dans la matrice contenant la moitié C-terminale de la GFP [91]. Lorsqu'une protéine de fusion exprimée contient les informations de ciblage vers la matrice mitochondriale, une GFP complète est générée par épissage protéique. Ainsi, le tri automatisé des cellules fluorescentes puis l'identification des protéines de fusion de l'ADNc exprimées sont devenus possibles. Il sera alors intéressant de voir quelles nouvelles protéines mitochondriales seront identifiées par cette approche. Dans sa forme actuelle, elle est limitée à la détection de protéines exposant leur extrémité C-terminale à la matrice mitochondriale.

Toutes les approches de marquage exigent que les niveaux d'expression soient suffisamment élevés et que les marqueurs n'interfèrent pas avec le ciblage correct de leur localisation. En effet, la modification de la séquence originale des acides aminés d'une protéine (par insertion d'un marqueur) ainsi que la surexpression peuvent conduire à une mauvaise localisation [68]. D'autres expériences indépendantes sont nécessaires pour considérer comme fiable la localisation d'une protéine. Par exemple, la similitude du phénotype d'une souche dont le gène candidat a été supprimé avec le phénotype d'une souche déficiente en une protéine mitochondriale connue. Le déficit respiratoire et la croissance ralentie sur des sources de carbone non fermentescible sont deux de ces phénotypes exploités dans les études à grande échelle [79, 80]. Ces méthodes sont réservées aux gènes non essentiels et ne prouvent manifestement pas qu'une protéine soit mitochondriale, car, par exemple, les facteurs de transcription nucléaires ou les régulateurs cytosoliques affectent également la fonction mitochondriale. La létalité provoquée par la suppression d'un gène candidat, de concert avec la suppression d'un gène codant pour une protéine mitochondriale connue, indique également une interaction fonctionnelle et, par conséquent, souvent une interaction physique. Même si une étude à grande échelle sur la détermination de la létalité synthétique chez la levure de boulanger a été réalisée [92], les résultats à ce jour n'ont pas été analysés pour la prédiction de protéines mitochondriales.

1.3.3.2 Transcriptomique

Une autre approche pour détecter les protéines mitochondriales est d'étudier le niveau d'expression des ARNm sous diverses conditions de croissance. Les réponses similaires des niveaux des transcrits de différents gènes permettent leur regroupement : les groupes corégulés sont plus susceptibles de représenter des réseaux de protéines interagissant

physiquement et fonctionnellement que les transcrits qui sont différentiellement régulés. Plusieurs études portant sur la synexpression ont été rapportées, y compris sur des gènes impliqués dans des fonctions mitochondriales [87, 88, 93, 94]. Deux d'entre elles sont décrites dans le Tableau 6. Ces études permettent de prévoir la localisation subcellulaire des protéines, mais seulement avec une spécificité faible à moyenne.

1.3.3.3 Bio-informatique

Une autre possibilité pour identifier les protéines localisées dans les mitochondries est l'utilisation de prédictions *in silico*. Certaines prédictions reposent sur le fait qu'un grand nombre de protéines destinées à la matrice ou à la membrane interne contient un peptide signal N-terminal chargé positivement et bipolaire [83-85]. Plusieurs algorithmes permettent une détection rapide de la présence d'une telle séquence (Tableau 6 et Tableau 7). En revanche, les protéines adressées au réticulum endoplasmique ou très chargées positivement, telles que les sous-unités ribosomales, donnent souvent des scores élevés, ce qui explique la spécificité faible ou moyenne de ces approches. Les protéines mitochondriales contenant des signaux d'adressage avec d'autres caractéristiques, en particulier les signaux internes, ne sont pas détectées. Ce dernier groupe comprend les protéines de la famille des transporteurs de la membrane interne (une classe de protéines contenant six hélices transmembranaires et impliquées dans le transport de petites molécules comme l'ADP ou l'ATP à travers la membrane interne), de nombreuses protéines intermembranaires, et probablement toutes les protéines de la membrane externe. Une approche bio-informatique plus globale a intégré des informations provenant de la génétique, de prévisions *in silico* et de corrélations entre des jeux de données d'expression des ARNm en utilisant un algorithme probabiliste de Bayes [86] (Tableau 6 et Tableau 7). Un avantage de ce type d'analyse réside dans la possibilité de facilement y intégrer d'autres types d'informations tels que des données de protéomiques.

Tableau 7. Bases de données et outils relatifs au protéome mitochondrial.

Base de données	Description	Site Internet
MITOP2 [67, 78]	Base de données contenant un grand nombre de données provenant de diverses études sur les mitochondries de <i>Saccharomyces cerevisiae</i> , <i>Neurospora crassa</i> et <i>Homo sapiens</i>	http://ihg.gsf.de/mitop2
MitoProteome [69]	Base de données sur le protéome mitochondrial humain	http://www.mitoproteome.org
TRIPLES [20]	Base de données pour la localisation des protéines dans la levure de boulanger par immunofluorescence en microscopie après marquage par épitope	http://ygac.med.yale.edu/triples
<i>Yeast deletion project and proteomics of mitochondria database</i> (YDPM) [67, 80]	Base de données contenant les informations sur des études protéomiques et d'expression à grande échelle sur la levure de boulanger	http://www-deletion.stanford.edu/YDPM/
<i>Yeast GFP fusion localization database</i> [82]	Base de données pour la localisation des protéines dans la levure de boulanger par microscopie en immunofluorescence de protéines de fusion GFP	http://yeastgfp.ucsf.edu
GOBASE [95]	Compilation de génomes mitochondriaux	http://gobase.bcm.umontreal.ca/
MITOMAP [96]	Base de données sur les séquences de l'ADNmt humain, incluant les mutations et polymorphismes connus	http://www.mitomap.org
Prévision bio-informatiques des protéines mitochondriales		
Mitoprot [83]	Analyse du signal d'adressage N-terminal	http://ihg.gsf.de/ihg/mitoprot.html
Predotar [85]	Analyse du signal d'adressage N-terminal	http://www.inra.fr/predotar/index.html
PSORT et autres [84]	Analyse du signal d'adressage N-terminal pour les bactéries, les plantes et d'autres séquences eucaryotes	http://www.psort.org

Base de données	Description	Site Internet
Analyse bayésienne [86]	Prédiction de la localisation subcellulaire basée sur 30 éléments différents	http://bioinfo.mbb.yale.edu/genome/localize
Mitopred [97]	Prédiction basée sur la présence de motifs de domaines protéiques, la composition en acides aminés et la <i>pI value</i>	http://bioapps.rit.albany.edu/MITOPRED/

1.3.4 L'origine et l'évolution du protéome mitochondrial

Comme nous l'avons vu précédemment (voir Le génome mitochondrial, p. 1), il est largement accepté qu'un seul événement endosymbiotique est à l'origine des mitochondries actuelles, à partir duquel un hôte eucaryote ancestral sans mitochondrie a intégré un endosymbionte α -protéobactérien [98, 99]. Bien que la plupart des gènes mitochondriaux aient été transférés au cours de l'évolution vers le noyau, la localisation et la fonction des protéines mitochondriales codées par ces gènes semblent souvent être maintenues. Par conséquent, une protéine candidate qui montrerait une homologie avec *Rickettsia prowazekii*, le plus proche descendant actuellement connu de la mitochondrie ancestrale, aurait plus de chances d'être adressée à la mitochondrie qu'une protéine qui n'a pas d'homologue dans cet organisme. Cela ne s'applique cependant pas à toutes les protéines mitochondriales, car certaines sont clairement codées par des gènes nucléaires qui étaient déjà présents dans l'hôte eucaryote ancestral. En outre, certains produits de gènes d'origine eubactérienne ont été recrutés par d'autres compartiments cellulaires que les mitochondries au cours de l'évolution [98]. Par conséquent, ce genre d'analyse peut, au mieux, être considéré comme une indication en faveur d'une localisation mitochondriale de la protéine correspondante.

1.3.5 Les approches intégratives

1.3.5.1 Les réseaux d'interactions protéine-protéine

L'étude des interactions protéine-protéine offre une nouvelle possibilité pour déchiffrer le protéome mitochondrial. Le principe est que toute interaction d'une protéine non identifiée avec une protéine mitochondriale connue révèle sa localisation mitochondriale. En plus de l'identité de la protéine candidate, des informations sur la(les) protéine(s) partenaire(s) d'interaction apportent des éléments sur sa fonction physiologique. Plusieurs grandes études ont été réalisées pour déterminer les réseaux d'interactions protéine-protéine chez la levure. Elles sont basées sur le criblage double-hybride à haut débit, l'identification par spectrométrie de masse de protéines copurifiées après isolement de complexes protéiques, la corrélation de l'expression des ARNm, la létalité synthétique de deux gènes, et les prédictions *in silico* à travers l'analyse du génome. Ces approches ont été analysées par von Mering *et coll.* [89]. Lors de l'utilisation de la purification d'affinité des complexes protéiques, seules les interactions stables sont détectées ainsi que les interactions qui ne sont pas perturbées par le marqueur d'affinité. Les dosages double-hybrides sont en mesure de trouver les interactions les plus labiles, mais sont particulièrement sujets aux faux positifs. Les faux positifs sont un inconvénient des méthodes à haut débit lorsqu'ils interviennent au

début de la chaîne des analyses. Par conséquent, les données issues de ces différentes études sur le génome entier ont été intégrées et les interactions protéine-protéine ont été regroupées en classes de moyenne et haute confiance [89]. Pour attribuer de nouvelles protéines à la mitochondrie par des méthodes non-protéomiques, les administrateurs de la base de données MITOP2 (<http://ihg.gsf.de/mitop2> [67, 77]) ont utilisé ces classes de confiance pour déduire de nouvelles protéines en fonction de la présence d'au moins une protéine mitochondriale connue dans ces interactions protéine-protéine (Tableau 6 et Tableau 7). Ainsi, la localisation mitochondriale de respectivement 815 et 154 protéines a été prédite. La sensibilité et la spécificité sont affectées différemment selon la catégorie. Lors de l'analyse de la classe des interactions protéine-protéine de haute confiance, une faible sensibilité, mais une haute spécificité ont été observées. Inversement, lors de l'analyse de la classe des interactions protéine-protéine de moyenne de confiance, une sensibilité élevée, mais une faible spécificité ont été observées (Tableau 6). Cela illustre bien le compromis entre la sensibilité et de spécificité.

1.3.5.2 Intégration de la transcriptomique et de la protéomique appliquée à la mitochondrie animale

Mootha *et coll.* ont récemment décrit un type d'analyse relativement intégratif et sophistiqué [64]. Ils montrent de manière exemplaire ce que l'on peut obtenir en combinant la protéomique basée sur les tissus et l'analyse à grande échelle de l'expression des ARNm. Des protéines hautement purifiées obtenues à partir de mitochondries de foie, cerveau, cœur et rein ont été séparées par filtration sur gel et chaque fraction a été soumise à digestion et LC-MS/MS. Ainsi, 399 protéines non redondantes présentes dans au moins un des tissus ont été obtenues. En combinant ces 399 protéines aux protéines mitochondriales connues, ils ont obtenu un ensemble de 591 protéines associées à la mitochondrie qui a été utilisé pour des analyses ultérieures. Parmi les 399 protéines détectées, 163 représentaient des protéines nouvellement identifiées.

La sensibilité et la spécificité pour chaque tissu peuvent être déterminées en analysant les données communiquées par les auteurs [64]. En dehors de l'ensemble de référence de 409 protéines mitochondriales, 217 protéines (53,1 % de sensibilité) ont été identifiées dans au moins un des tissus (Tableau 6). Pour le cœur, le cerveau, le foie et les reins, la sensibilité serait respectivement de 32,5 %, 31,5 %, 42,8 % et 37,9 % (Tableau 6). En outre, parmi les protéines qui ont été identifiées et qui étaient également dans l'ensemble de référence des protéines mitochondriales connues, seulement environ 40 % seraient présentes dans les quatre tissus [64]. Ce faible pourcentage ne semble pas être dû à des limitations méthodologiques telles que le sous-échantillonnage, mais peut être attribué, comme les auteurs le suggèrent, à la tissu-spécificité.

Cette interprétation qui intègre désormais des données de transcriptomique et de protéomique est basée sur les arguments suivants. Les auteurs ont testé si une protéine dans un tissu se comporte d'une manière concordante ou non avec son ARNm correspondant. Par exemple, une protéine qui est identifiée dans les mitochondries du cœur, mais pas dans celles du foie est considérée comme se comportant d'une manière concordante lorsque le niveau d'expression de ses ARNm est plus élevé dans le cœur que dans le foie. En effet, il y aurait une forte concordance entre l'identification des protéines et l'abondance des ARNm. Les auteurs ont utilisé cinq échantillons de tissus, deux de foie et un de cœur, rein et cerveau. Ensuite, ils ont déterminé la probabilité que la protéine détectée dans le premier échantillon de foie soit détectée dans le deuxième échantillon de foie et dans les échantillons provenant des autres tissus. Pour le deuxième échantillon de foie, cette probabilité serait de 92 %, mais 79 % en moyenne pour les autres tissus. Les auteurs proposent alors un modèle dans lequel les mitochondries comporteraient un sous-ensemble d'environ 50 % de protéines ubiquitaires et 50 % de protéines tissu-spécifiques, dont la moitié serait partagée entre deux tissus. Par conséquent, deux types de tissus devraient partager environ 75 % des protéines mitochondriales ce qui, au moins pour le foie, est proche des 79 % observés.

En outre, Mootha *et coll.* ont déterminé une liste de gènes qui sont régulés transcriptionnellement de la même manière que les gènes codant des protéines mitochondriales identifiées par protéomique. Ces gènes coréglés étendent la série des protéines « associées à la mitochondrie » à un « voisinage mitochondrial » comprenant 643 gènes au total. Ce travail montre bien la force des approches intégratives dans lesquels des données tissu-spécifiques de protéomique et de transcriptomique se complètent les unes les autres.

La plus complète intégration de protéomique, de transcriptomique, de bio-informatique et de données génétiques a été rapportée dans une analyse du protéome mitochondrial de *S. cerevisiae* [67]. Au total, 22 approches ont été intégrées, incluant une nouvelle étude protéomique et plusieurs autres études qui ont déjà été examinées en détail précédemment. Les chevauchements entre les différentes approches ont été analysés. Pour chaque protéine, une valeur appelée score MITOP2 a été calculée, représentant la spécificité de la meilleure combinaison des méthodes (<http://ihg.gsf.de/mitop2>). Le score MITOP2 varie de 0 à 100, indiquant une faible à forte probabilité qu'une protéine soit localisée dans la mitochondrie. La puissance de ce type d'intégration est mise en évidence lorsque le score MITOP2 est utilisé comme un critère pour prédire les protéines mitochondriales. Les auteurs ont calculé la spécificité et la sensibilité de différents ensembles de protéines avec un score MITOP2 supérieur à 90, 93, 95, 96 ou 97. Cela permet d'obtenir une bien meilleure

sensibilité (81,1 %) par rapport aux approches appliquées individuellement (Tableau 6), tout en conservant une haute spécificité (69,5 %).

1.3.6 Dysfonctionnements de la mitochondrie dans les pathologies humaines

Les pathologies mitochondriales regroupent de nombreuses affections héréditaires qui résultent d'un dysfonctionnement du métabolisme énergétique mitochondrial [100]. Elles sont très hétérogènes sur le plan de leur origine génétique et de leur expression phénotypique. La présentation clinique des maladies mitochondriales est principalement neuromusculaire et neurosensorielle, mais la plupart des grandes fonctions physiologiques peuvent être affectées. Elles impliquent classiquement l'association de plusieurs pathologies d'organes et s'aggravent progressivement le plus souvent. Les affections mitochondriales peuvent débuter à tout âge, des pics d'apparition de ces maladies sont cependant observés en période néonatale, dans l'enfance et chez l'adulte jeune.

Sur le plan physiopathologique, ces maladies sont très souvent la conséquence d'un défaut de la chaîne respiratoire mitochondriale qui est composée de 5 complexes multiprotéiques enchâssés dans la membrane interne et qui assurent le couplage entre l'oxydation des nutriments et la phosphorylation de l'ADP en ATP. La perturbation d'autres fonctions mitochondriales peut aussi être à l'origine de déficits énergétiques d'origine génétique (complexe de la PDH, cycle de Krebs, β -oxydation des acides gras, dynamique du réseau mitochondrial, transport transmembranaire, métabolisme du fer...). Outre le déficit énergétique, la surproduction d'espèces réactives de l'oxygène, les déséquilibres redox et une susceptibilité accrue à l'apoptose pourraient participer à la physiopathologie des cytopathies mitochondriales [101-103].

Sur le plan génétique, les maladies mitochondriales sont extrêmement hétérogènes. Plus d'une centaine de mutations pathogènes de l'ADNmt (environ 16,5 kb, 37 gènes) ont été rapportées depuis leur première description en 1988 [45] et de nouvelles mutations sont encore régulièrement décrites de nos jours (<http://www.mitomap.org/>) [42]. Ce sont principalement des mutations ponctuelles, de grandes délétions et des déplétions (réduction du taux d'ADNmt). Certaines mutations de l'ADNmt ont une transmission maternelle (mutations ponctuelles), d'autres apparaissent plutôt de manière sporadique (délétions simples) et d'autres enfin sont sous la dépendance de gènes nucléaires à transmission mendélienne (délétions multiples et déplétions). Depuis leur première description en 1995 [104], une cinquantaine de gènes nucléaires a été impliquée dans les cytopathies mitochondriales et la liste de ces gènes ne cesse de s'accroître. Ils codent soit des sous-unités de la chaîne respiratoire, soit des facteurs d'assemblage des complexes de la chaîne respiratoire, soit des facteurs nécessaires à la maintenance et à l'expression de l'ADNmt, mais de nombreuses autres fonctions mitochondriales peuvent être concernées. À l'heure

actuelle, la cause génétique d'une majorité d'affections mitochondriales demeure inconnue, ce qui rend difficile le conseil génétique.

L'épidémiologie des affections mitochondriales est mal connue et peu de données sont aujourd'hui disponibles pour estimer la prévalence des pathologies de la chaîne respiratoire [105]. Les études les plus récentes suggèrent que la prévalence des déficits de la chaîne respiratoire pourrait concerner plus de 1 individu sur 5 000 habitants en Europe. Mais la rareté de chaque pathologie prise séparément, les difficultés diagnostiques, le nombre important de gènes nucléaires impliqués, le fait que certains dysfonctionnements mitochondriaux puissent être secondaires à d'autres affections, rendent ces études épidémiologiques extrêmement délicates à entreprendre.

Le diagnostic des affections mitochondriales est particulièrement délicat. La présentation clinique, la génétique (mode de transmission), l'imagerie (IRM) et les dosages biochimiques (lactate, pyruvate, corps cétoniques, acides organiques, cycles redox...) permettent une première orientation diagnostique. Les dosages biochimiques spécialisés sont, dans un second temps et le plus souvent, nécessaires pour poser le diagnostic. Ces dosages concernent principalement la mesure de l'activité enzymatique des différents complexes de la chaîne respiratoire, mais aussi les dosages de la PDH et d'autres enzymes impliquées dans la maintenance de l'ADNmt. Enfin, si un déficit enzymatique est identifié, la recherche de mutations de l'ADNmt ou de gènes nucléaires est réalisée.

Les mutations de l'ADNmt, qui affectent des sous-unités essentielles de la chaîne respiratoire ou des gènes à ARNt mitochondriaux, sont associées à un grand nombre de maladies mitochondriales telles que le syndrome MELAS (*mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes* ou myopathie mitochondriale, encéphalopathie, acidose lactique et tableaux neurologiques aigus ressemblant à des accidents ischémiques cérébraux, MIM# 540000), le syndrome MERRF (*myoclonic epilepsy with ragged red fibers* ou épilepsie myoclonique avec fibres rouges déchiquetées, MIM #545000), la myocardiopathie, la neuropathie optique héréditaire de Leber (NOHL, MIM #535000), le syndrome de Leigh (MIM #256000) et beaucoup d'autres [102]. De plus, des altérations de l'ADNmt ont été observées dans des tumeurs du foie, de la prostate, de la vessie et de la tête et du cou, ainsi que dans les astrocytomes, les adénocarcinomes et les tumeurs primaires du poumon et l'œsophage de Barrett [106].

Des mutations dans des gènes nucléaires codant pour des protéines mitochondriales peuvent aussi être la cause de maladies humaines. Par exemple :

- L'ataxie de Friedreich (MIM #229300) est causée par des mutations de la frataxine, une protéine impliquée dans la biogenèse des protéines fer-souffre [107, 108].

- La protéine *Deafness dystonia peptide 1* (DDP1), impliquée dans l'importation de protéines dans les mitochondries, est associée au syndrome de Mohr-Tranebjaerg (MIM #304700) [109-112].
- L'atrophie optique de type I (MIM #165500), transmise selon un mode autosomique dominant, dans laquelle le gène codant pour la protéine Opa1 (MIM *605290) est affecté [113, 114]. Opa1 est impliquée dans le maintien de la morphologie des mitochondries et de son orthologue chez la levure, Mgm1, est essentiel pour la fusion mitochondriale [115, 116]. Dans la suite de ce document, nous allons nous intéresser plus particulièrement à la protéine Opa1 humaine.
- Récemment, une autre protéine mitochondriale impliquée dans la fusion, Mfn2 (MIM *608507), a été trouvée mutée dans des familles atteintes de la neuropathie de Charcot-Marie-Tooth de type 2A (MIM #118210, #609260) [117].

En plus de ces maladies bien documentées, plusieurs autres maladies humaines communes ont des causes mitochondriales. La déficience en glucose et l'altération du métabolisme des acides gras dus à des dysfonctionnements mitochondriaux sont des facteurs importants du diabète et de l'obésité [118]. En outre, la diminution des niveaux d'ATP cellulaire et de production des ROS due à un dysfonctionnement mitochondrial serait un facteur majeur contribuant aux maladies de Parkinson et d'Alzheimer [119, 120], mais le lien entre ce dysfonctionnement mitochondrial et ces maladies est loin d'être connu.

1.4 LES PATHOLOGIES ASSOCIÉES À OPA1 : DES MITOCHONDRIOPATHIES À PART ENTIÈRE

Le nerf optique transmet les informations visuelles de la rétine au cortex visuel. Chez l'homme, chaque nerf optique, qui comprend environ 1,2 million d'axones émis des cellules ganglionnaires de la rétine (CGR), reçoit l'information visuelle des photorécepteurs de la rétine. À la surface de cette dernière, les axones des CGR forment la couche des fibres nerveuses rétinienne (CFNR) et convergent vers la tête du nerf optique ou disque optique (Figure 7). Sortis de l'œil, ces axones sont myélinisés et transportent l'information visuelle au corps géniculé latéral du thalamus. Une perte des CGR ou de leurs axones conduit à une neuropathie optique, qui se manifeste à l'examen ophtalmologique du fond d'œil par une pâleur du disque optique ou à un stade plus grave, par une atrophie optique (Figure 8). Toutefois, l'atrophie optique n'est pas spécifique des neuropathies optiques héréditaires et de nombreux autres facteurs, comme la compression, l'infiltration, l'inflammation, l'ischémie, la toxicité ou les carences nutritionnelles peuvent donner lieu à des manifestations ophtalmologiques similaires.

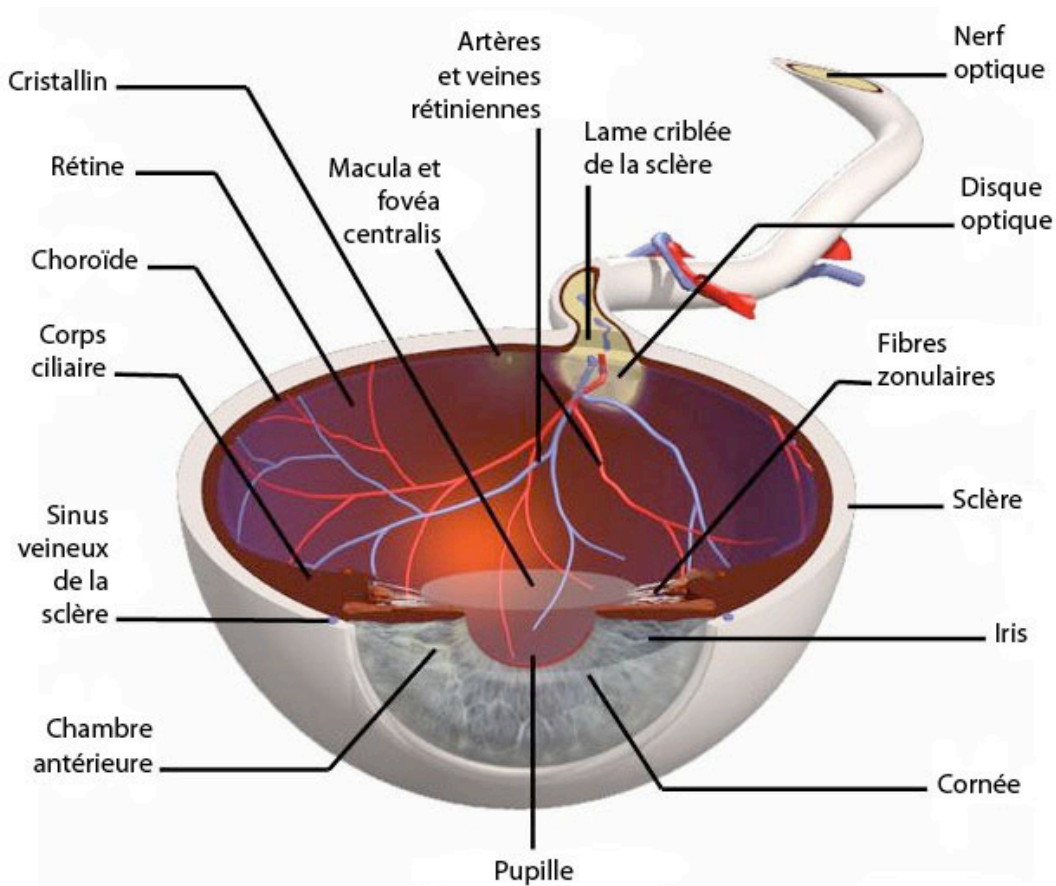


Figure 7. Coupe axiale de l'œil droit. D'après *l'Atlas d'anatomie 3D Primal Pictures* (usage éducation seul autorisé).

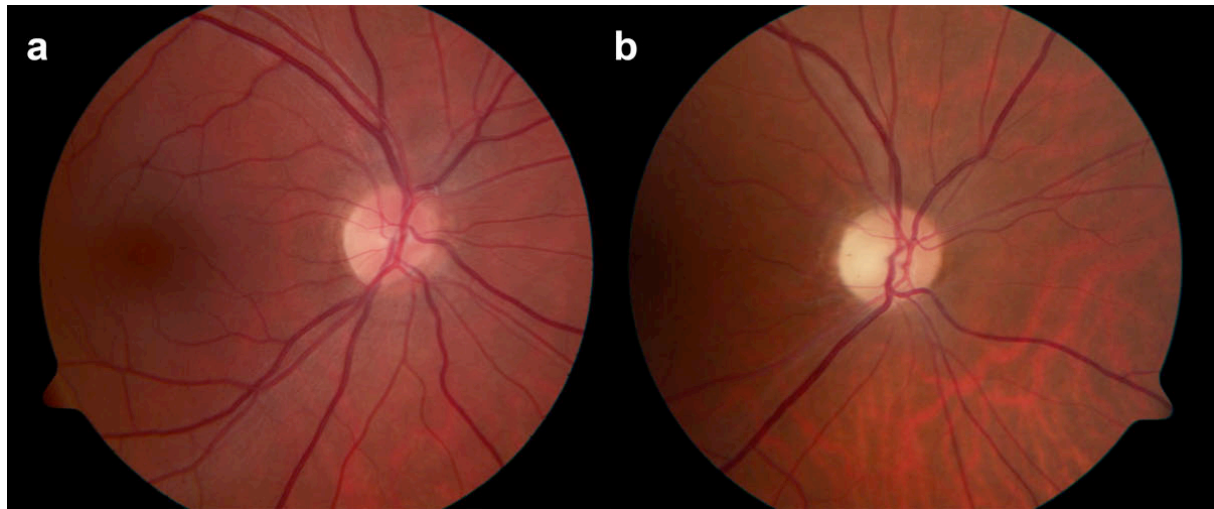


Figure 8. Fond d'œil : a) d'un disque optique droit normal ; et b) d'une atrophie diffuse du disque droit chez un patient atteint d'AOAD avec une vision de 20/200.

Les atrophies optiques héréditaires ont d'abord été décrites à la fin du 19^e siècle par deux ophtalmologistes britanniques, Frederick Batten et Simeon Snell [121, 122]. Un quart de siècle plus tôt, l'ophtalmologiste allemand Théodore Leber a signalé un trouble ophtalmologique [123], aujourd'hui connu sous le nom de neuropathie optique héréditaire de Leber (NOHL). Moins d'un siècle plus tard, l'ophtalmologiste danois Poul Kjer rapporta 19 familles souffrant d'une atrophie optique héréditaire dominante [124], maintenant appelée atrophie optique autosomique dominante (AOAD, MIM #165500) ou atrophie optique de Kjer. Le locus du principal gène responsable de l'AOAD, *optic atrophy 1* (*OPA1*, MIM *605290), a été cartographié sur le chromosome 3q28–29 [125, 126]. Par la suite, l'hétérogénéité génétique de l'AOAD a été mise en évidence puisque deux autres locus ont été identifiés, à savoir *OPA4* (MIM %605293) et *OPA5* (MIM %610708), localisés respectivement sur les chromosomes 18q12.2–q12.3 [127] et 22q12.1–q13.1 [128]. Il existe également d'autres formes rares de neuropathies optiques héréditaires : l'atrophie optique liée à l'X (*X-linked optic atrophy*, XLOA), localisée sur le locus Xp11.4–p11.21 (*OPA2*, MIM %311050) [129], et les atrophies optiques autosomiques récessives (AOAR) *OPA3* (#258501) et *OPA6* (MIM %258500), localisées respectivement sur les chromosomes 19q13.2–13.3 [130] et 8q21–q22 [130–132], et l'AOAR non syndromique *OPA7*, localisée sur le locus 11q14.1–q21 (*TMEM126A*, *612988) [132]. Actuellement, aucun des gènes responsables des atrophies optiques *OPA2*, *OPA4*, *OPA5* et *OPA6* n'a été identifié à ce jour (Tableau 8).

Tableau 8. Loci d'atrophies optiques héréditaires. AD : autosomique dominante ; RLX : récessive liée à l'X ; AR : autosomique récessive.

Symbole	Transmission	Identifiant MIM	Locus	Gène
OPA1	AD	*605290	3q28-q29	<i>OPA1</i>
OPA2	RLX	%311050	Xp11	Non identifié
OPA3	AR/AD	*606580	19q13	<i>OPA3</i>
OPA4	AD	%605293	18q12	Non identifié
OPA5	AD	%610708	22q12.1-q13.1	Non identifié
OPA6	AR	%258500	8q21-q22	Non identifié
OPA7	AR	*612988	11q14.1-q21	<i>TMEM126A</i>

La NOHL, avec une prévalence allant de 1/30 000 à 1/50 000 [133, 134], et l'AOAD, avec une prévalence de 1/50 000 [135], sont les formes de neuropathies optiques héréditaires les plus communes. Il est intéressant de noter qu'en raison d'un effet fondateur, l'AOAD est plus fréquente au Danemark (1/10 000) que dans toute autre partie du monde [136, 137]. En effet, la mutation c.2826delT qui est très répandue dans la population danoise aurait été introduite il y a au moins 69 générations, avant l'époque viking [137].

La NOHL, qui a un mode de transmission maternel, a été la première maladie pour laquelle une mutation dans l'ADNmt, la mutation m.11778G>A, a été identifiée [45]. Par la suite, une douzaine d'autres mutations de l'ADNmt causant des NOHL, affectant principalement les gènes codant pour les sous-unités du complexe I de la chaîne respiratoire, la NADH:ubiquinone oxydoréductase, ont été identifiées (<http://www.mitomap.org>) [42], trois d'entre elles représentant environ 95 % des cas décrits (m.11778G>A, m.3460G>A et m.14484T>C). L'implication des dysfonctionnements de la chaîne respiratoire dans les NOHL souligne la forte dépendance des CGR en besoins énergétiques.

En 2000, le lien entre l'atrophie optique héréditaire et les mitochondries a été renforcé par l'identification du gène *OPA1*, qui code une protéine mitochondriale de la famille des GTPases s'apparentant aux dynamines, comme étant le principal gène impliqué dans l'AOAD [113, 114]. La mise en évidence du rôle des protéines Opa1 dans la structure et la morphologie du réseau mitochondrial a ouvert de nouvelles voies pour la compréhension des mécanismes des maladies neurodégénératives. En 2004, de nouveaux arguments en faveur de l'implication des mitochondries dans la physiopathologie des neuropathies optiques héréditaires ont été fournis par la découverte de mutations dominantes dans *OPA3* (*606580), un autre gène codant pour une protéine mitochondriale interne, dans un syndrome associant l'AOAD et la cataracte (AOADC, MIM #165300) [138].

Depuis 2000, le séquençage systématique du gène *OPA1* chez les patients présentant différents aspects cliniques de neuropathie optique a révélé un large spectre de phénotypes cliniques liés aux mutations d'*OPA1*. Aujourd'hui, jusqu'à 10 % des patients porteurs de mutations d'*OPA1* sont affectés par des anomalies extraophtalmologiques en association avec une neuropathie optique, ce qui montre que Opa1 peut être responsable d'un continuum de phénotypes allant de légers troubles affectant seulement les CGR à des phénotypes sévères et complexes de maladies multisystémiques.

Dans les parties suivantes, nous passerons en revue l'éventail des présentations cliniques des mutations d'*OPA1* ainsi que les mécanismes moléculaires et cellulaires impliqués dans la physiopathologie des troubles associés à Opa1.

1.4.1 Description clinique des AOAD

L'AOAD est classiquement décrite comme une neuropathie optique isolée, avec un début insidieux au cours des deux premières décennies de la vie. L'atteinte généralement simultanée des deux nerfs optiques aboutit à une perte visuelle bilatérale, habituellement progressive, due à des scotomes centrocaux, centraux ou paracentraux, bien que les atteintes bitemporales et paracentrales soient une caractéristique commune de la maladie (Figure 9). L'âge exact d'apparition de la maladie n'est presque jamais clairement établi, la plupart des patients ayant été diagnostiqués lors de l'entrée à l'école, ou seulement au cours de l'examen d'autres membres de la famille. Ensuite, la vision se détériore lentement chez certaines personnes avec l'âge ; aucune récupération spontanée de la vue n'a été signalée, sauf dans un cas atypique [139]. La tritanopie, c'est-à-dire la non-perception du bleu, est le défaut de vision des couleurs classiquement observé, mais une dyschromatopsie non spécifique diffuse, c'est-à-dire une difficulté de discrimination des couleurs, n'est pas rare. L'examen du fond d'œil révèle une pâleur temporale ou diffuse des disques optiques (Figure 8b), parfois associée à l'excavation du disque optique. Le bord neurorétinal montre une pâleur dans la plupart des cas, parfois associée à une pigmentation temporale grise courbée.

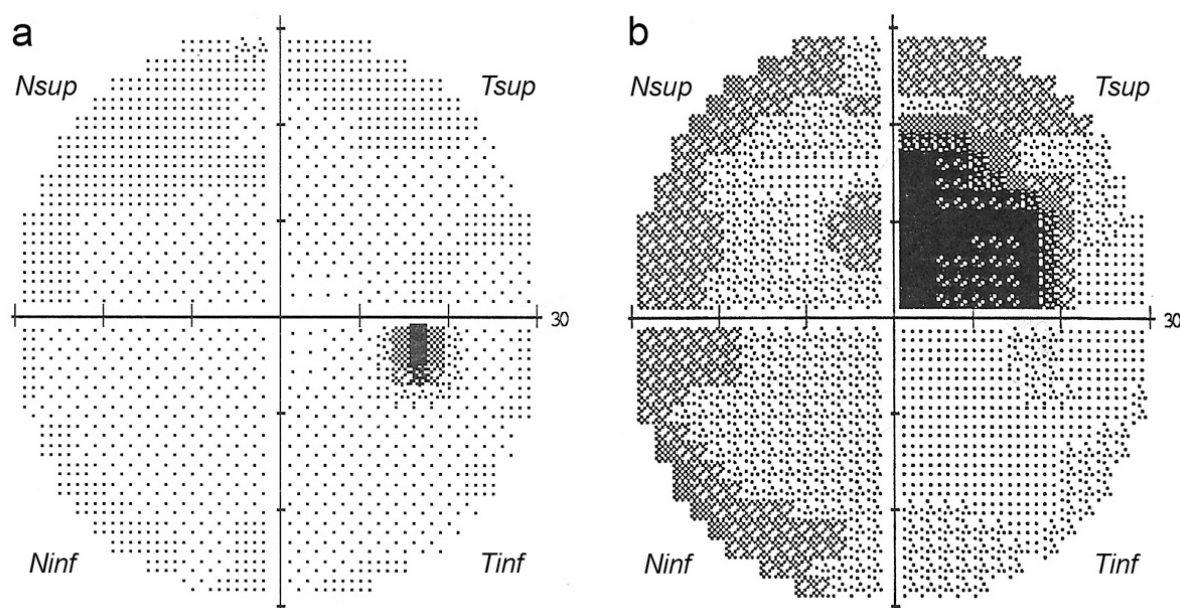


Figure 9. Champs visuels^a, explorés par périmétrie 30-2 Humphrey automatisée statique. Seuls les enregistrements de l'œil droit sont montrés : a) chez un patient ayant une vision normale et un fond d'œil normal (Figure 8a) ; et b) chez un patient atteint d'AOAD (fond de l'œil Figure 8b) montrant une grande irrégularité temporelle paracentrale.

L'AOAD a une pénétrance incomplète et son expressivité clinique est très hétérogène au sein même des familles [140, 141]. La perte visuelle est généralement modérée, bien que dans certains cas, elle puisse s'étendre à la cécité légale ou ne permettre que la perception de la lumière [141]. Étant donné que la perte visuelle n'est souvent pas un symptôme fonctionnel majeur, l'AOAD pourrait être sous-diagnostiquée. Une histoire familiale de neuropathie optique bilatérale, touchant autant les hommes que les femmes, appuiera le mode de transmission autosomique dominant, indiquant ainsi une AOAD.

L'électrophysiologie ne contribue pas au diagnostic de l'AOAD, car le tracé des potentiels évoqués visuels, montrant un retard et des réponses non spécifiques, est visible dans toute neuropathie optique. Toutefois, les tracés électrorétinographiques peuvent révéler une amplitude réduite de la forme d'onde N95, suggérant un dysfonctionnement des CGR [141].

Les études histopathologiques sur les AOAD sont assez rares, mais elles sont compatibles avec l'hypothèse de perte de CGR. Seules deux études post-mortem ont été publiées à ce jour, décrivant une perte sélective des CGR et des fibres nerveuses dans le faisceau papillomaculaire [142, 143]. Récemment, la tomographie de cohérence optique (TCO) *in vivo* a permis une meilleure évaluation de l'épaisseur de la couche des fibres

^a Champ visuel : *Nsup* : nasal supérieur ; *Ninf* : nasal inférieur ; *Tsup* : temporal supérieur ; *Tinf* ; temporal inférieur.

nerveuses rétiniennes (CNFR) chez les patients atteints de différents types de neuropathies optiques, y compris l'AOAD [144, 145]. Ces études ont confirmé l'amincissement de la CNFR et de la couche contenant les cellules ganglionnaires chez les patients atteints d'AOAD, tandis que les couches des cellules photoréceptrices seraient intactes. En outre, l'épaisseur maculaire totale serait considérablement réduite, particulièrement à la périphérie de la fovéa où la densité des CGR est la plus élevée [145]. Cependant, d'autres études en TCO seront nécessaires pour déterminer si la CNFR est mince dès la naissance chez les patients atteints d'AOAD par rapport à des individus en bonne santé.

La présentation clinique classique des AOAD est distincte de celle des NOHL [146], qui se caractérise par l'apparition de graves problèmes de perte visuelle séquentielle dans les deux yeux à un âge plus avancé (généralement 18–35 ans). La NOHL, qui est transmise exclusivement par la mère au travers des mutations de l'ADNmt, affecte surtout les hommes, alors que dans l'AOAD les hommes et les femmes sont également touchés.

1.4.2 OPA1 est le gène majeur associé à l'AOAD

Bien que les mutations du gène *OPA1* soient la principale cause génétique d'AOAD, l'étiologie génétique de la maladie reste hétérogène. Le taux de détection des mutations d'*OPA1* dans les familles atteintes d'AOAD est très variable, allant de 32 % à 89 % selon les études. Cette variation est probablement due aux critères cliniques utilisés pour la sélection des patients [147-153].

Les AOAD liées à *OPA1* ont une pénétrance incomplète, certains porteurs de mutations d'*OPA1* étant asymptomatiques. Initialement, la pénétrance de la maladie a été estimée à 98 % [154], mais l'identification des mutations d'*OPA1* a conduit à une réévaluation avec des résultats allant de 43 % (3/7) et 62 % (8/13) dans deux familles [153] à 82,5 % dans une grande enquête portant sur 17 familles [148], et 90 % (27/30) dans une étude comprenant 10 familles [155]. Cette pénétrance variable et incomplète suggère que le contexte génétique et les facteurs environnementaux sont susceptibles d'interférer avec l'expression de la maladie. D'ailleurs, il a été montré que l'expression clinique de l'AOAD serait influencée par le patrimoine génétique de l'ADNmt dans un sous-ensemble de cas, avec une surreprésentation de trois fois de l'haplogroupe mitochondrial J chez les patients atteints d'AOAD par rapport aux contrôles [156].

Compte tenu de la pénétrance variable de l'AOAD, de l'existence de cas sporadiques et de la petite taille de certaines familles, les principaux critères de diagnostic de la transmission autosomique dominante font souvent défaut [147]. En effet, au cours des dernières années, nous avons constaté que, dans une population de patients non liés adressés à notre laboratoire pour une suspicion de neuropathie optique héréditaire, environ la moitié de ceux présentant une mutation d'*OPA1* serait des cas sporadiques [157]. Cela

renforce la nécessité d'étudier ce gène lorsque le clinicien est confronté à des cas d'atrophie optique sans preuve de la transmission familiale de la maladie. En outre, l'AOAD peut parfois affecter des patients à la naissance, comme en témoigne l'identification d'une mutation d'*OPA1* chez un patient souffrant d'une forme congénitale grave de la maladie [147].

1.4.3 Expressivité variable de l'AOAD

Le dépistage moléculaire systématique d'*OPA1* chez les patients atteints d'AOAD a révélé un large éventail de variations phénotypiques de la maladie. La détection des mutations d'*OPA1* a montré que de nombreux cas d'atrophie optique, qui n'auraient pas été classés comme AOAD sur la base des critères cliniques, seraient en fait liés à la maladie. Les différentes présentations cliniques de l'AOAD sont décrites ci-après et les caractéristiques principales sont résumées dans le Tableau 9.

Tableau 9. Présentations cliniques associées aux mutations d'Opa1. CGR : cellules ganglionnaires de la rétine ; LCR : liquide céphalorachidien ; NOHL : neuropathie optique héréditaire de Leber ; OXPHOS : phosphorylation oxydative (*oxidative phosphorylation*) ; CPEO : ophtalmoplégie chronique progressive externe (chronique *progressive external ophthalmoplegia*) ; RRF : Fibres rouges déchiquetées (*ragged red fibers*) ; SNC : système nerveux central.

Pathologie associée à Opa1	Organes/tissus/cellules affectés	Symptômes et signes	Pathomécanisme	Références
AOAD type Kjer (MIM #165500)	CGR et nerf optique	Neuropathie optique	Haplo-insuffisance Défaut OXPHOS + Fragmentation mitochondriale ±	Alexander <i>et coll.</i> , 2000 [113] Delettre <i>et coll.</i> , 2000 [114] Lodi <i>et coll.</i> , 2004 [158] Chevrollier <i>et coll.</i> , 2008 [159]
AOAD réversible	CGR et nerf optique	AOAD sévère	Découplage mitochondrial	Cornille <i>et coll.</i> , 2008 [139]
AOAD sévère avec apparition tardive	CGR et nerf optique	Atrophie optique sévère Apparition tardive (62 ans)	Défaut OXPHOS + Fragmentation mitochondriale +	Nochez <i>et coll.</i> , 2009 [160]
AOADS (MIM #125250)	CGR et nerf optique Nerf auditif ± Muscle	Neuropathie optique Neuropathie auditive ± CPEO	Mutation p.R445H +++ Effet dominant négatif ? Défaut OXPHOS ++ Fragmentation mitochondriale +++	Shimizu <i>et coll.</i> , 2003 [161] Amati-Bonneau <i>et coll.</i> , 2003 [162] Payne <i>et coll.</i> , 2004 [163] Puomila <i>et coll.</i> , 2005 [155] Li <i>et coll.</i> , 2005 [164] Amati-Bonneau <i>et coll.</i> , 2005 [165] Ke <i>et coll.</i> , 2006 [166]

Pathologie associée à Opa1	Organes/tissus/ cellules affectés	Symptômes et signes	Pathomécanisme	Références
Syndrome AOAD « plus »	CGR et nerf optique Nerf auditif Cerveau Nerfs périphériques Muscle	Neuropathie optique Surdité de perception Ataxie Polyneuropathie sensorimotrice CPEO and mitochondrial Myopathie RRF et fibres COX-négatives	Haplo-insuffisance Défaut OXPHOS + Fragmentation mitochondriale ± Délétions ADNmt dans le muscle	Amati-Bonneau <i>et coll.</i> , 2008 [167] Hudson <i>et coll.</i> , 2008 [168] Liguori <i>et coll.</i> , 2008 [169]
AOAD et neuropathie périphérique	CGR et nerf optique Nerfs périphériques	Neuropathie optique Polyneuropathie sensorimotrice	Pas de défaut OXPHOS Fragmentation mitochondriale ++	Spinazzi <i>et coll.</i> , 2008 [170]
AOAD et symptômes évoquant la sclérose en plaques	CGR et nerf optique SNC	Atrophie optique bilatérale Hyperintensités de la substance blanche du cerveau et dans la moelle épinière Bandes oligoclonales dans le LCR à l'électrophorèse	Défaut OXPHOS ++ Similaire à la NOHL avec des symptômes évoquant la sclérose en plaques	Verny <i>et coll.</i> , 2008 [171]

1.4.3.1 L'AOAD réversible

L'AOAD typique est associée à une perte progressive et irréversible de la vision. Toutefois, le cas d'un homme de 23 ans qui présentait une neuropathie optique bilatérale inexplicée, isolée, progressive et indolore due à des scotomes centraux (acuité visuelle de 20/200 à l'œil droit et 20/100 à l'œil gauche, trois mois après les premiers signes de la perte visuelle), suivie ensuite d'une récupération visuelle partielle, spontanée et durable 6 mois plus tard (20/30 à l'œil droit et 20/25 à l'œil gauche) a été rapporté [139]. Le patient hétérozygote était porteur d'une mutation dans l'exon 5b (c.740G>A du transcrit variant 8 d'*OPA1*, RefSeq : NM_130837.1), qui a été la première mutation décrite dans l'un des trois exons alternatifs d'*OPA1*, entraînant une modification des acides aminés N-terminaux du domaine superhélice (*coiled-coil*) (p.R247H de l'isoforme 8 d'Opa1, RefSeq : NP_570850.1). Cette mutation induit des changements inhabituels du réseau mitochondrial, avec des mitochondries tubulaires, contrastant avec la fragmentation du réseau généralement observée dans les fibroblastes de patients atteints d'un syndrome Opa1 (patients Opa1). Une augmentation significative de la susceptibilité à l'apoptose, un découplage de la phosphorylation oxydative conduisant à un défaut de production d'ATP, et une altération du profil de la protéine Opa1 ont également été associés au phénotype des fibroblastes de patients. Cette observation a élargi le spectre de l'AOAD, suggérant que la récupération partielle pourrait intervenir dans cette maladie comme cela a été observé dans certains cas rares de NOHL. Ainsi, tous ces résultats montrent que les mécanismes pathologiques qui sous-tendent l'AOAD peuvent varier largement selon le domaine d'Opa1 défectueux ou les exons concernés.

1.4.3.2 L'AOAD évoquant une NOHL

Une atrophie optique aiguë et d'apparition tardive a été diagnostiquée chez une femme de 62 ans qui a été adressée au service d'ophtalmologie pour une baisse d'acuité visuelle brutale et sévère de l'œil droit, suivie un an plus tard par la perte d'acuité visuelle de son œil gauche [160]. La patiente était porteuse d'une nouvelle mutation faux-sens hétérozygote (c.2794C>T du transcrit variant 1, RefSeq : NM_015560.1) dans l'exon 27 d'*OPA1*, donnant un changement d'acide aminé très conservé (p.R932C de l'isoforme 1, RefSeq : NP_056375.1). L'étude des fibroblastes a révélé un défaut de couplage de la phosphorylation oxydative mitochondriale et une fragmentation partielle du réseau. Ces résultats suggèrent que, contrairement à l'apparition classique de l'AOAD, c'est-à-dire une atrophie optique insidieuse et progressive au cours de l'enfance, des présentations cliniques aiguës ou retardées de la maladie peuvent se produire. Dans notre laboratoire, plusieurs patients adressés pour la recherche de mutations NOHL de l'ADNmt s'avéraient être

porteurs d'une mutation d'*OPA1* (Amati-Bonneau *et coll.*, données non publiées), suggérant un chevauchement entre les manifestations cliniques de ces deux maladies. Ainsi, dans le cas d'une atrophie optique qui n'implique pas une transmission père-fils, ce qui exclut la NOHL où la transmission est maternelle depuis plusieurs générations et rend peu probable l'AOAD, les bases moléculaires de ces deux maladies devraient être étudiées.

1.4.3.3 L'AOAD associée à des symptômes évoquant la sclérose en plaques

Le cas d'un homme de 44 ans a récemment été décrit avec une atrophie optique et des symptômes évoquant une sclérose en plaques. Il est porteur d'une nouvelle mutation faux-sens d'*Opa1* (p.S646L de l'isoforme 1, RefSeq : NP_056375.1) dans le domaine dynamine qui est hautement conservé [171]. Le défaut énergétique, consistant en une réduction de la synthèse d'ATP et une diminution de la fonction respiratoire, observé dans les fibroblastes du patient par rapport à des fibroblastes d'autres patients *Opa1* ou de témoins suggérerait une relation entre la dysfonction mitochondriale et la fragilité du système nerveux central. D'ailleurs, plusieurs cas de patients atteints de NOHL avec des caractéristiques identiques à la sclérose en plaques (SEP) ont été signalés [172]. Bien que d'autres éléments soient nécessaires pour établir un lien entre les atrophies optiques liées à *Opa1* et les symptômes évoquant la sclérose en plaques, ces observations tendent à renforcer encore davantage le chevauchement entre les manifestations cliniques de la NOHL et de l'AOAD.

1.4.3.4 L'AOAD et surdité (AOADS) un nouveau syndrome dû à une mutation spécifique d'*Opa1*

L'association entre une surdité congénitale et une atrophie optique progressive, d'un mode de transmission dominant, a été signalée avant la découverte d'*OPA1* [140, 173-175]. En 2003, deux publications décrivent des cas sporadiques de patients souffrant de surdité modérée associée à une atrophie optique et présentant les mêmes mutations d'*OPA1*. Cette mutation faux-sens hétérozygote dans l'exon 14 (c.1334G>A du transcrit variant 1, RefSeq : NM_015560.1) conduit à la modification d'un acide aminé extrêmement conservé (p.R445H de l'isoforme 1, RefSeq : NP_056375.1) dans le domaine GTPase de *Opa1* [161, 162]. Plus tard, la même mutation a été impliquée dans un syndrome observé chez les membres d'une grande famille de l'Utah, associant AOAD, surdité, ptôsis et ophtalmoplégie [163]. Une autre famille de l'Ohio a aussi été décrite avec une AOADS et la mutation p.R445H [164]. Également, quatre patients porteurs de la mutation p.R445H d'*Opa1* avec un syndrome associant une AOAD typique et une surdité de perception modérée, en raison d'une neuropathie auditive, ont été rapportés [165]. Seules deux autres mutations d'*OPA1* (c.970delCGTTCTCCA et p.G401D de l'isoforme 1, RefSeq : NP_056375.1) ont été associées au phénotype AOADS [155, 166], ce qui suggère une étroite corrélation

génotype/phénotype entre la mutation p.R445H et ce syndrome et soulève des interrogations sur le mécanisme sous-jacent à cette condition clinique.

Les fibroblastes p.R445H ont le phénotype le plus marqué, en comparaison avec d'autres mutations d'Opa1, combinant une fragmentation du réseau mitochondrial, une altération de la phosphorylation oxydative et une altération importante de la synthèse d'ATP [165]. La déstructuration du réseau mitochondrial pourrait donc conduire à un défaut de couplage mitochondrial responsable de la diminution de la production d'ATP. Les nerfs optiques et auditifs pourraient être particulièrement sensibles à cette perte d'énergie, une hypothèse qui est également soutenue par le fait que l'atrophie optique et la surdité sont souvent présentes simultanément ou séparément dans d'autres maladies mitochondriales.

1.4.3.5 L'atteinte multisystémique complexe « AOAD plus » associée à des mutations d'OPA1 et à une instabilité et de l'ADNmt

En 2008, sept familles étudiées par les laboratoires de Newcastle, Bologne, Madrid, Montpellier et Angers ont mis en évidence une grave atteinte multisystémique associée à des mutations spécifiques dans le domaine GTPase d'OPA1 [167, 168, 176]. Cette atteinte, nommée syndrome "AOAD plus", analogue à celle décrite par Payne *et coll.* en 2004 [163] et par Liguori *et coll.* en 2008 [169], comporte une atrophie optique apparaissant dans l'enfance, suivie par une CPEO, une ataxie, une surdité neurosensorielle, une neuropathie sensorimotrice et une myopathie à l'âge adulte. La myopathie est typique des pathologies de la chaîne respiratoire mitochondriale avec la présence de fibres cytochrome *c* oxydase (COX) négatives et de fibres rouges déchiquetées (*Ragged Red Fibers*, RRF). Le plus intéressant est que tous les patients possèdent des délétions multiples de l'ADNmt dans le muscle squelettique, ce qui suggère que Opa1 joue un rôle clé dans la maintenance de l'ADNmt. L'instabilité structurelle de l'ADNmt est une cause majeure de maladies mitochondriales et, souvent, résulte de la mutation de gènes impliqués dans le métabolisme des nucléotides ou de l'ADNmt [176]. Ainsi, OPA1 est le cinquième gène responsable des délétions multiples de l'ADNmt après ceux codant pour les polymérase gamma 1 et 2 (Polg1, Polg2) et pour les protéines Twinkle et Ant1. Par la suite, le dépistage de 21 propositus avec des pathologies similaires liées aux délétions de l'ADNmt a permis d'identifier trois patients avec des mutations d'OPA1, ce qui montre que ce gène pourrait être impliqué relativement fréquemment dans ces troubles [177].

Ainsi, les syndromes AOAD, AOADS et « AOAD plus » sont des entités cliniques distinctes partageant une atrophie optique et des mutations d'OPA1. Concernant l'« AOAD plus », deux familles ont été présentées portant les mêmes mutations délétères p.R445H, alors que trois autres familles ont des mutations hétérozygotes dans le domaine GTPase. Récemment, une délétion dans le domaine GTPase d'Opa1 a également été trouvée dans

une famille présentant une neuropathie optique, associée à un syndrome limité à une AOAD et une neuropathie périphérique [170].

Ainsi, le dépistage génétique systématique d'*OPA1* a confirmé la forte variabilité des AOAD liées à *Opa1* et a révélé de nouvelles présentations cliniques inattendues. En effet, plusieurs formes multisystémiques d'AOAD ont été décrites montrant que les CGR, bien que particulièrement sensibles, ne sont pas la seule cible du mécanisme physiopathologique. Le nombre croissant de présentations cliniques atypiques de patients porteurs de mutations d'*OPA1* indique que ces mutations sont responsables d'un large spectre de maladies assez similaires à d'autres maladies affectant la chaîne respiratoire mitochondriale.

Il est intéressant de noter que toutes les présentations cliniques des AOAD associées à des mutations d'*OPA1* décrites jusqu'à présent incluent des neuropathies optiques. Seule une publication récente présente le cas d'un syndrome mitochondrial qui n'est pas associé à une atrophie optique [178].

1.4.4 Spectre des mutations du gène *OPA1*

Le gène *OPA1*, localisé sur le chromosome 3q28–29, s'étend sur environ 100 kb et est composé de 30 exons codants, dont trois exons alternatifs (4, 4b et 5b), résultants en huit variants de transcription [149]. L'isoforme principale (l'isoforme 1), codée par un cadre ouvert de lecture (ORF pour *open reading frame*) de 2 880 nucléotides de long, excluant les exons 4b et 5b, est une protéine de 960 acides aminés, alors que le plus long ORF, comprenant tous les exons alternatifs, code pour une protéine prédite de 1 015 acides aminés. La protéine *Opa1* est ubiquitairement exprimée et est principalement localisée sur la face externe de la membrane interne mitochondriale. Les principaux domaines d'*Opa1* sont la séquence d'adressage mitochondrial N-terminale, les domaines superhélice (*coiled-coil*) et GED (*GTPase effector domain*) putatif C-terminaux, le domaine GTPase et la région centrale dynamique (Figure 10).

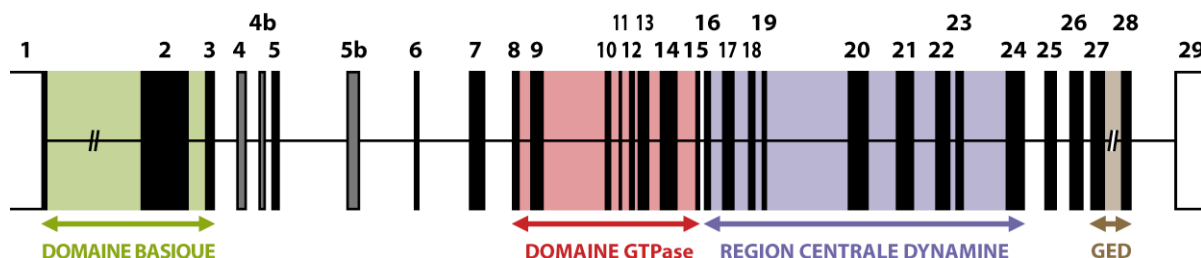


Figure 10. Représentation schématique de la structure du gène *OPA1* humain. Les exons sont représentés par des rectangles noirs ; ceux soumis à un épissage alternatif apparaissent en gris. Les domaines de la protéine *Opa1* sont représentés en arrière-plan. GED : domaine GED (*GTPase effector domain*) putatif.

Nous avons développé une base de données spécifique d'*OPA1*, nommée eOPA1 que nous décrirons plus loin (voir Second article : « eOPA1: une base de données pour les mutations du gène *OPA1* », p. 106), afin de recenser les variations de séquence de ce gène (<http://lbbma.univ-angers.fr/eOPA1/>) [179]. Toutes les mutations et variants de séquence non pathogènes d'*OPA1* sont inclus dans cette base de données selon les données publiées ou soumises directement par les laboratoires. Deux-cent-quatre mutations pathogènes d'*OPA1*, souvent famille-spécifiques, sont recensées dans la base de données (juillet 2009) [179]. Ces mutations sont réparties tout au long de la protéine, mais le domaine GTPase, la région centrale dynamine et le domaine C-terminal sont plus fréquemment touchés (Figure 11) avec 14 mutations (6,9 %) dans le domaine basique d'adressage mitochondrial, 76 (37,2 %) dans le domaine GTPase, 67 (32,8 %) dans la région centrale dynamine, et 37 (18,1 %) dans le domaine C-terminal.

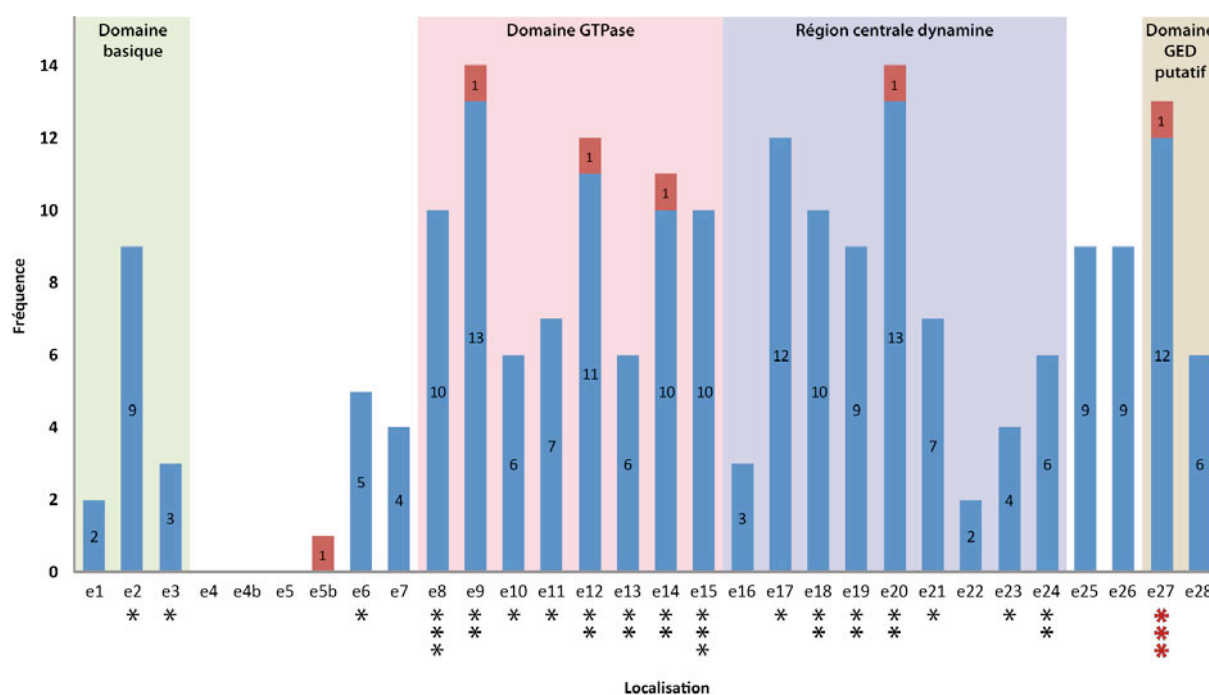


Figure 11. Données provenant de la base de données eOPA1 (juillet 2009) [179]. Distribution des 204 mutations pathogènes dans le gène *OPA1* en fonction des exons impliqués, comprenant leur voisinage intronique. La fréquence des mutations d'*OPA1* chez les patients est symbolisée par des étoiles, le nombre d'étoiles étant corrélé à la fréquence des mutations dans chaque exon. GED : *GTPase effector domaine*.

Au niveau de la protéine, 82 mutations (40 %) causent une terminaison prématurée de la traduction (33 mutations non-sens et 49 décalant le cadre de lecture), soutenant l'hypothèse de l'haplo-insuffisance comme mécanisme pathogène, 55 (27 %) sont des mutations faux-sens, 55 (27 %) sont des variants d'épissage et 12 (6 %) sont des délétions ou des duplications (Figure 12). Ces mutations sont pour la plupart famille-spécifiques, mais

une mutation dans l'exon 27 (c.2708_2711delTTAG), retrouvée dans environ 17 % des cas, représente un point chaud de mutation. La fréquence des mutations d'*OPA1* chez les patients est indiquée dans la Figure 11. En outre, une grande délétion hétérozygote d'*OPA1* a été rapportée [150] et des réarrangements génomiques, habituellement non détectés par séquençage, ont été récemment montrés dans 8 sur 42 patients atteints d'AOAD par la technique de MLPA (*multiplex ligation probe amplification*) [180].

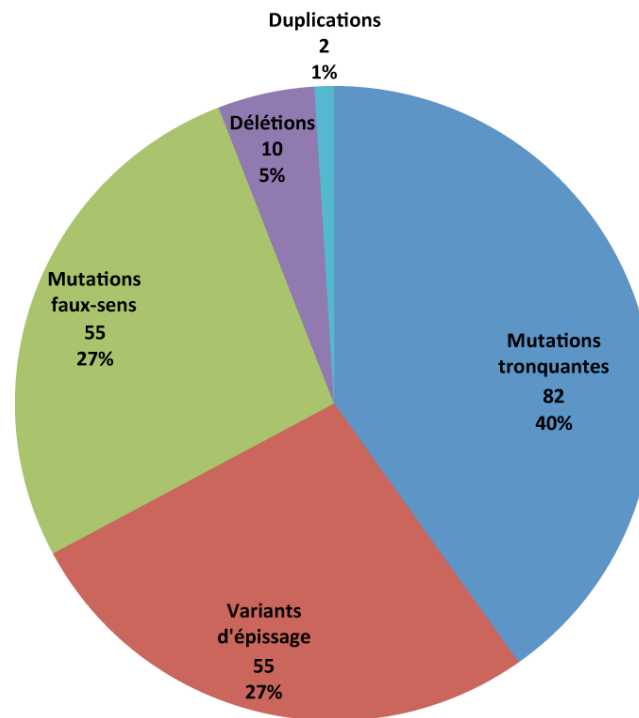


Figure 12. Données provenant de la base de données eOPA1 (juillet 2009) [179]. Répartition des 204 mutations pathogènes du gène *OPA1* en fonction de leur conséquence au niveau de la protéine.

Aucune corrélation génotype/phénotype n'a été observée dans les AOAD isolées puisque ni le type de mutation, ni sa localisation dans la protéine n'entraînent de résultats significatifs, même dans les grandes études cliniques [148, 151, 155]. Toutefois, comme mentionnée précédemment, une forte corrélation génotype/phénotype associe la mutation p.R445H avec l'AOAD et certaines mutations affectant le domaine GTPase avec des présentations cliniques d'« AOAD plus » ont été trouvées.

1.4.5 Fonction de la protéine Opa1

La protéine Opa1 est localisée dans la mitochondrie, spécifiquement dans l'espace intermembranaire (IMS pour *inter membrane space*) [181-183]. Des analyses par *western blot* d'Opa1 en conditions physiologiques ont montré cinq isoformes : deux isoformes longues (Opa1_L) ancrées à la membrane et trois isoformes courtes (Opa1_S) résultant d'un

processus de clivage protéolytique secondaire [184]. Lenaers *et coll.* suggèrent que Opa1 contribuerait aux structures de la membrane interne mitochondriale (IMM pour *inner mitochondrial membrane*), à savoir les crêtes, les jonctions des crêtes et les domaines d'interaction avec la membrane externe mitochondriale (OMM pour *outer mitochondrial membrane*) (Figure 13a) [185].

La protéine Opa1 joue un rôle essentiel dans la respiration mitochondriale. Elle serait nécessaire pour maintenir l'intégrité de la chaîne respiratoire [186-188]. La fonction centrale d'Opa1 consisterait donc à maintenir l'intégrité de l'IMM pour limiter les fuites de protons et faciliter le transport des électrons au cœur de cette membrane, entre les complexes de la chaîne respiratoire (figure 13b). Opa1 est également impliquée dans la fusion du réseau mitochondrial par l'intermédiaire de deux phénomènes : d'abord, la modification d'Opa1 perturberait le potentiel de membrane mitochondrial qui est nécessaire pour le processus de fusion (Figure 13b) [189] ; ensuite, Opa1 interagirait dans l'IMS avec les mitofusines, des protéines partenaires pro-fusion de l'OMM (Figure 13c) [190, 191]. Ainsi, les mutations affectant le gène *OPA1* pourraient soit dissiper le potentiel de membrane et envoyer un message pro-fusion, soit directement influencer la machinerie nécessaire pour l'entretien d'un réseau fusionné. En outre, des données récentes suggèrent qu'Opa1 aurait un rôle central dans le contrôle de l'apoptose. Certaines isoformes pourraient former un goulot d'étranglement à la jonction des crêtes et agir comme un « bouchon », séquestrant le cytochrome c dans le volume des crêtes (Figure 13d) [139, 192]. Les jonctions des crêtes sont la cible de protéines proapoptotiques, comme tBid [193], qui pourraient perturber certaines formes particulières d'Opa1, induisant un remodelage subtil des crêtes et la libération complète du cytochrome c [194, 195].

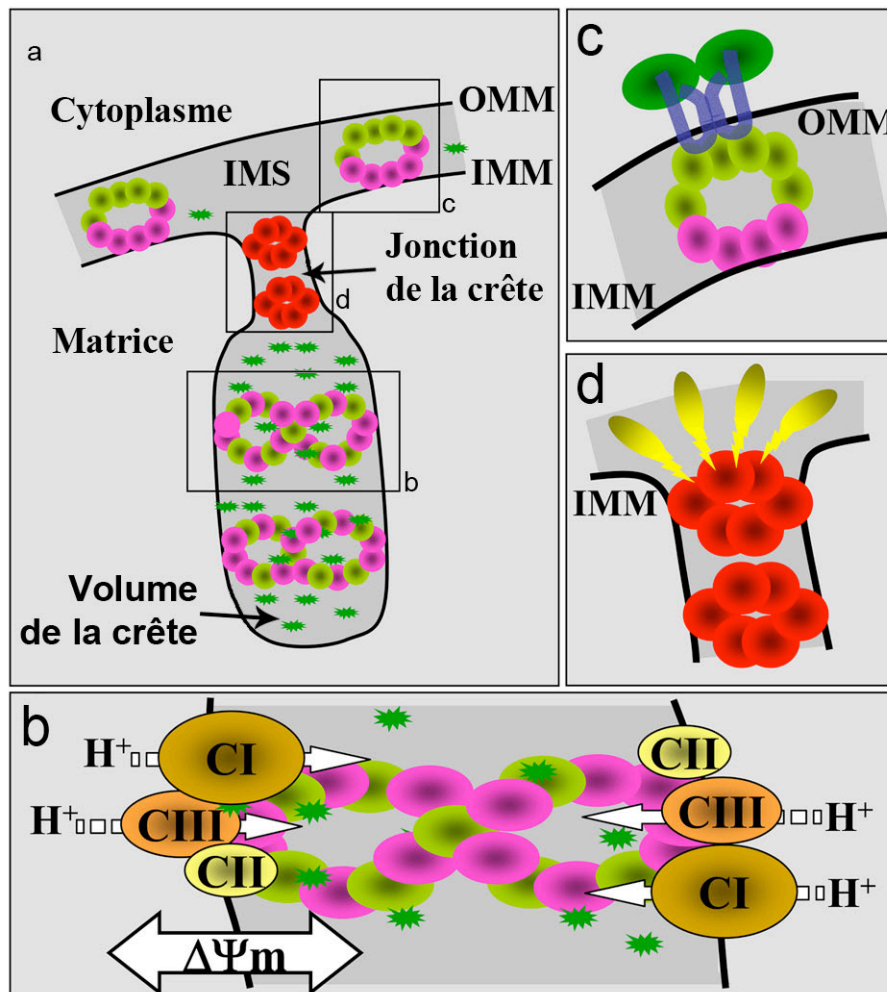


Figure 13. Structure et fonction d'Opa1 dans les mitochondries. (a) Représentation schématique des rôles d'Opa1 dans la structuration des crêtes, la jonction des crêtes et l'interaction entre les membranes externe (OMM) et interne (IMM) mitochondriales. Les points roses représentent les isoformes Opa1_L ancrées à la membrane, les points verts représentent les isoformes Opa1_S qui n'interagissent pas avec la membrane, les deux étant nécessaires pour la structure de crêtes et les ponts entre l'OMM et l'IMM. Les points rouges représentent les isoformes d'Opa1 faisant la jonction des crêtes. Le cytochrome c, représenté par des étoiles vertes, est le plus souvent ségrégué dans le volume des crêtes. (b) Les isoformes Opa1_L et Opa1_S, pouvant inclure l'exon 4, sont nécessaires pour les structures membranaires des crêtes et interagissent avec les complexes I (CI), II (CII) et III (CIII) de la chaîne (ellipses brun clair) pour promouvoir le transfert d'électrons dans l'IMM et maintenir le potentiel de membrane ($\Delta\Psi_m$). (c) Les isoformes OPA1_S (points verts) interagissent avec des protéines mitofusines (en vert et bleu dans l'OMM) pour promouvoir la fusion du réseau mitochondrial avec les isoformes Opa1_S (points verts) et éventuellement les isoformes Opa1_L (points violets). (d) Les isoformes antiapoptotiques d'Opa1 (points rouges) structurent la jonction de crêtes pour empêcher la mobilisation cytochrome c. Elles pourraient être ciblées par des protéines proapoptotiques (en jaune) pour démanteler la structure des jonctions des crêtes. D'après Lenaers et coll. (*Int J Biochem Cell Biol*, 2009) [185].

1.4.6 Physiopathologie

1.4.6.1 Haplo-insuffisance ou effet dominant négatif ?

La génétique des AOAD liées à *OPA1* est plus complexe et diversifiée que ce qui était attendu initialement. La prépondérance des mutations d'*OPA1* conduisant à une terminaison prématurée de la traduction suggère fortement que le mécanisme pathogène principal serait l'haplo-insuffisance. La description d'une grande délétion de 560–860 kb, révélée par hybridation *in situ* fluorescente (*fluorescence in situ hybridization*, FISH), entraînant la perte complète d'une copie du gène *OPA1*, a fourni une preuve supplémentaire que l'haplo-insuffisance est probablement responsable de l'AOAD [150]. Il a également été démontré que les transcrits d'*OPA1* contenant des codons de terminaison prématurés seraient instables et probablement dégradés par le mécanisme *nonsense-mediated mRNA decay* (NMD) [196], soutenant encore l'hypothèse de l'haplo-insuffisance. Cependant, toutes les formes multisystémiques de la maladie (AOADS et « AOAD plus ») sont liées à des mutations affectant le domaine GTPase [167], suggérant qu'un effet négatif dominant contribue à la pathogenèse de ces formes graves de la maladie. En outre, un patient atteint d'AOAD, hétérozygote composite de deux mutations dans l'exon 8 d'*OPA1*, est sévèrement touché par la maladie [151], alors que ses parents et ses frères et sœurs hétérozygotes sont moins gravement touchés, suggérant dans ce cas un mode de transmission semi-dominant.

1.4.6.2 Le déficit métabolique dans l'AOAD

Le mécanisme conduisant à la dégénérescence des CGR chez les patients atteints de mutations d'*OPA1* est encore largement inconnu. L'expression d'*OPA1* est ubiquitaire, ce qui est compatible avec sa localisation mitochondriale, et, même s'il est fortement exprimé dans les CGR, Opa1 n'est pas plus abondante dans ces cellules que dans les autres cellules rétinienne [197-200]. Les CGR ont des besoins en énergie particulièrement élevés, car la première partie de leurs axones, de la tête du nerf optique à la lame criblée de la sclère, est démyélinisée et particulièrement riche en mitochondries qui fournissent l'ATP nécessaire à la transmission des signaux visuels (Figure 14) [146]. Au-delà de la lame criblée, les axones myélinisés ont moins besoin d'énergie, car la présence de nœuds de Ranvier facilite la transmission de la dépolarisation, nécessitant donc un plus petit nombre de mitochondries dans cette région. Les CGR semblent être particulièrement vulnérables à la perte d'énergie dans la NOHL [146] et dans d'autres mitochondriopathies impliquant neuropathie optique et baisse d'activité de la chaîne respiratoire [201]. Compte tenu des expressions cliniques similaires de la NOHL et de l'AOAD, ainsi que la localisation d'Opa1 dans la membrane interne mitochondriale qui contient la chaîne respiratoire, il est probable que les mutations d'*OPA1* conduisent, directement ou indirectement, à une perte d'énergie. En effet, la

technique de spectroscopie du phosphore en résonance magnétique nucléaire a révélé une altération de la phosphorylation oxydative avec une diminution de production d'ATP dans les tissus musculaires de patients atteints d'AOAD liée à *OPA1* [158].

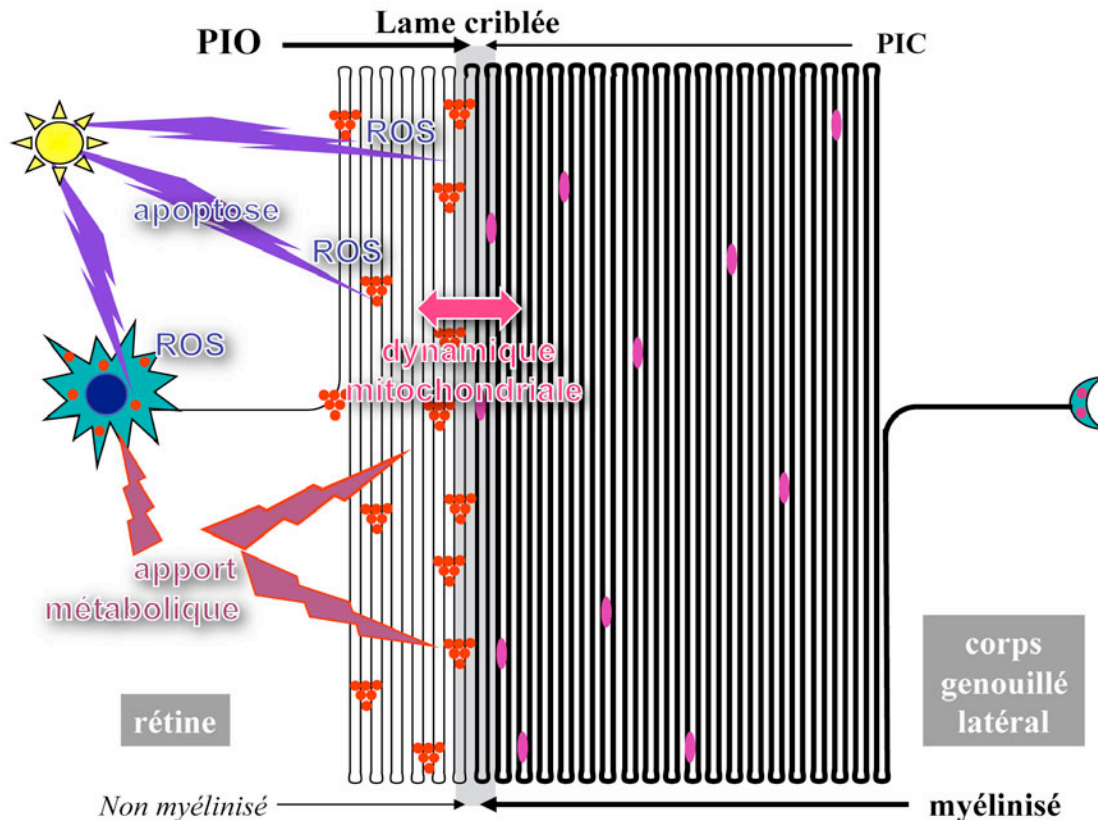


Figure 14. Représentation schématique d'un neurone de cellule ganglionnaire de la rétine (CGR) avec son réseau mitochondrial et des contraintes environnementales. Le soma de la CGR est localisé à gauche, dans la rétine, tandis que la synapse de la CGR est située dans le corps genouillé latéral. Entre les deux, l'axone est démyélinisé dans le globe oculaire, pour devenir myélinisé à partir de la lame criblée de la sclère. La région démyélinisée de l'axone et le soma subissent l'assaut de la lumière du jour (soleil jaune et étincelles violettes) et de la pression intraoculaire (PIO), alors que l'axone myélinisé est protégé de la lumière et des contraintes par la pression intracrânienne (PIC). Ce stress favoriserait la production de dérivés réactifs de l'oxygène (ROS) et l'apoptose. La surproduction de ROS augmenterait l'activité de fusion du réseau mitochondrial (double flèche rose). Pour fournir l'ATP nécessaire à la transmission des signaux visuels, les mitochondries (points rouges) ont des besoins en énergie particulièrement élevés (étincelles roses) dans la région démyélinisée de l'axone dans le globe oculaire. Le réseau mitochondrial est inégalement réparti, avec de nombreuses varicosités constituées de mitochondries agrégées (points rouges) dans la région démyélinisée de l'axone, particulièrement autour de la lame criblée de la sclère, alors que les mitochondries sont rares et allongées (ovales roses) dans la région myélinisée de l'axone dans le nerf optique. D'après une communication orale de Guy Lenaers (Colloque MeetOchondrie, La Grande-Motte, Hérault, 4 mai 2009).

Pour confirmer cette observation, notre laboratoire a utilisé des cultures primaires de fibroblastes de peau de patients pour explorer le métabolisme énergétique. Les fibroblastes, qui sont riches en mitochondries, sont particulièrement adaptés à l'exploration du métabolisme mitochondrial, car ces cellules sont recueillies aisément et sans douleur, et résistent à la transformation vers un métabolisme glycolytique non mitochondrial anaérobie. Les mesures de l'efficacité de la phosphorylation oxydative (*oxidative phosphorylation*, OXPHOS) ont révélé un défaut de couplage de celle-ci dans les fibroblastes de patients atteints de NOHL et d'AOAD liée à *OPA1* et *OPA3* [159]. Cependant, le mécanisme de découplage diffère selon le type de neuropathie optique puisque le déficit du complexe I constaté dans les fibroblastes de NOHL (diminution moyenne : 39 %) n'a pas été trouvé dans des fibroblastes *Opa1* qui, en revanche, montrent une réduction de 25 % de l'activité du complexe IV et une augmentation de 60 % de l'activité du complexe V. Fait intéressant, la production d'ATP n'est réduite que dans les fibroblastes de NOHL mais pas dans les fibroblastes d'AOAD isolées, suggérant un lien entre le niveau de déficit énergétique et la sévérité de l'expression clinique de la maladie. Cette relation est également confirmée par le fait que les NOHL et les formes « AOAD plus » ont un plus mauvais profil énergétique que les AOAD isolées. L'instabilité de l'ADNmt, au moins chez les patients atteints d'« AOAD plus », suggère un lien possible entre le déficit de l'OXPHOS et le dysfonctionnement d'*Opa1* [167, 168]. De plus, le contenu en ADNmt du sang serait plus faible chez les patients atteints d'AOAD, fournissant d'autres preuves liant *Opa1* avec l'ADNmt et l'OXPHOS [202].

Toutefois, l'existence et la nature du déficit énergétique des patients *Opa1* restent controversées, car d'autres auteurs explorant des cellules de la lignée lymphoblastique de patients atteints d'AOAD ne les ont pas retrouvées [203], ou dans des cellules musculaires, des fibroblastes ou des myotubes de patients atteints d'AOAD avec symptômes extraophtalmologiques [170]. Ainsi, une réduction significative de la synthèse d'ATP par les substrats du complexe I a été identifiée dans des fibroblastes de patients atteints d'AOAD [188]. Toutefois, les présentations cliniques et biochimiques des patients *Opa1* syndromiques ont de nombreuses similitudes avec celles observées dans les maladies de l'OXPHOS, renforçant ainsi la notion d'un défaut de la conversion énergétique dans ces troubles.

1.4.6.3 La dynamique du réseau mitochondrial dans l'AOAD

Contrairement au défaut énergétique partagé par l'ensemble des lignées cellulaires fibroblastiques étudiées, aucune altération du réseau mitochondrial commune aux différentes formes de neuropathies optiques n'a été mise en évidence [159]. Les fibroblastes des patients atteints de NOHL présentent un réseau de mitochondries fortement interconnectées, alors que les fibroblastes de patients atteints d'AOAD possèdent, en comparaison avec les

témoins, un réseau mitochondrial normal à très fragmenté. Selon l'expérience de notre laboratoire, les patients Opa1 présentant les plus sévères phénotypes sont ceux qui possèdent le réseau mitochondrial le plus fragmenté, suggérant un lien entre l'expression clinique, le niveau du déficit de l'OXPHOS et l'altération de la structure des mitochondries. Des modifications de la structure mitochondriale ont été fréquemment signalées dans les fibroblastes [165, 167, 187, 188], les myotubes [170], le muscle squelettique de patients Opa1 [167] ainsi que dans des modèles de souris Opa1 [204, 205]. Des exemples de fragmentation du réseau mitochondrial observés dans les fibroblastes de patients Opa1 sont illustrés dans la Figure 15. Ces anomalies morphologiques sont clairement liées à la fonction d'Opa1, identifiée pour la première fois par analogie avec les orthologues de *Saccharomyces cerevisiae* et *Schizosaccharomyces pombe*, en agissant comme une mécano-enzyme sur la membrane mitochondriale interne, favorisant la fusion mitochondriale et maintenant la structure des crêtes [186].

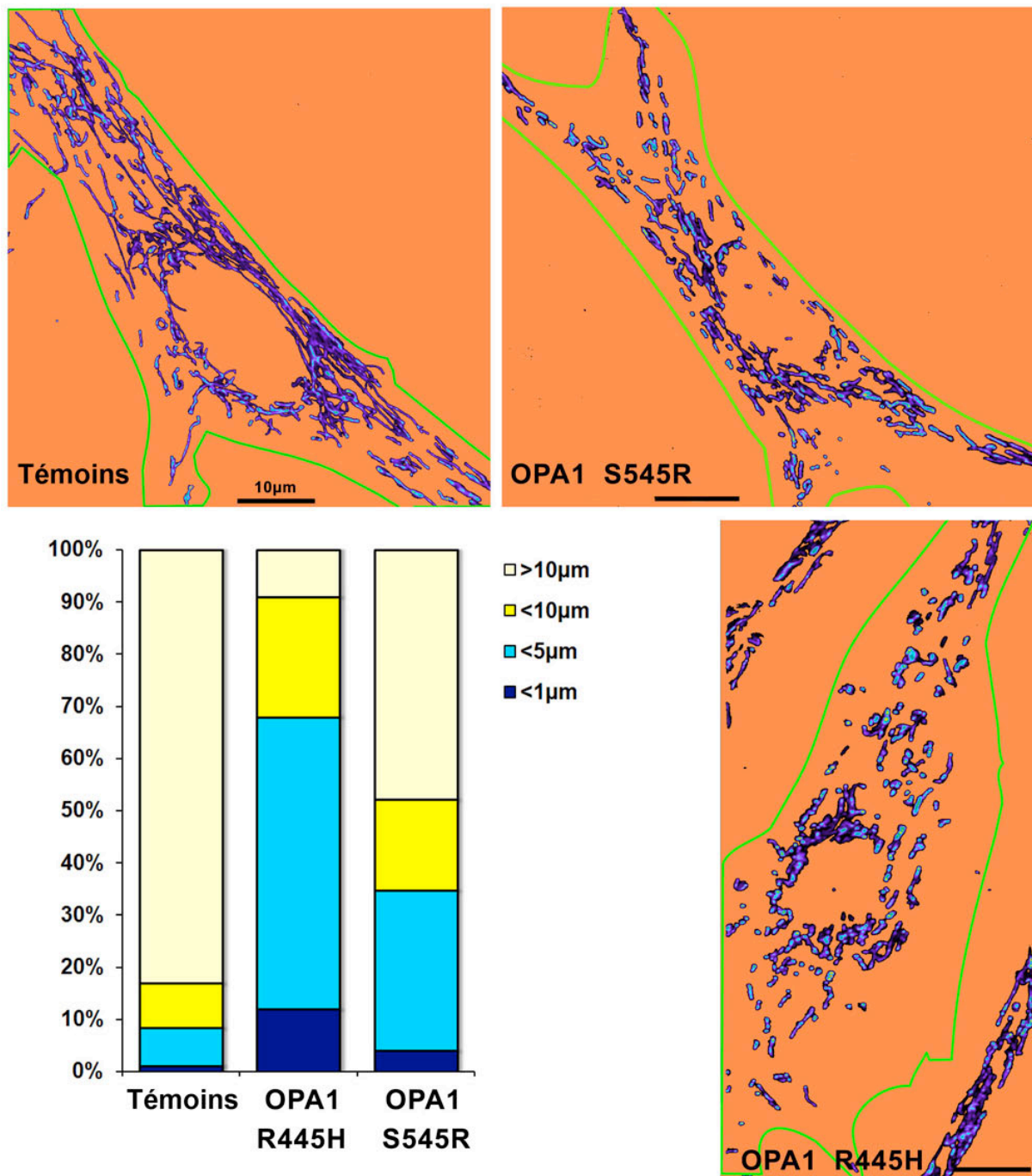


Figure 15. Fragmentation complète ou partielle du réseau mitochondrial dans des cellules mutées Opa1 par rapport aux témoins. Les mitochondries ont été marquées avec la sonde fluorescente Mitotracker®. Vingt et un plans-images ont été acquis le long de l'axe Z par incréments de 0,3 µm avec un microscope Leica DMI6000B (Microsystems GmbH, Wetzlar, Allemagne) et un appareil d'acquisition Roper CoolSnap HQ2 (Roper Scientific GmbH, Allemagne). Les piles d'images ont été déconvoluées à l'aide d'un algorithme de Meinel et du logiciel MetaMorph (Molecular Devices, CA, USA). Le signal fluorescent mitochondrial a été seuillé et l'image binarisée. La longueur des mitochondries et leur nombre ont été déterminés à l'aide d'une analyse morphométrique intégrée des régions créées autour des mitochondries. Cette séquence d'analyses a été utilisée pour créer l'histogramme. Les images en contraste de

phase ont été utilisées pour déterminer les contours de cellules (trait vert). Les images montrent le réseau mitochondrial (en pseudocouleurs) de deux fibroblastes de peau comportant des mutations d'*OPA1* par rapport à un témoin.

1.4.6.4 Apoptose dans l'AOAD

Le fait que le dysfonctionnement d'Opa1 augmente la sensibilité à l'apoptose induite par la staurosporine ajoute une hypothèse concernant la physiopathologie de l'AOAD. Le rôle crucial d'Opa1 dans la régulation de l'apoptose a été ensuite confirmé par des études montrant qu'Opa1 contrôlerait le remodelage des crêtes mitochondriales apoptotiques et que la protéase Parl régulerait la libération du cytochrome c par l'intermédiaire des protéines Opa1 [187, 194, 206]. De plus, une sensibilité accrue à l'apoptose a été systématiquement observée dans les fibroblastes de tous les patients porteurs de mutations d'*OPA1* [187]. Ainsi, il apparaît que les CGR sont particulièrement sensibles à la mort cellulaire et que le dysfonctionnement et l'expression réduite d'*OPA1* exacerbent la dégénérescence de ces cellules.

1.4.7 Modèles animaux

Il existe désormais deux modèles de souris AOAD, avec des mutations hétérozygotes dans l'exon 8 (c.1051C>T) et l'intron 10 (c.1065+5G>A) du gène *OPA1* [204, 205]. Ces deux mutations sont tronquantes, réduisant de 50 % l'expression de la protéine Opa1 et, par conséquent, représentant un mécanisme d'haplo-insuffisance de la maladie. Dans les deux modèles, les souris mutantes homozygotes (*OPA1*^{-/-}) sont mortes *in utero* au cours de l'embryogenèse, soulignant le rôle central joué par la protéine Opa1 au début du développement. Les souris hétérozygotes *OPA1*^{+/-} reproduisent fidèlement le phénotype humain en présentant une neuropathie optique progressive lente et en montrant une réduction de leur fonction visuelle sur les tests psychophysiques. Il y a une perte progressive des CGR, conduisant à l'amincissement de la couche des fibres nerveuses rétiniennes, et les axones du nerf optique indemnes ont une morphologie anormale avec des gonflements et des déformations, des zones irrégulières de démyélinisation et des agrégats de myéline. Les mitochondries à l'intérieur de ces axones présentent, en microscopie électronique en transmission (MET), des structures de crêtes désorganisées et les fibroblastes en culture montrent une fragmentation du réseau mitochondrial. Ces deux modèles de souris *OPA1* sont de puissants moyens d'étudier les voies menant à la perte préférentielle des CGR dans l'AOAD, en permettant d'effectuer des études fonctionnelles directement sur ces cellules spécialisées, ce qui n'est pas possible chez les humains étant donné l'indisponibilité de tissus oculaires. Ces souris mutantes pourraient également être utiles pour étudier le

potentiel thérapeutique de molécules biologiques à venir qui, en étant injectées dans le corps vitré, permettraient un accès direct à la couche des CGR.

Un autre développement récent concerne un mutant drosophile homozygote touché spécifiquement à l'œil, qui exprime aussi un phénotype lié à une mort cellulaire anormale et des niveaux élevés de ROS dans l'œil des mouches adultes [207]. Fait intéressant, le phénotype de l'œil de la drosophile a été partiellement annulé par un traitement antioxydant, ce qui suggère que les ROS pourraient jouer un rôle déterminant dans la pathogenèse de l'atrophie optique.

1.4.8 Conseil génétique

L'AOAD est transmise selon un mode autosomique dominant de sorte que chaque enfant d'une personne portant une mutation d'*OPA1* possède un risque de 50 % d'hériter de l'allèle mutant. Toutefois, le conseil génétique est compliqué par la pénétrance incomplète et la variabilité de l'expression inter- et intrafamiliale de la maladie. En outre, à l'exception des AOAD avec des manifestations extraoculaires, les corrélations génotype-phénotype sont très rares. Par conséquent, la probabilité qu'un porteur d'une mutation d'*OPA1* soit touché par de graves déficiences visuelles est presque impossible à évaluer. Une autre difficulté se pose pour les généticiens cliniques confrontés à des cas apparemment isolés d'AOAD dans des familles : l'histoire d'une famille peut sembler négative à cause des difficultés à reconnaître une atrophie optique parmi ses membres, compte tenu de l'expressivité variable de la maladie. C'est pourquoi il est impossible d'affirmer qu'une mutation est *de novo* sans une étude moléculaire de la famille.

1.4.9 Perspectives d'avenir

Le gène *OPA1* joue un rôle important dans la fusion des mitochondries, dans l'efficacité du métabolisme énergétique, dans l'apoptose et dans la maintenance de l'ADNmt. Les mutations de ce gène sont responsables d'un large spectre de pathologies allant de formes cliniques simples et modérées jusqu'à des atteintes sévères et multisystémiques probablement en relation avec la gravité du déficit énergétique. Il reste à mieux comprendre le mécanisme de perturbation des mitochondries afin d'essayer d'identifier de nouvelles stratégies thérapeutiques pharmacologiques pour ces maladies mitochondriales. La description récente de modèles murins de l'AOAD liée à *OPA1* présentant une perte progressive des cellules ganglionnaires de la rétine devrait permettre le développement de cette recherche physiopathologique et thérapeutique. De manière plus générale, les pathologies liées à *Opa1* font partie d'une nouvelle classe de maladies dues à une perturbation de la plasticité mitochondriale dont l'expression est principalement neurologique. Ainsi les protéines de *mitofusin 2* (*Mfn2*, MIM *608507) et *dynamine-like protein*

1 (Dlp1, MIM +603850), acteurs clefs de la fusion et de la fission mitochondriale, sont respectivement associées à la maladie de Charcot-Marie-Tooth de type 2A (MIM #609260) et à un syndrome multisystémique complexe [208]. Des études récentes ont aussi démontré qu'une perturbation de la dynamique mitochondriale participait aussi probablement à la physiopathologie de la maladie d'Alzheimer et des formes héréditaires de la maladie de Parkinson. Dans ce contexte, les pathologies associées à *OPA1* constituent un excellent modèle de cette nouvelle classe de maladies neurodégénératives.

CHAPITRE 2 : OBJECTIF DE LA THÈSE

Ce projet de thèse a été réalisé dans la continuité des travaux qui avaient été engagés dans notre laboratoire, au début des années 2000, par Yves Tourmen *et coll.* sur l'étude de l'origine évolutive et phylogénétique du protéome mitochondrial humain. Les auteurs avaient montré que le transfert de gènes mitochondriaux vers le génome nucléaire est un phénomène fréquent et encore en cours par l'identification et la datation de plus de 300 pseudogènes mitochondriaux dans le génome humain [10]. En outre, la littérature scientifique suggérait que le protéome mitochondrial (plus d'un millier de protéines) serait composé de protéines dérivant de l'endosymbionte protomitochondrial procaryote et également d'autres protéines qui auraient été intégrées plus récemment au cours de l'évolution, à partir des génomes des hôtes eucaryotes. L'étude de la double origine évolutive du protéome mitochondrial paraissait particulièrement intéressante dans la perspective d'applications médicales.

Le projet de thèse que nous avons développé consistait donc en l'étude de l'origine évolutive du protéome mitochondrial grâce aux techniques bio-informatiques afin d'exploiter et organiser les masses de données qui devenaient disponibles. Pour répondre à cette problématique, il fallait, d'une part, adopter une approche fondamentale d'analyse de l'origine évolutive des protéines mitochondriales et, d'autre part, étudier leur versant pathologique, notamment en se focalisant sur l'origine évolutive des protéines mitochondriales mises en cause dans les pathologies.

Parallèlement à cette analyse globale du protéome mitochondrial, une étude spécialisée devait également être menée. Un des principaux sujets d'étude du laboratoire est l'exploration des mécanismes physiopathologiques mis en jeu dans les neuropathies héréditaires associées aux mutations de gènes impliqués dans la dynamique mitochondriale et, en particulier, à la protéine mitochondriale Opa1 dont les mutations sont la principale cause génétique d'AOAD. Il fallait donc mettre en place une base de données internationale dédiée à l'étude des mutations du gène *OPA1* dans le cadre des activités de diagnostic et de recherche du laboratoire afin de conduire une étude génétique portant sur les patients porteurs d'une atrophie optique.

CHAPITRE 3 : TRAVAIL DE THÈSE

3.1 ANALYSE DU PROTÉOME MITOCHONDRIAL HUMAIN À TRAVERS L'ÉVOLUTION

Dans la première partie de ce travail de thèse, nous avons étudié le protéome mitochondrial et son origine évolutive. Cette partie comporte un article publié et une étude complémentaire non publiée :

1. ***Mitochondrial diseases preferentially involve proteins with prokaryote homologues.***

Tourmen Y., Ferre M., Malthiery Y., Dessen P., and Reynier P., C R Biol, 2004. **327**(12): p. 1095-101.

2. **Méthode de prédiction à haut débit de nouvelles protéines mitochondriales par génomique comparative.** Données non publiées.

La disponibilité des séquences génomiques d'un nombre croissant d'organismes a contribué à relancer la recherche sur l'origine et l'évolution de la mitochondrie moderne. Selon la théorie endosymbiotique, les mitochondries possèderaient une origine monophylétique unique. Une cellule eucaryote primitive (ou une archéobactérie) aurait intégré un endosymbionte procaryote il y environ 1,5 à 2 milliards d'années, lorsque que l'atmosphère primitive s'est enrichie en oxygène [2, 5]. Les études phylogénétiques indiquent que cet endosymbionte est apparenté aux α -protéobactéries, le plus proche parent de la mitochondrie connu actuellement étant *Rickettsia prowazekii*, une bactérie intracellulaire obligatoire [2]. Au cours de l'évolution, la majorité des gènes de l'endosymbionte originel aurait été perdue ou bien transférée vers le noyau de la cellule eucaryote hôte [5, 9]. En effet, les nombreux pseudogènes mitochondriaux présents dans le génome attestent d'un processus de transfert tout au long de l'évolution [10, 11].

La taille du protéome mitochondrial humain est estimée à plus d'un millier de protéines. Seules 13 protéines sont codées par l'ADNmt, vestige du génome de l'endosymbionte. Toutes les autres protéines sont codées par le génome nucléaire. Chez la levure 50–60 % des protéines mitochondriales ont des homologues chez les procaryotes alors que 40–50 % n'en ont pas [5]. Les protéines mitochondriales possédant un homologue procaryote résultent probablement du transfert des gènes de l'endosymbionte vers le noyau, tandis que les protéines mitochondriales non homologues a des protéines procaryotes résultent d'un

phénomène « d'enrichissement » du protéome mitochondrial par de nouvelles protéines et donc de nouvelles fonctions.

Dans un premier article, nous avons analysé *in silico* les 393 protéines mitochondriales humaines répertoriées (à l'époque de l'étude) dans Swiss-Prot en vue de déterminer leur affiliation procaryote ou non (origine eucaryote présumée). Nous avons comparé la taille, la localisation mitochondriale et la fonction des protéines selon leur origine. Enfin, nous avons examiné l'implication de ces protéines dans les maladies mitochondriales (voir Premier article : « Les mitochondriopathies impliquent majoritairement des protéines d'origine procaryote », p. 74).

En particulier, dans cette précédente étude, nous avons montré que 88 % des protéines mitochondriales humaines possédant un homologue procaryote sont significativement plus grandes que dans ce dernier. Cette différence s'explique principalement par la présence d'une extension N-terminale. Ces séquences additionnelles ne présentent pas d'homologie avec des protéines procaryotes ; elles seraient probablement d'origine eucaryote. La majorité des protéines homologues à des protéines procaryotes serait donc des protéines chimères eucaryotes-procaryotes. Étant donné que peu de protéines mitochondriales sont actuellement connues (Tableau 6), la majorité de celles-ci reste à identifier.

Le fait que les protéines mitochondriales d'origine procaryote possèdent dans leur grande majorité une séquence additionnelle N-terminale (par rapport à leur ancêtre procaryote) nous procure un mode de criblage potentiel de nouvelles protéines mitochondriales non identifiées à ce jour. Afin de confirmer cette hypothèse, nous proposons de comparer l'ensemble des protéines procaryotes connues contre le génome humain et sélectionner celles qui possèdent cette extension N-terminale particulière aux protéines mitochondriales. Nous avons réalisé un test sur le chromosome 19 qui permet d'évaluer la capacité de cette démarche à identifier les protéines mitochondriales déjà connues (contrôle positif). (voir Étude complémentaire non publiée : méthode de prédiction à haut débit de nouvelles protéines mitochondriales par génomique comparative, p. 83)

3.1.1 Premier article : « Les mitochondriopathies impliquent majoritairement des protéines d'origine procaryote »



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

C. R. Biologies 327 (2004) 1095–1101



<http://france.elsevier.com/direct/CRASS3/>

Genetics / Génétique

Mitochondrial diseases preferentially involve proteins with prokaryote homologues

Yves Tourmen^{a,*}, Marc Ferré^a, Yves Malthièry^a, Philippe Dessen^b, Pascal Reynier^a

^a Laboratoire de biochimie et biologie moléculaire, INSERM E0018, CHU d'Angers, 4, rue Larrey, 49033 Angers, France

^b Laboratoire de génétique oncologique, UMR 8125 CNRS, Institut Gustave-Roussy, 94805 Villejuif, France

Received 5 July 2004; accepted 8 September 2004

Presented by Jean Rosa

Abstract

The comparison of each of the 393 nuclear-encoded human mitochondrial proteins annotated in the SwissProt databank with 256,953 proteins from 94 prokaryote species showed that two thirds of the mitochondrial proteome were homologous with prokaryotic proteins, whereas one third was not. Prokaryotic mitochondrial proteins differ markedly from eukaryotic proteins, particularly in regard to their size, localization, function, and mitochondrial-targeting N-terminal sequence. Remarkably, the majority of nuclear genes implicated in respiratory chain mitochondrial diseases were found to be of prokaryotic ancestry. Our study indicates that the investigation of the co-evolution of eukaryotic and prokaryotic mitochondrial proteins should lead to a better understanding of mitochondrial diseases. *To cite this article: Y. Tourmen et al., C. R. Biologies 327 (2004).*

© 2004 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Résumé

Les mitochondriopathies impliquent majoritairement des protéines d'origine procaryote. Nous avons comparé les séquences des 393 protéines mitochondriales humaines répertoriées dans la banque de données SwissProt avec celles de 256 953 protéines procaryotes. Seules 64% des protéines mitochondriales sont homologues à des protéines procaryotes, ce qui témoigne de leur double origine évolutive procaryote et eucaryote. La structure, la localisation et la fonction des protéines mitochondriales diffèrent fortement selon leur origine. De plus, les protéines impliquées dans les pathologies de la chaîne respiratoire sont majoritairement homologues à des protéines procaryotes. Ces résultats montrent que l'étude de la double origine évolutive du protéome mitochondrial pourrait contribuer à la compréhension de la physiopathologie mitochondriale. *Pour citer cet article : Y. Tourmen et al., C. R. Biologies 327 (2004).*

© 2004 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Keywords: evolution; mitochondria; mitochondrial diseases; proteome

Mots-clés : évolution ; mitochondries ; maladies mitochondriales ; protéome

* Corresponding author.

E-mail address: yvtourmen@chu-angers.fr (Y. Tourmen).

1. Introduction

The origin and evolution of modern mitochondria are matters of considerable biological interest. According to the endosymbiotic theory, mitochondria have a unique origin, arising from a symbiont closely related to an ancestral alpha-proteobacteria. This endosymbiont is supposed to have lived in a nucleus-containing host cell – either an amitochondriate eukaryote or an Archea-related cell – between 1.5 and 2 billion years ago when the oxygen content of the atmosphere started to increase [1–7]. Indeed, the obligate intracellular symbiont *Rickettsia prowazekii*, which belongs to the alpha-proteobacteria group, was previously described as the closest known eubacterial relative of mitochondria identified by the phylogenetic analysis of mitochondrial DNA encoded genes [4]. During the course of evolution, most of the endosymbiotic genes may have been lost or transferred to the nucleus of the eukaryotic host cell [5–9]. Numerous mitochondrial pseudogenes currently present in the modern human nuclear genome attest to the massive and ongoing transfer process of mitochondrial genes to the nucleus in the course of evolution [10–12]. It is believed that because of these complex evolutionary processes, only 13 protein-encoding genes may have persisted in the modern human mitochondrial DNA.

The human mitochondrial proteome is believed to contain over a thousand proteins [13]. However, the size of the mitochondrial proteome differs markedly among species [14]. Recent studies have demonstrated the dual origin of the mitochondrial proteome in yeast; 50–60% of the mitochondrial proteins have homologues in prokaryotic species, whereas 40–50% do not [7,15,16]. The current hypothesis is that these proteins may have been recruited from pre-existing nuclear genes and targeted toward mitochondria [4,15]. The large number of eukaryote-derived genes suggests that numerous specific functions of modern mitochondria were absent in prokaryotic mitochondrial ancestors. Thus, the modern mitochondrial proteome is composed of proteins with a dual eukaryotic and prokaryotic origin, the eubacterial proteins being encoded by two genomes, nuclear, and mitochondrial [17].

We have analysed *in silico* all the 393 human mitochondrial proteins annotated in SwissProt in order to determine their prokaryotic or eukaryotic affiliation. We have compared the size, the mitochondrial local-

ization, and the function of the proteins according to their origin. Finally, we have considered the implication of these proteins in mitochondrial diseases.

2. Material and methods

2.1. Data selection

A search of the SwissProt database (<http://www.ebi.ac.uk/swissprot/>) yielded data on 393 human mitochondrial proteins. We classified these proteins according to their mitochondrial function and localization. A search of the NCBI database covering 94 prokaryote species produced data on 256,953 proteins.

2.2. BLAST analysis

The sequences of each of the 393 human mitochondrial proteins were compared with those of the 256,953 prokaryotic proteins using the BLASTP program [18]. We used $E < 1 \times 10^{-4}$ as the criterion to avoid false positives in establishing similarity.

2.3. Mitochondrial N-terminal targeting prediction

To test the proteins listed without a transit peptide in the SwissProt databank, we used two independent N-terminal mitochondrial-targeting predicting methods: Mitoprot (<http://ihg.gsf.de/ihg/mitoprot.html>) [19], and TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) [20]. The putative existence of the N-terminal transit peptide was admitted only when the results of both tests were positive.

2.4. Statistical analysis

Statistical analysis was performed using the Mann–Whitney and Kolmogorov–Smirnov tests. Differences were considered significant at $p < 0.05$.

3. Results

The BLASTP comparison of the 393 human proteins annotated as mitochondrial proteins in the SwissProt database with 256,953 proteins from 94 prokaryotic species allowed us to identify two distinct groups

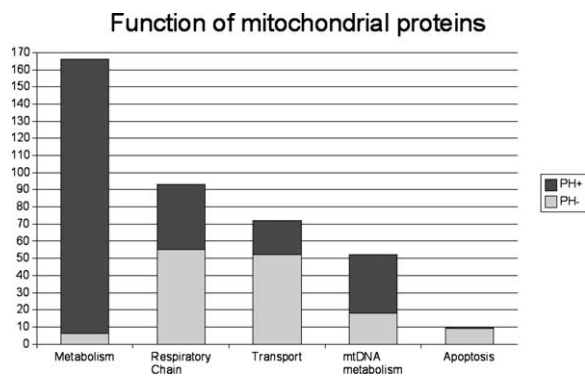


Fig. 1. Functional categories of the 393 known human mitochondrial proteins and functions of the proteins according to their eukaryotic (PH⁻) or prokaryotic (PH⁺) origin.

of human mitochondrial proteins. The first group of proteins with prokaryotic homologues, noted PH⁺, comprising 253 out of the 393 proteins (64%), included human mitochondrial proteins displaying a high score of homology with one or more prokaryotic proteins ($E < 1 \times 10^{-4}$). A more drastic cut-off value ($E < 1 \times 10^{-10}$) did not greatly affect this result, since the number of mitochondrial proteins with prokaryotic homologues dropped from 253 to 233. The second group of proteins without prokaryotic homologues, noted PH⁻, contained the remaining 140 proteins (36%) that were non-homologous with any of the known prokaryotic proteins.

The proportion of PH⁺ and PH⁻ mitochondrial proteins involved in the main mitochondrial functions varied considerably (Fig. 1). Remarkably, 160 out of the 166 proteins involved in metabolism belonged to the PH⁺ group, whereas nine out of the 10 mitochondrial proteins involved in apoptosis belonged to the PH⁻ group. The other functional classes were associated with both the PH⁺ and the PH⁻ groups of proteins, attesting to their dual evolutionary origin. The metabolite carriers were evenly divided in the two groups: 12 were PH⁺ and 12 were PH⁻, whereas 12 out of the 13 ion carriers and 15 out of the 20 protein carriers belonged to the PH⁻ group. As expected, a majority of proteins involved in mitochondrial DNA maintenance and expression, i.e. 34 out of the 52 proteins, or 65%, were of prokaryotic origin. Most of the matrix proteins, i.e. 175 out of the 200 proteins, or 88%, were PH⁺ (Fig. 2). The majority of the inner membrane proteins, i.e. 92 out of the 153 proteins, or 60%, was

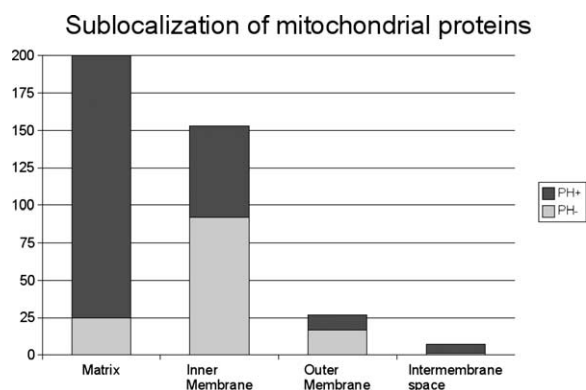


Fig. 2. Distribution of the 393 known human mitochondrial proteins according to their localization and sublocalization of the proteins according to their eukaryotic or prokaryotic origin.

PH⁻, as was the majority of the outer membrane proteins, i.e. 17 out of the 27 proteins, or 63%. Out of the seven proteins in the intermembrane space, six were PH⁺ and only one was PH⁻.

Fig. 3 shows the composition of the respiratory chain. Out of the 73 respiratory chain proteins annotated in the SwissProt database, only 25 were PH⁺. This analysis confirmed the dual origin of complex I subunits that contain 18 proteins known to originate from eubacterial ancestors. These include 7 mitochondrial DNA-encoded proteins and 11 nuclear DNA-encoded proteins. The other 23 proteins were contributed by eukaryotic ancestors. Furthermore, the study of complex IV (cytochrome c oxidase, COX) revealed that all the subunits originated from eukaryotic ancestors, except for the three mitochondrial DNA-encoded subunits known to originate from prokaryotic ancestors.

The comparison of the size of the mitochondrial proteins according to their origin revealed that the proteins of prokaryotic origin were significantly larger ($p < 0.0001$). The PH⁺ proteins were made up on average of 435 amino acids (range: 69–1500), whereas the PH⁻ proteins had an average size of 238 amino acids (range: 50–1816). As Fig. 4 shows, this difference was mainly due to the larger size of the PH⁺ respiratory chain components and transport proteins.

We found that 223 out of the 253 human mitochondrial PH⁺ proteins (88%) were significantly larger (average: 435 amino acids, range: 69–1500) than their prokaryotic homologues (average: 400 amino acids, range: 38–1353) ($p < 0.0001$). This difference was

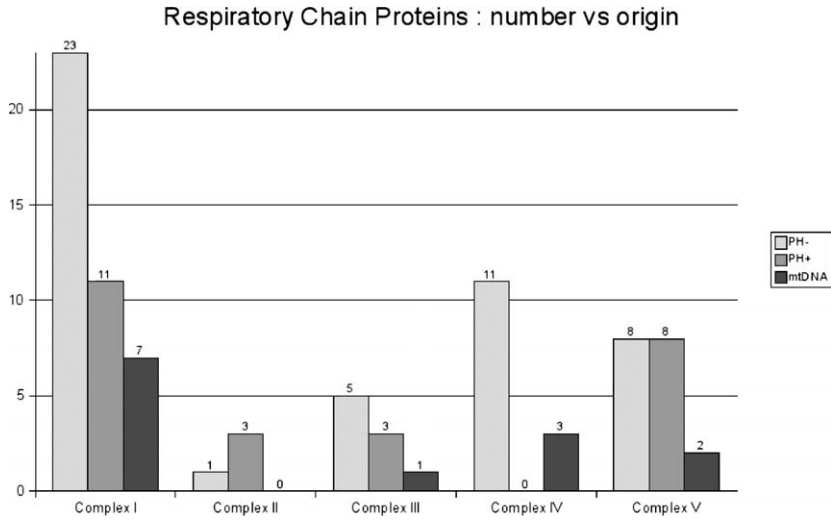


Fig. 3. Seventy-three nuclear-encoded human mitochondrial proteins of the respiratory chain according to their eukaryotic or prokaryotic origin. The 13 proteins of prokaryote origin encoded by mitochondrial DNA (mtDNA) are also shown.

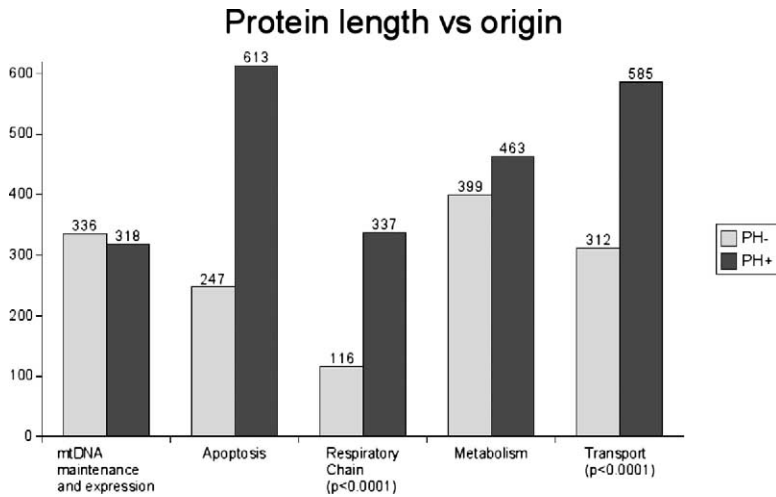


Fig. 4. Average protein length of human mitochondrial proteins according to their function and origin.

mainly due to the presence of a supplementary N-terminal sequence. These additional sequences displayed no homology with prokaryotic proteins and were probably of eukaryotic origin. Thus, most of the PH+ proteins are in fact probably prokaryotic-eukaryotic chimeral proteins. The use of TargetP and Mitoprot software showed that 213 out of the 253 mitochondrial PH+ proteins (84%) either had, or were predicted to have, a mitochondrial-specific N-terminal targeting sequence. In contrast, only 54 out of the 140 mito-

chondrial PH- proteins (39%) possessed this targeting sequence. A possible explanation is that proteins of eukaryotic origin may have mainly contributed to the membrane compartment in which proteins frequently lack the typical N-terminal targeting sequence.

Lastly, we examined the nuclear-encoded proteins that have been previously linked to human mitochondrial pathology in terms of their evolutionary origin (Table 1). Strikingly, 18 out of the 20 mitochondrial proteins known to be involved in diseases associated

Table 1
Origin of nuclear genes associated with human mitochondrial diseases [29–32]

Gene	OMIM	Prokaryotic homologue
Respiratory chain and Krebs cycle		
NDUFS1 (complex I)	157 655	+
NDUFS2 (complex I)	602 985	+
NDUFS3 (complex I)	603 846	+
NDUFS4 (complex I)	602 694	+
NDUFS7 (complex I)	601 825	+
NDUFS8 (complex I)	602 141	+
NDUFV1 (complex I)	161 015	+
NDUFV2 (complex I)	600 532	+
SDHA (complex II and Krebs cycle)	600 857	+
SDHB (complex II and Krebs cycle)	115 310	+
SDHC (complex II and Krebs cycle)	605 373	+
SDHD (complex II and Krebs cycle)	168 000	–
HUMQPC (complex III)	191 330	–
BCS1L (complex III assembly factor)	603 647	+
SURF1 (complex IV assembly factor)	185 620	+
SCO1 (complex IV assembly factor)	603 644	+
SCO2 (complex IV assembly factor)	604 377	+
COX10 (complex IV assembly factor)	602 125	+
COX15 (complex IV assembly factor)	603 646	+
FH (Krebs cycle)	150 800	+
MtDNA maintenance and expression		
TP (thymidine phosphorylase)	603 041	+
DGUOK (deoxyguanosine kinase)	251 880	+
TWINKLE (DNA helicase)	157 640	+
TK2 (thymidine kinase 2)	251 880	+
DNC (deoxynucleotide carrier)	607 196	–
ANT1 (adenine nucleotide translocator 1)	157 640	–
POLG1 (DNA polymerase γ)	157 640	–
Other mitochondrial functions		
SPG7 (paraplegin)	607 259	+
FRDA (frataxin)	229 300	+
ABC7 (ABC transporter)	301 310	+
DDP1 (deafness-dystonia protein)	304 700	–
OPA1 (optic atrophy 1)	165 500	–

OMIM, Online Mendelian Inheritance in Man (<http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=OMIM>).

with the respiratory chain and the Krebs cycle belonged to the PH+ group, whereas only two belonged to the PH– group ($p < 0.0001$).

4. Discussion

We found that 64% of the proteins of the human mitochondrial proteome had prokaryote homologues, whereas 36% of the proteins were non-homologous

with existing prokaryotic proteins. Mitochondrial proteins involved in energetic metabolism, biosynthetic metabolism, and mitochondrial DNA maintenance and expression were mainly of prokaryotic origin, while those involved in the transport and control functions originated from eukaryotes. The majority of the matrix proteins originated from prokaryotic ancestors, whereas the proteins of the inner and outer membrane compartments were of eukaryotic origin. Most of the eukaryotic proteins were probably targeted to the en-

dosymbiont in order to develop communication with the host cell (protein, ion and metabolite transport), to regulate ancestral mitochondrial functions (such as ATP production and regulation of membrane potential) and to establish the more recent biological functions (such as apoptosis or androgen synthesis).

Mitochondrial proteins of eubacterial origin are significantly larger than proteins of eukaryotic origin. The difference is particularly striking in the case of proteins involved in membrane transport and the respiratory chain. We hypothesise that the endosymbiont contributed the genes of the large core enzymes of the respiratory chain complexes and that the genes of the more recently evolved accessory proteins of smaller size came from the nuclear genome of the eukaryotic host. This hypothesis is supported by the fact that the three mitochondrial-encoded subunits of complex IV (cytochrome c oxidase) are known to be the largest proteins of the complex. It has been shown in *Paracoccus denitrificans* that two of these subunits were sufficient to ensure oxygen reduction in water and proton transport [21]. Thus, the other nuclear-encoded subunits (originating from eukaryotic hosts), which may not be necessary to ensure the basic catalytic function of complex IV, might be implicated in the assembly of the complex or the modulation and stabilization of its activity.

Interestingly, we found that 89% of the proteins encoded by nuclear genes involved in respiratory chain deficiencies have prokaryotic homologues. This is an unexpected result, since 59% of the respiratory chain subunits and assembly factors are of eukaryotic origin (38 of the subunits belonging to the PH⁻ group and 55 to the PH⁺ group). In addition to these nuclear-encoded proteins, two other main categories of proteins of eubacterial origin are involved in human mitochondrial pathologies. The first consists of mitochondrial DNA-encoded proteins, which are also derived from alpha-proteobacteria. The second includes proteins implicated in several mitochondrial diseases associated with matrix metabolic enzyme deficiencies. Our study shows that these proteins are of prokaryotic origin in 96% of the cases. Taken together, these data indicate that the majority of human mitochondrial pathologies may involve proteins of eubacterial origin.

For instance, respiratory complex IV deficiency fits in with this hypothesis. The disorder has been attributed to mitochondrial DNA mutations (3 sub-

units) or to mutations in the nuclear genes such as *SURF1*, *Sco1*, *Sco2* and *COX10*, encoding mitochondrial assembly factors [22–26], all of which were found to have eubacterial homologues in the present study. However, several authors failed to detect any mutations [27,28] in the 10 nuclear-encoded structural subunits of complex IV in other respiratory chain complex deficiencies. Our observation that none of these nuclear-encoded proteins of complex IV has a prokaryotic homologue reinforces our hypothesis of the specific involvement of proteins of eubacterial origin in human mitochondrial pathologies.

Our study indicates that the investigation of the evolution of mitochondrial proteins should lead to a better understanding of mitochondrial diseases. It is in accordance with previous study that showed that 18/100 human mitochondrial proteins that were strongly conserved among eukaryotic species were directly associated with human disease [14]. It should be emphasized that our study bears on the 393 known human mitochondrial proteins, whereas the mitochondrial proteome is estimated to contain over a thousand proteins. Thus, several hundred other proteins remain to be identified and investigated. To date, the nuclear genetic origin of many mitochondrial diseases remains to be identified and numerous genes await analysis. Our results suggest that, considering the implication of mitochondrial proteins in various pathologies, the screening of mitochondrial proteins with eubacterial homologues should be given the attention it deserves.

Acknowledgements

We thank Kanaya Malkani for critical reading of the manuscript. This work was supported by grants from INSERM, the 'CER Pays de la Loire', 'CHU d'Angers', and the University of Angers (France).

References

- [1] L. Margulis, Origin of eukaryotic cells, Yale University Press, New Haven, CT, USA, 1970.
- [2] M.W. Gray, The endosymbiont hypothesis revisited, *Int. Rev. Cytol.* 141 (1992) 233–357.
- [3] T. Sicheritz-Ponten, C.G. Kurland, S.G. Andersson, A phylogenetic analysis of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae, *Biochim. Biophys. Acta* 1365 (1998) 545–551.

- [4] S.G. Andersson, A. Zomorodipour, J.O. Andersson, T. Sicheritz-Ponten, U.C. Alsmark, R.M. Podowski, A.K. Naslund, A.S. Eriksson, H.H. Winkler, C.G. Kurland, The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria, *Nature* 396 (1998) 133–140.
- [5] M. Martin, M. Müller, The hydrogen hypothesis for the first eukaryote, *Nature* 392 (1998) 37–41.
- [6] M.W. Gray, G. Burger, B.F. Lang, Mitochondrial evolution, *Science* 283 (1999) 1476–1481.
- [7] M.W. Gray, G. Burger, B.F. Lang, The origin and early evolution of mitochondria, *Genome Biol.* 2 (2001) 1018.
- [8] C.G. Kurland, S.G. Andersson, Origin and evolution of the mitochondrial proteome, *Microbiol. Mol. Biol. Rev.* 64 (2000) 786–820.
- [9] K. Henze, W. Martin, How do mitochondrial genes get into the nucleus?, *Trends Genet.* 17 (2001) 383–387.
- [10] T. Mourier, A.J. Hansen, E. Willerslev, P. Arctander, The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus, *Mol. Biol. Evol.* 18 (2001) 1833–1837.
- [11] M. Woischnik, C.T. Moraes, Pattern of organization of human mitochondrial pseudogenes in the nuclear genome, *Genome Res.* 12 (2002) 885–893.
- [12] Y. Tourmen, O. Baris, P. Dessen, C. Jacques, Y. Malhière, P. Reynier, Structure and chromosomal distribution of human mitochondrial pseudogenes, *Genomics* 80 (2002) 71–77.
- [13] B. Westermann, W. Neupert, ‘Omics’ of the mitochondrion, *Nat. Biotechnol.* 21 (2003) 239–240.
- [14] E. Richly, P.F. Chinnery, D. Leister, Evolutionary diversification of mitochondrial proteomes: implications for human disease, *Trends Genet.* 19 (2003) 356–362.
- [15] O. Karlberg, B. Canback, C.G. Kurland, S.G. Andersson, The dual origin of the yeast mitochondrial proteome, *Yeast* 17 (2000) 170–187.
- [16] E.M. Marcotte, I. Xenarios, A.M. van Der Blik, D. Eisenberg, Localizing proteins in the cell from their phylogenetic profiles, *Proc. Natl Acad. Sci.* 97 (2000) 12115–12120.
- [17] J.N. Timmis, M.A. Ayliffe, C.Y. Huang, W. Martin, Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes, *Nat. Rev. Genet.* 5 (2004) 123–135.
- [18] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402.
- [19] M.G. Claros, P. Vincens, Computational method to predict mitochondrially imported proteins and their targeting sequences, *Eur. J. Biochem.* 241 (1996) 779–786.
- [20] O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, *J. Mol. Biol.* 300 (2000) 1005–1016.
- [21] B. Ludwig, G. Schatz, A two-subunit cytochrome c oxidase (cytochrome aa3) from *Paracoccus denitrificans*, *Proc. Natl Acad. Sci.* 77 (1980) 196–200.
- [22] V. Tiranti, K. Hoertnagel, R. Carozzo, C. Galimberti, M. Munaro, M. Granatiero, L. Zelante, P. Gasparini, R. Marzella, M. Rocchi, M.P. Bayona-Bafaluy, J.A. Enriquez, G. Uziel, E. Bertini, C. Dionisi-Vici, B. Franco, T. Meitinger, M. Zeviani, Mutations of SURF-1 in Leigh disease associated with cytochrome c oxidase deficiency, *Am. J. Hum. Genet.* 63 (1998) 1609–1621.
- [23] Z. Zhu, J. Yao, T. Johns, K. Fu, I. De Bie, C. Macmillan, A.P. Cuthbert, R.F. Newbold, J. Wang, M. Chevrette, G.K. Brown, R.M. Brown, E.A. Shoubridge, SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome, *Nat. Genet.* 20 (1998) 337–343.
- [24] I. Valnot, S. Osmond, N. Gigarel, B. Mehaye, J. Amiel, V. Cormier-Daire, A. Munnich, J.-P. Bonnefond, P. Rustin, A. Rotig, Mutations of the SCO1 gene in mitochondrial cytochrome c oxidase deficiency with neonatal-onset hepatic failure and encephalopathy, *Am. J. Hum. Genet.* 67 (2000) 1104–1109.
- [25] L.C. Papadopoulou, C.M. Sue, M.M. Davidson, K. Tanji, I. Nishino, J.E. Sadlock, S. Krishna, W. Walker, J. Selby, D.M. Glerum, R.V. Coster, G. Lyon, E. Scalais, R. Lebel, P. Kaplan, S. Shanske, D.C. De Vivo, E. Bonilla, M. Hirano, S. DiMauro, E.A. Schon, Fatal infantile cardioencephalomyopathy with COX deficiency and mutations in SCO2, a COX assembly gene, *Nat. Genet.* 23 (1999) 333–337.
- [26] I. Valnot, J.C. von Kleist-Retzow, A. Barrientos, M. Gorbatyuk, J.W. Taanman, B. Mehaye, P. Rustin, A. Tzagoloff, A. Munnich, A. Rotig, Mutation in the human heme A:farnesyltransferase gene (COX10) causes cytochrome c oxidase deficiency, *Hum. Mol. Genet.* 9 (2000) 1245–1249.
- [27] P.L. Adams, R.N. Lightowers, D.M. Turnbull, Molecular analysis of cytochrome c oxidase deficiency in Leigh’s syndrome, *Ann. Neurol.* 41 (1997) 268–270.
- [28] M. Jaksch, S. Hofmann, S. Kleinle, S. Liechti-Gallati, D.E. Pongratz, J. Muller-Hocker, K.B. Jedele, T. Meitinger, K.D. Gerbitz, A systematic mutation screen of 10 nuclear and 25 mitochondrial candidate genes in 21 patients with cytochrome c oxidase (COX) deficiency shows tRNA(Ser)(UCN) mutations in a subgroup with syndromal encephalopathy, *J. Med. Genet.* 35 (1998) 895–900.
- [29] M. Zeviani, A. Spinazzola, V. Carelli, Nuclear genes in mitochondrial disorder, *Curr. Opin. Genet. Dev.* 13 (2003) 262–270.
- [30] S. Haut, M. Brivet, G. Touati, P. Rustin, S. Lebon, A. Garcia-Cazorla, J.-M. Saudubray, A. Boutron, A. Legrand, A. Slama, A deletion in the human QP-C gene causes a complex III deficiency resulting in hypoglycaemia and lactic acidosis, *Hum. Genet.* 113 (2003) 118–122.
- [31] P. Benit, R. Beugnot, D. Chretien, I. Giurgea, P. De Lonlay-Debeney, J.-P. Issartel, M. Corral-Debrinbski, S. Kerscher, P. Rustin, A. Rotig, A. Munnich, Mutant NDUFV2 subunit of mitochondrial complex I causes early onset hypertrophic cardiomyopathy and encephalopathy, *Hum. Mutat.* 21 (2003) 582–586.
- [32] P. Benit, A. Slama, F. Cartault, I. Giurgea, D. Chrétien, S. Lebon, C. Marsac, A. Munnich, A. Rotig, P. Rustin, Mutant NDUFS3 subunit of mitochondrial complex I causes Leigh syndrome, *J. Med. Genet.* 41 (2004) 14–17.

3.1.2 Étude complémentaire non publiée : méthode de prédiction à haut débit de nouvelles protéines mitochondriales par génomique comparative

3.1.2.1 Introduction

À ce jour, un seul type d'adressage des protéines mitochondriales est connu [209, 210], sur lequel reposent la majorité des méthodes bio-informatiques de prédiction (Tableau 6 et Tableau 7). Elles sont basées sur la reconnaissance d'un peptide signal N-terminal chargé positivement et bipolaire, contenu par un grand nombre de protéines destinées à la matrice ou à la membrane interne [83-85].

Dans l'article précédent (Premier article : « Les mitochondriopathies impliquent majoritairement des protéines d'origine procaryote », p. 74), les protéines mitochondriales humaines connues ont été analysées afin de déterminer leur origine procaryote ou non (origine eucaryote présumée). Elles ont été classées en deux groupes : celles qui ont un homologue procaryote, notées PH+, et celles qui n'ont pas d'homologue procaryote, notées PH-, [211]. Il a été montré que les protéines PH+ sont préférentiellement impliquées dans les pathologies de la chaîne respiratoire. Les gènes codants ces protéines PH+ seraient donc des candidats potentiels pour découvrir de nouveaux gènes impliqués dans des pathologies.

Également, il a été montré suite à ce travail que 88 % des protéines mitochondriales humaines possédant un homologue procaryote sont significativement plus grandes (moyenne : 435 acides aminés ; amplitude : 69–1 500) que leur homologue procaryote (moyenne : 400 acides aminés ; amplitude : 38–1 353 ; $p < 0,0001$). Cette différence s'explique principalement par la présence d'une extension N-terminale ne présentant pas d'homologie avec des protéines procaryotes ; cette séquence serait probablement d'origine eucaryote. La majorité des protéines homologues à des protéines procaryotes serait donc des protéines chimères eucaryotes-procaryotes. Nous faisons alors l'hypothèse que l'on peut utiliser ce motif pour rechercher dans le génome humain des gènes mitochondriaux encore inconnus.

Nous avons comparé les séquences de 291 protéomes d'archéobactéries et d'eubactéries (disponibles complets en juillet 2006, à l'époque de l'étude) contre le génome humain sur l'outil informatique ReMIX [212]. Nous utilisons le programme *tblastn* de la suite iBLAST, combiné à un outil de filtrage que nous avons développé afin de ne retenir que les séquences d'intérêt. Un tel calcul génère beaucoup de résultats, entre autres en raison des nombreux pseudogènes mitochondriaux présents dans le génome nucléaire [10, 11]. Nous avons alors recherché les meilleurs paramètres de filtrage afin de limiter les faux positifs, tout en retrouvant les gènes mitochondriaux connus (témoins). Si nous obtenons ces

gènes mitochondriaux déjà référencés dans les bases de données, nous pensons également être en mesure d'identifier de nouveaux gènes mitochondriaux (classés PH+).

3.1.2.2 Matériel et méthodes

Banques de données

La localisation mitochondriale des protéines a été obtenue à partir des annotations de la banque de données Swiss-Prot (<http://www.uniprot.org>) [213]. L'annotation PH+ ou PH- de ces protéines a été ajoutée d'après l'étude effectuée précédemment [211]. La séquence du génome humain est celle de la banque de données *UCSC Genome Browser Database*, version hg18 (mars 2006) [214]. La base de données relationnelle associée, contenant les informations sur la position des gènes, a été utilisée pour notre étude. Des gènes sont considérés comme connus s'ils sont référencés dans la table *knownGenes* de cette base de données. Les 270 protéomes d'archéobactéries et d'eubactéries (disponibles complets en 2005, à l'époque de l'étude) ont été obtenus sur le site du *National Center for Biotechnology Information (NCBI Microbial Genomes Resources)*, http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html.

Alignements des séquences

Les séquences de chacune des 692 567 protéines bactériennes (291 protéomes) d'archéobactéries et d'eubactéries ont été comparées avec les séquences des 24 chromosomes haploïdes humains en utilisant le programme *tblastn* de la suite iBLAST, adaptation de NCBI BLAST [215] sur l'architecture ReMIX (voir ci-dessous) [212]. Nous avons utilisé $E < 1 \times 10^{-4}$ comme critère afin d'éviter les faux positifs dans les similarités.

Le projet ReMIX

La méthode de prédiction présentée dans ce travail de thèse est basée sur des alignements de séquences nécessitant des ressources informatiques importantes. La comparaison des environ 700 000 protéines bactériennes contre le génome humain a été réalisée dans le cadre d'une action de recherche collaborative de l'institut national de recherche en informatique et en automatique (ARC INRIA 2006–2007, <http://www.inria.fr/recherche/arc/actionsds06.fr.html>). Cette ARC réunissait quatre équipes ayant des compétences différentes sur un projet centré sur la recherche par le contenu dans les banques de séquences d'ADN. Elle s'est appuyée sur le projet *A Reconfigurable Memory for Indexing Mass of Data* (ReMIX, <http://www.irisa.fr/remix/>) démarré en 2004 qui consistait à concevoir un système ayant un accès rapide à un espace mémoire important. Dans ce contexte, l'équipe Symbiose (IRISA, Rennes) a développé un prototype possédant une mémoire FLASH de 512 giga-octets pilotée par des composants reconfigurables qui, suivant

l'application, reconfigurent les accès mémoire et le traitement associés aux données. Concrètement, le prototype se présente sous un petit cluster de 4 PC équipés chacun de 2 cartes PCI sur lesquelles se trouvent 64 Go de mémoire FLASH connectée à un composant reconfigurable Xilinx Virtex-2 Pro (Figure 16). Un environnement de programmation permet d'avoir une vision unifiée de l'espace mémoire disponible.

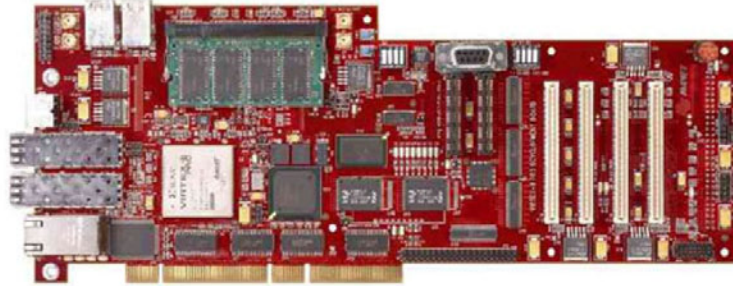


Figure 16. Carte PCI ReMIX comportant 64 Go de mémoire FLASH et un composant reconfigurable Xilinx Virtex-2 Pro.

L'originalité de ReMIX réside dans le fait de disposer de capacité de calcul au plus près des données. Pour des applications ciblées sur la recherche dans les masses de données, cela permet d'éliminer extrêmement rapidement (au vol) une grande quantité de données et, ainsi, de consacrer les ressources en calcul du processeur hôte uniquement aux informations pertinentes. Grâce à sa capacité mémoire importante et à son accès rapide, ReMIX peut stocker des structures de données sophistiquées qui évitent de balayer l'ensemble des données pour rechercher de l'information. Dans le cas des banques d'ADN, plutôt que de stocker linéairement les séquences, on peut imaginer un schéma d'indexation qui permet de ne pointer que vers des sous-ensembles de séquences — ou portions de séquences. Il faut alors définir des critères permettant d'indexer ces sous-ensembles. Ce type de schéma d'indexation est employé dans le logiciel BLAST [215] : la confrontation d'une séquence d'ADN et d'une banque consiste d'abord à indexer la requête sur la base de mots (graines) de 10 à 12 caractères. La taille des mots définit la sensibilité et la rapidité de la recherche.

Sur ReMIX, c'est la banque qui est indexée de manière permanente dans la mémoire FLASH. Le choix des graines est donc extrêmement important pour garantir une certaine sensibilité. De plus, à sensibilité fixée, il peut exister plusieurs graines qui peuvent influencer sur la taille de l'index. En général, les études portent sur l'optimisation des graines pour favoriser une sensibilité maximale. Ici, la composante « taille de l'index » doit être largement considérée pour minimiser l'espace de stockage. Un travail préliminaire a donc été de considérer le problème de l'optimisation de graines sous un angle différent : pour une sensibilité donnée, quelles sont les graines qui procurent une indexation minimale [216, 217].

Nous avons ensuite expérimenté les divers modèles d'indexation sur notre étude qui, en terme informatique, revient à comparer plusieurs centaines de milliers de protéines d'origine procaryote contre le génome humain. Avec le logiciel BLAST, combiné à un outil de filtrage maison, le temps de calcul est estimé à plus de dix jours sur le supercalculateur scalaire IBM-SP4 de l'IDRIS (12 nœuds SMP P690+ avec 32 processeurs / RAM 128 Go, <http://www.idris.fr/>).

L'usage d'un système comme ReMIX est intéressant à double titre. D'une part, l'usage de gros moyens de calcul (tels que ceux disponibles à l'IDRIS) est une procédure relativement lourde et qui supporte mal les phases de mise au point pour des codes non disponibles en standard. La mise à disposition de ReMIX, aussi bien pour calibrer l'application que pour exécuter un long calcul, était immédiate. D'autre part, son architecture optimisée permet de réduire considérablement le temps de calcul qui a été au final d'une dizaine d'heures.

Filtrage des résultats

Le programme *tblastn* de la suite iBLAST génère les mêmes résultats que celui de la suite NCBI BLAST traditionnelle. Dans le cadre de notre étude, le programme *tblastn* présente les résultats sous la forme d'un tableau, en ne conservant que les résultats nécessaires. Le détail des alignements n'est pas conservé, réduisant ainsi très sensiblement la quantité de données générées. Nous avons développé trois filtres, appelés *mitopid* et écrits en langage Perl, pour traiter les résultats obtenus (Figure 17) :

1. Le premier filtre (*mitopid1*) traite la sortie iBLAST en fusionnant les alignements qui se recouvrent puis en les sélectionnant en fonction de leur couverture et de leurs segments N- et C-terminaux. L'algorithme de *mitopid1* traite les données selon les principes généraux suivants :
 - a. Les résultats élémentaires obtenus dans les fichiers de sortie du programme iBLAST qui correspondent à des alignements (HSP, high-scoring segment pairs) considérés contigus, c'est-à-dire suffisamment proches, sont concaténés. Les HSP non contigus sont conservés tels quels.
 - b. Ne seront ensuite conservés en sortie du filtre que les HSP qui ont une couverture suffisante et qui débutent et se terminent autour de positions fixées en partie C- et N-terminale de la protéine alignée. La couverture est calculée pour chaque HSP contigu ou non. Elle est définie comme le rapport en pour cent entre la taille de l'alignement fusionné et la taille de la protéine bactérienne.
 - c. Au final, le fichier de sortie du filtre *mitopid1* contient des HSP contigus qui peuvent correspondre aux exons contigus d'un gène sur un chromosome

humain. Son algorithme détaillé et son code source sont donnés en annexe (Annexe A : Le filtre *mitopid1*, p. 173).

2. Le second filtre (*mitopid2*) prend les résultats du premier et interroge la base de données de l'UCSC pour identifier les positions des séquences alignées et les rattacher à des gènes connus. Les alignements correspondants à des gènes déjà connus sont identifiés et les résultats du filtre *mitopid1* sont enrichis d'informations sur ces gènes. La recherche de gènes témoins est effectuée à partir des résultats de ce deuxième filtre. Son algorithme détaillé et son code source sont donnés en annexe (Annexe B : Le filtre *mitopid2*, p. 189).
3. Le troisième filtre (*mitopid3*) identifie les gènes candidats mitochondriaux selon l'hypothèse que nous avons formulée (voir Introduction, p. 83). Ce filtre compare les caractéristiques de chaque alignement retenu par le deuxième filtre avec son gène (ou son gène prédit) correspondant dans la base de données de l'UCSC. Un gène candidat est retenu par ce filtre s'il valide les paramètres correspondants à notre hypothèse. Son algorithme détaillé et son code source sont donnés en annexe (Annexe C : Le filtre *mitopid3*, p. 201).

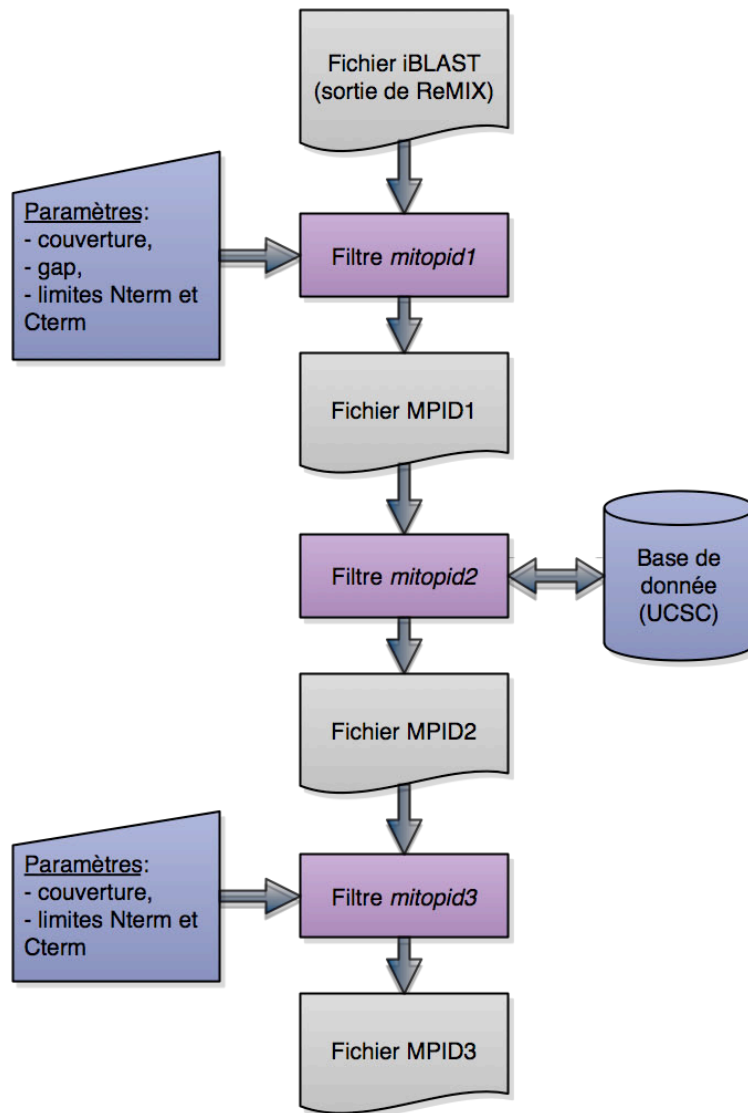


Figure 17. Flux de traitement d'une sortie iBLAST (ReMIX) avec les filtres *mitopid* et leurs différents paramètres (couverture : rapport en pour cent entre la taille de l'alignement et la taille de la protéine bactérienne ; gap : distance maximale entre deux HSP considérés contigus ; Nterm : limite en position N-terminale ; Cterm : limite en position C-terminale).

Dans notre étude, nous faisons varier exclusivement le paramètre de filtrage selon la couverture. Les limites N- et C-terminales de *mitopid1* ne sont pas utilisées. La distance maximale entre deux HSP considérés contigus est fixée à 10 000 bases.

Filtre supplémentaire pour l'étude

Le traitement des données en sortie d'iBLAST nécessite trois filtres (*mitopid1–3*). Un quatrième filtre (*mitopid11*) a été conçu pour déterminer la valeur optimale du paramètre de couverture. Il est utilisé pour une étape préliminaire qui consiste à faire varier de 0 à 100 % le paramètre de filtrage selon la couverture des alignements, 0 % correspondant à une absence de contrainte (tous les alignements sont retenus) et 100 % revenant à ne retenir

que les alignements qui englobent toute la longueur de la séquence requête. Le filtre *mitopid11* inclut l'algorithme de filtrage selon la couverture de l'alignement du filtre *mitopid1*. Il réalise la même tâche que *mitopid1* mais permet de déporter cette opération à la suite de *mitopid2* (Figure 18). En effectuant l'étape de mise au point sans *mitopid11* et en faisant varier le paramètre de couverture en entrée de *mitopid1* de 0 à 100 % par incréments de 1 %, la deuxième étape du filtrage serait répétée 101 fois, multipliant d'autant les requêtes de *mitopid2* sur la base de données. Pour notre étude, le filtrage selon la couverture est donc déplacé après *mitopid2* sans influencer les résultats. Le temps d'exécution est alors réduit de plus d'une semaine à environ une journée. Son algorithme détaillé et son code source sont donnés en annexe (Annexe D : Le filtre *mitopid11*, p. 209).

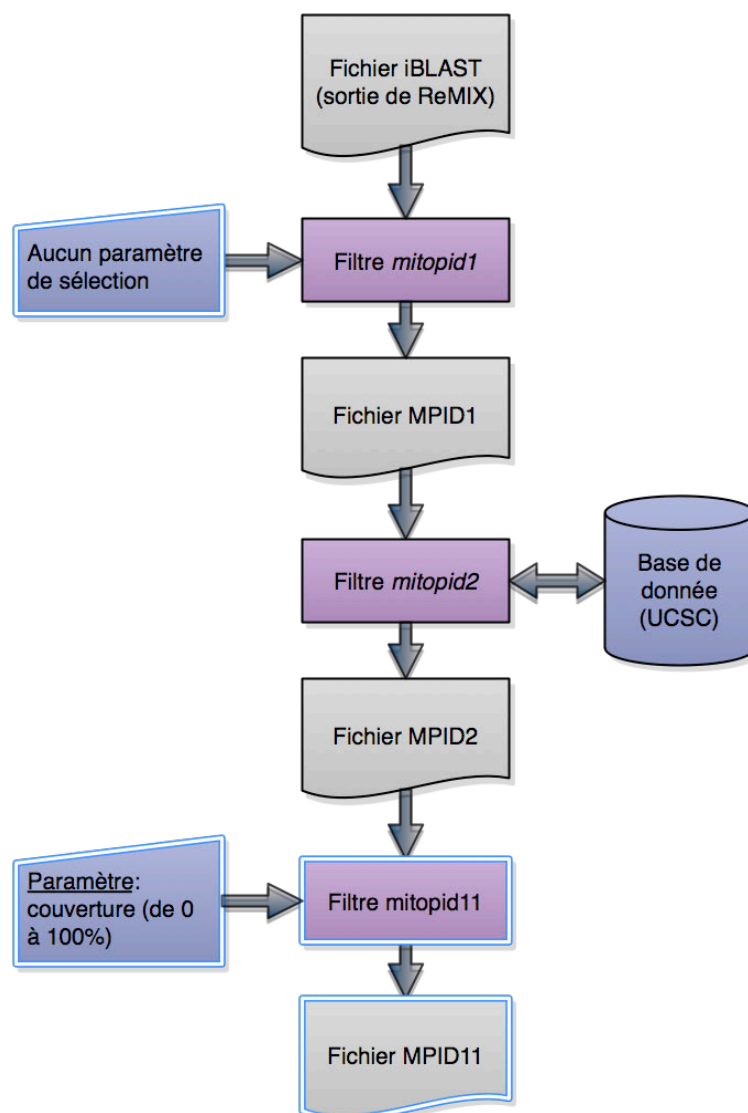


Figure 18. Flux de traitement d'une sortie iBLAST (ReMIX) intégrant le filtre *mitopid11* avec les différents paramètres (couverture : rapport en pour cent entre la taille de l'alignement et la taille de la protéine bactérienne ; Nterm : limite en position N-terminale ; Cterm : limite en position C-terminale).

Ensemble de référence

Trois tests sont mis en place pour estimer les paramètres optimaux des filtres *mitopid*, qui seront effectués seulement sur le chromosome 19 humain pour des raisons de temps de calcul. L'ensemble de référence pour notre étude sur le chromosome 19 humain contient les dix protéines mitochondriales PH+ connues (compte tenu de l'hypothèse de départ), codées par les gènes *BCAT2*, *ECH1*, *ETFB*, *GCDH*, *GPX4*, *NDUS7*, *ODBA*, *RM34*, *SYSM* et *TIM44*.

Sensibilité

La sensibilité est définie comme le pourcentage de gènes de l'ensemble de référence qui sont effectivement classés comme gènes mitochondriaux PH+ en sortie des filtres :

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

Avec :

- VP, le nombre de vrais positifs c'est-à-dire le nombre de gènes mitochondriaux PH+ détectés par les filtres comme PH+ ;
- FN, le nombre de faux négatifs c'est-à-dire le nombre de gènes mitochondriaux PH+ non détectés par les filtres comme PH+.

Le programme de test fait varier la couverture de 0 à 100 % et calcule la sensibilité. Sur le chromosome 19, l'ensemble de référence contient dix protéines et la formule devient :

$$\text{Sensibilité} = \frac{VP}{10}$$

Spécificité

La spécificité est définie comme le pourcentage de gènes n'appartenant pas à l'ensemble de référence qui sont effectivement classés comme non mitochondriaux en sortie des filtres :

$$\text{Spécificité} = \frac{VN}{FP + VN}$$

Avec :

- VN, le nombre de vrais négatifs, c'est-à-dire le nombre de gènes sur le chromosome 19 qui ne sont pas mitochondriaux PH+ et qui n'ont pas été détectés par les filtres comme étant mitochondriaux PH+.
- FP, le nombre de faux positifs, c'est-à-dire le nombre de gènes non mitochondriaux PH+ qui sont détectés par les filtres comme étant mitochondriaux PH+.

La clé primaire de la base de données de l'UCSC (utilisée par *mitopid2*) est un identifiant interne et un même gène peut être référencé par plusieurs clés. La spécificité est donc estimée en considérant les identifiants UCSC, soit un ensemble de référence de 17 identifiants parmi les 2 010 du chromosome (voir ci-dessous).

Identifiants UCSC et Swiss-Prot

La base de données Swiss-Prot répertorie des protéines et la base de données UCSC Genome Browser Database référence, entre autres, des gènes du génome humain. Pour récupérer certaines informations reliant ces deux bases, nous avons utilisé deux requêtes SQL (*Simple Query Language*).

Pour connaître les identifiants UCSC correspondants aux protéines mitochondriales PH+ codées par le chromosome 19 qui sont référencées dans Swiss-Prot, nous avons effectué la requête suivante :

```
SELECT spID, proteinID, alias FROM knownGene, kgSpAlias
WHERE chrom LIKE 'chr19'
      AND name=kgID
      AND (
          alias LIKE 'BCAT2%'
          OR alias LIKE 'ECH1%'
          OR alias LIKE 'ETFB%'
          OR alias LIKE 'GCDH%'
          OR alias LIKE 'GPX4%'
          OR alias LIKE 'NDUFS7%'
          OR alias LIKE 'ODBA%'
          OR alias LIKE 'RM34%'
          OR alias LIKE 'SYSM%'
          OR alias LIKE 'TIM44%'
      ) GROUP BY spid, proteinID;
```

Le résultat obtenu est présenté dans le Tableau 10. Ce tableau contient, pour chaque gène témoin, tous les *proteinID* à identifier en sortie des filtres pour les calculs de sensibilité et spécificité. Par exemple, le gène *GCDH* devra être recherché sous l'identifiant « GCDH_HUMAN », mais aussi « Q92947-2 ».

Tableau 10. Correspondance entre les identifiants *UCSC Genome Browser Database* et *Swiss-Prot*. *spID* : identifiant *Swiss-Prot* ; *proteinID* : identifiant *UCSC* ; *alias* : nom du gène donné par l'*UCSC*.

spID	proteinID	alias
O15382	BCAT2_HUMAN	BCAT2
Q13011	ECH1_HUMAN	ECH1
P38117	ETFB_HUMAN	ETFB
Q6ZQP2	Q6ZQP2_HUMAN	ETFB
Q71RF6	Q71RF6_HUMAN	ETFB
Q92947	GCDH_HUMAN	GCDH
Q92947-2	Q92947-2	GCDH
Q6PJX4	Q6PJX4_HUMAN	GPX4
O75251	NUKM_HUMAN	NDUFS7
Q6ZQU6	Q6ZQU6_HUMAN	NDUFS7
Q6ZS38	Q6ZS38_HUMAN	NDUFS7
Q7LD69	Q7LD69_HUMAN	NDUFS7
P12694	ODBA_HUMAN	ODBA
Q9BQ48	RM34_HUMAN	RM34
Q9NP81	SYSM_HUMAN	SYSM
O43615	TIM44_HUMAN	TIM44
Q9UPE4	Q9UPE4_HUMAN	TIM44

La requête SQL suivante est utilisée pour connaître le nombre total d'identifiants de protéines qui sont codées par des gènes localisés sur le chromosome 19 dans la base de données de l'*UCSC* :

```
SELECT DISTINCT proteinID FROM knownGene WHERE chrom LIKE 'chr19';
```

Le résultat de la requête comporte 2 010 identifiants. Cette valeur est supérieure au nombre de gènes codants dans la base de données, car, comme nous l'avons vu précédemment dans le Tableau 10, il y a de la redondance.

3.1.2.3 Résultats

Le nombre d'HSP contigus retenu par le filtre *mitopid1* (c'est-à-dire les gènes candidats suite à la première étape du filtrage) en fonction du paramètre de couverture est représenté Figure 19. Plus le pourcentage de couverture est important et plus le nombre d'HSP contigus diminue, selon une courbe en forme de sigmoïde. À 23 % de couverture, le filtre élimine la moitié des données de iBLAST.

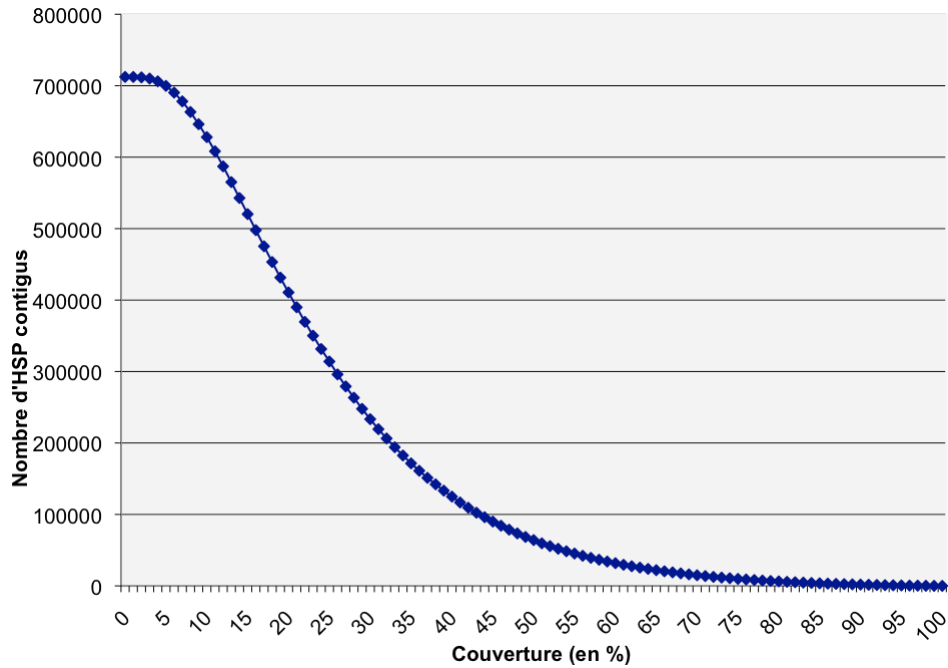


Figure 19. Performance du filtrage : nombre d'HSP contigus retenu en fonction de la couverture (en pour cent).

La sensibilité du filtrage diminue lorsque la contrainte sur la couverture fixée dans *mitopid1* augmente (Figure 20). À 0 % de couverture, la sensibilité est de 80 % : deux gènes sur dix (RM34 et TIM44) ne sont pas détectés, quelle que soit la couverture. À 23 % de couverture, la sensibilité est de 50 %.

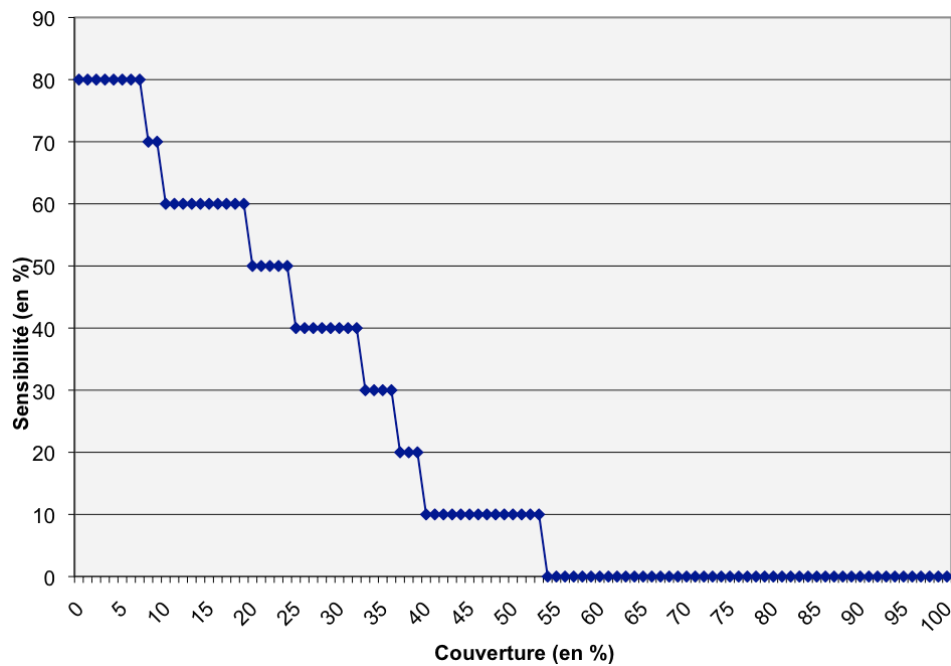


Figure 20. Sensibilité du filtrage en fonction de la couverture (en pour cent)

La spécificité du filtrage augmente lorsque la contrainte sur la couverture fixée dans *mitopid1* augmente (Figure 21). À 0 % de couverture, la spécificité est de 62,3 % ; à 23 % de couverture, la spécificité est de 79 %. Lorsque la couverture requise est de 100 %, la spécificité est de 100 % puisqu'aucun identifiant de gène n'est retrouvé.

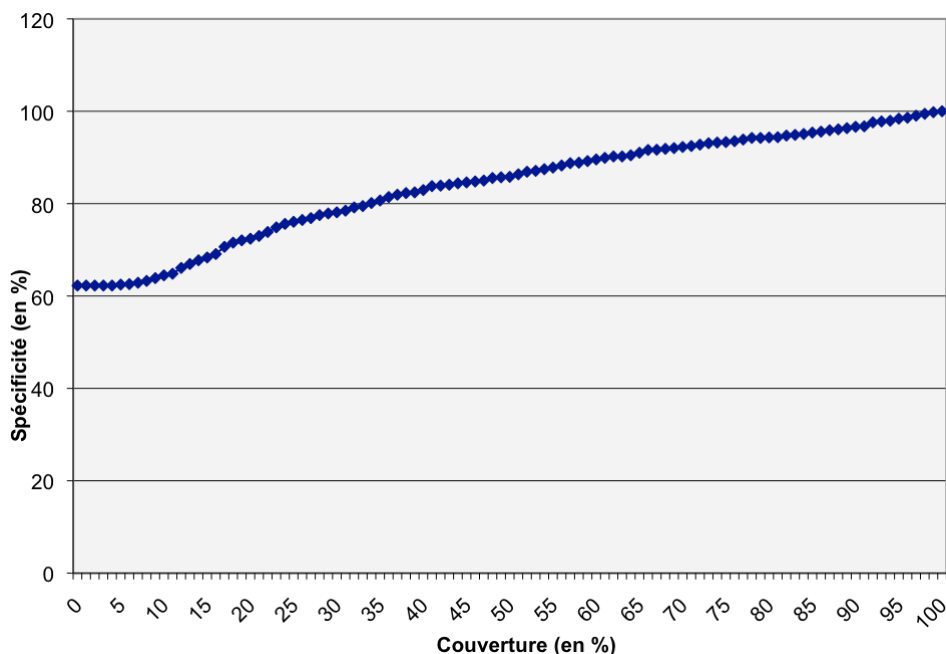


Figure 21. Spécificité du filtrage en fonction de la couverture (en pour cent).

À 23 % de couverture, notre outil retient alors 350 165 HSP contigus, correspondants à 511 protéines candidates (Tableau 11). Parmi celles-ci, il y a cinq résultats vrais positifs, cinq faux négatifs, 506 faux positifs et 2 005 vrais négatifs.

Tableau 11. Liste des 511 protéines mitochondriales PH+ candidates, identifiées avec un paramètre de couverture de 23 %. Les identifiants Swiss-Prot sont séparés par des virgules, avec les résultats vrais positifs en gras et les redondances regroupées par les couleurs de police bleue, rouge et verte.

O95996_HUMAN, CLPP_HUMAN, G3PT_HUMAN, Q9NVU6_HUMAN, Q96EY9_HUMAN, Q9BV20_HUMAN, RS15_HUMAN, AAAT_HUMAN, ANGL6_HUMAN, TA2R_HUMAN, O75228_HUMAN, Q6DK52_HUMAN, **Q7LD69_HUMAN**, **Q6ZQU6_HUMAN**, **NUKM_HUMAN**, PRS6B_HUMAN, P43686-2, Q96GZ7_HUMAN, CAF1A_HUMAN, SIRT6_HUMAN, Q8N6T7-4, Q8NDC9_HUMAN, RS5_HUMAN, Q8NAA1_HUMAN, Q7RTU6_HUMAN, Q7Z5V7_HUMAN, Q5EBL2_HUMAN, Q9H6K5_HUMAN, Q9ULT2_HUMAN, Q711P8_HUMAN, PPP5_HUMAN, Q9BPW0_HUMAN, UCRI_HUMAN, CPNS1_HUMAN, CCG8_HUMAN, **Q6PJX4_HUMAN**, PRDX2_HUMAN, Q6P390_HUMAN, SHAN1_HUMAN, Q8N6W7_HUMAN, DNJB1_HUMAN, MLL4_HUMAN, BRD4_HUMAN, RKHD1_HUMAN, Q6DD92_HUMAN, Q6ZS25_HUMAN, Q8N6I1_HUMAN, K1543_HUMAN, O00455_HUMAN, RFX1_HUMAN, BTEB4_HUMAN, FUT1_HUMAN, Q99450_HUMAN, FUT2_HUMAN, RS16_HUMAN, HCN2_HUMAN, Q86VE0_HUMAN, NOTC3_HUMAN, NMDE4_HUMAN, Q6ZQQ7_HUMAN, Q7Z4V5_HUMAN, Q9BW08_HUMAN, CS016_HUMAN, FKRP_HUMAN, NOVA2_HUMAN, SIX5_HUMAN, RYR1_HUMAN, SPTN4_HUMAN, Q9HCE9_HUMAN, APOE_HUMAN, Q6QNY0_HUMAN, ZBT7A_HUMAN, Q9BRY0_HUMAN, PALM_HUMAN, ZN759_HUMAN, Q6ZTZ0_HUMAN, Q6UWP8_HUMAN, Q6ZQR3_HUMAN, Q6NY19_HUMAN, Q6T4P5_HUMAN, Q9HAJ4_HUMAN, SF3A2_HUMAN, NUP62_HUMAN, Q96B18_HUMAN, KLF2_HUMAN, CAC1A_HUMAN, SELV_HUMAN, Q3KQZ3_HUMAN, MBRL_HUMAN, O75257_HUMAN, Q3KQX0_HUMAN, RU17_HUMAN, P08621-2, Q9UFS1_HUMAN, Q6ZS42_HUMAN, Q6ZMX8_HUMAN, ZN324_HUMAN, Q9H861_HUMAN, ILF3_HUMAN, Q6XCG5_HUMAN, Q96AS7_HUMAN, NPAS1_HUMAN, RPOM_HUMAN, P2Y11_HUMAN, JUNB_HUMAN, CILP2_HUMAN, ZN575_HUMAN, ATF5_HUMAN, Q96SJ0_HUMAN, Q8N594_HUMAN, Q6ZS72_HUMAN, Q4VHE4_HUMAN,

HAS1_HUMAN, Q9BSF4_HUMAN, Q86X18_HUMAN, Q8TCM5_HUMAN, HNRPM_HUMAN, Q7KYM9_HUMAN, Q8N7U7_HUMAN, MAST1_HUMAN, IASPP_HUMAN, LRFN3_HUMAN, SFR16_HUMAN, Q8N2M8-3, SEM6B_HUMAN, CNOT3_HUMAN, PE2R1_HUMAN, CIC_HUMAN, ZNF42_HUMAN, NMD3B_HUMAN, GPR41_HUMAN, VASP_HUMAN, ZN499_HUMAN, Q6ZQW4_HUMAN, Q8N7K9_HUMAN, Q96PG8_HUMAN, Q6UUV9-3, MECT1_HUMAN, ETV2_HUMAN, S11L3_HUMAN, Q8WU30_HUMAN, Q8IWX8_HUMAN, ZN576_HUMAN, SAFB2_HUMAN, Q86Y97_HUMAN, Q9BRZ6_HUMAN, ZN580_HUMAN, PRAX_HUMAN, MADCA_HUMAN, DAZP1_HUMAN, LIPS_HUMAN, O15446_HUMAN, Q9UPF6_HUMAN, EDG6_HUMAN, Q8N825_HUMAN, Q8TDY7_HUMAN, LONM_HUMAN, NUMBL_HUMAN, FUBP2_HUMAN, OPA3_HUMAN, ZSWM4_HUMAN, PHLB3_HUMAN, Q5U682_HUMAN, Q6ZTE5_HUMAN, Q6PEZ8_HUMAN, DAPK3_HUMAN, Q8WVWF5_HUMAN, O75866_HUMAN, Q9HAE9_HUMAN, PLVAP_HUMAN, Q96LJ4_HUMAN, KCNC3_HUMAN, Q4G188_HUMAN, Q8N8V1_HUMAN, Q8N1H8_HUMAN, FOSB_HUMAN, Q49AD7_HUMAN, PAR4_HUMAN, Q96B70_HUMAN, CEBPA_HUMAN, Q96CN4_HUMAN, TCGAP_HUMAN, Q7Z7M0-2, EGFL4_HUMAN, ERF_HUMAN, Q9H024_HUMAN, ZN497_HUMAN, HUMAN, DBP_HUMAN, FIG1_HUMAN, Q6ZLN8_HUMAN, ZN541_HUMAN, PAK4_HUMAN, Q8NCH5_HUMAN, Q8N4E1_HUMAN, HNF3G_HUMAN, Q95783_HUMAN, BRSK1_HUMAN, O15090_HUMAN, Q8NCA9_HUMAN, ZN358_HUMAN, Q8IU81_HUMAN, Q6NXF1_HUMAN, Q9NUX3_HUMAN, ZNF71_HUMAN, Q6ZWC9_HUMAN, Q6ZUL8_HUMAN, NAC2_HUMAN, KLF1_HUMAN, MARK4_HUMAN, Q6IPE9_HUMAN, Q8N2N5_HUMAN, Q6X3Y3_HUMAN, MIS_HUMAN, Q8N6N2_HUMAN, C5ARL_HUMAN, ZN444_HUMAN, Q66K64_HUMAN, Q53EK9_HUMAN, VMDL1_HUMAN, DMWD_HUMAN, Q8WUW6_HUMAN, Q95784_HUMAN, CHST8_HUMAN, FIZ1_HUMAN, CCDC8_HUMAN, Q96M06_HUMAN, M3K10_HUMAN, Q8NA69_HUMAN, EAR2_HUMAN, A2GL_HUMAN, Q9H0W8_HUMAN, Q9H9A9_HUMAN, SMCA4_HUMAN, Q6NXC4_HUMAN, Q4G0X8_HUMAN, O75532_HUMAN, EPN1_HUMAN, Q8IY88_HUMAN, **GCDH_HUMAN**, Q6ZUR2_HUMAN, Q2YDY3_HUMAN, Q9H5H8_HUMAN, **ETFB_HUMAN**, **Q71RF6_HUMAN**, Q9NUP0_HUMAN, Q9H8S4_HUMAN, Q6ZSA3_HUMAN, **ODBA_HUMAN**, Q8N188_HUMAN, Q8NBS4_HUMAN, DLL3_HUMAN, Q32W58_HUMAN, Q6UXC7_HUMAN, Q32W66_HUMAN, Q32W67_HUMAN, Q8IYK2_HUMAN, Q8NEY2_HUMAN, Q8IX33_HUMAN, ZN342_HUMAN, BTBD2_HUMAN, Q8WTR8_HUMAN, Q6NXR0_HUMAN, Q96S99_HUMAN, Q6B089_HUMAN, LYL1_HUMAN, EF2_HUMAN, MYO9B_HUMAN, DIRA1_HUMAN, Q6ZVX7_HUMAN, Q9H7S7_HUMAN, CDN2D_HUMAN, Q96IQ9_HUMAN, Q96MG2_HUMAN, Q6PEZ9_HUMAN, GDF1_HUMAN, PI2R_HUMAN, GPR4_HUMAN, LRP3_HUMAN, Q13459-2, O10H1_HUMAN, S27A1_HUMAN, RENT1_HUMAN, Q6IPE0_HUMAN, Q6ZN10_HUMAN, Q96C55_HUMAN, BBC3_HUMAN, SRCH_HUMAN, Q8TB06_HUMAN, Q96M54_HUMAN, Q8NE63_HUMAN, EAA4_HUMAN, Q8N753_HUMAN, ENL_HUMAN, BORG3_HUMAN, RAIN_HUMAN, Q8N115_HUMAN, LPHN1_HUMAN, FBX27_HUMAN, RCN3_HUMAN, Q6ZW31_HUMAN, Q7L218_HUMAN, Q96AB3_HUMAN, Q9H838_HUMAN, Q6ZTW0_HUMAN, Q6ZV13_HUMAN, Q8TC94_HUMAN, PKN1_HUMAN, GRIK5_HUMAN, Q69YU8_HUMAN, Q32MH8_HUMAN, NALP7_HUMAN, NALP2_HUMAN, Q9NX02-2, Q6ZS11_HUMAN, Q6UXY4_HUMAN, SPHK2_HUMAN, Q8N2M3_HUMAN, Q8N9G5_HUMAN, GDF15_HUMAN, CSO21_HUMAN, Q59FK8_HUMAN, Q96G46_HUMAN, Q8IZ63_HUMAN, Q6ZS82_HUMAN, RD23A_HUMAN, O14526_HUMAN, TTP_HUMAN, RDH13_HUMAN, KISSR_HUMAN, Q5PRF9_HUMAN, Q8IXP7_HUMAN, Q8N2T9_HUMAN, Q9H735_HUMAN, Q8NCJ3_HUMAN, EHD2_HUMAN, EDG8_HUMAN, Q9NX00_HUMAN, Q68CT3_HUMAN, DHX34_HUMAN, TSH3_HUMAN, Q7L8P4_HUMAN, Q6ZS20_HUMAN, Q9HAL3_HUMAN, Q6GMU2_HUMAN, Q8NAL1_HUMAN, TM86B_HUMAN, Q6ZS40_HUMAN, Q96RP8_HUMAN, LAM2_HUMAN, Q8N3H4_HUMAN, Q86V59_HUMAN, Q14576-2, ELAV3_HUMAN, Q96J71_HUMAN, ELAV1_HUMAN, NAL12_HUMAN, P59046-4, P59046-2, FUT5_HUMAN, FUT3_HUMAN, Q6P7E6_HUMAN, GSK3A_HUMAN, NRTN_HUMAN, Q9BZC4_HUMAN, Q9NR73_HUMAN, Q96S58_HUMAN, O10H2_HUMAN, CRX_HUMAN, Q7Z7M8_HUMAN, GPR43_HUMAN, CRSP7_HUMAN, KLK11_HUMAN, Q9H253_HUMAN, Q9HA55_HUMAN, Q75253_HUMAN, MAG_HUMAN, O75265_HUMAN, Q96RE7_HUMAN, THAP8_HUMAN, Q8WVP5_HUMAN, Q499Z8_HUMAN, Q8NBM0_HUMAN, O00426_HUMAN, Q9BT33_HUMAN, Q5XKB9_HUMAN, Q6IPW8_HUMAN, SAFB1_HUMAN, Q8TBC5_HUMAN, AKAP8_HUMAN, HNRPL_HUMAN, Q969G4_HUMAN, Q8TCY6_HUMAN, DOT1L_HUMAN, DYR1B_HUMAN, GRLF1_HUMAN, PPA5_HUMAN, Q9HAD2_HUMAN, CSPG3_HUMAN, Q6PMT2_HUMAN, PDE4A_HUMAN, Q86V67_HUMAN, K1683_HUMAN, Q96BI4_HUMAN, CDC37_HUMAN, DHR10_HUMAN, Q6P6B0_HUMAN, Q96SQ6_HUMAN, Q6ZN24_HUMAN, RS9_HUMAN, RS28_HUMAN, GLU2B_HUMAN, **Q6ZS38_HUMAN**, SIGL6_HUMAN, Q8N9Y3_HUMAN, Q6ZN01_HUMAN, Q3KQU9_HUMAN, Q8WU63_HUMAN, Q96BF3_HUMAN, Q6UW59_HUMAN, Q8N1L2_HUMAN, Q3SX65_HUMAN, AKP8L_HUMAN, IRK14_HUMAN, Q9Y463-2, Q9H166_HUMAN, Q8N7H5_HUMAN, LIG4_HUMAN, ZN446_HUMAN, PKHA4_HUMAN, Q86YF0_HUMAN, TIF1B_HUMAN, Q9GZY2_HUMAN, Q5H9P9_HUMAN, P78418_HUMAN, Q96BU2_HUMAN, Q9H699_HUMAN, Q9UQ10_HUMAN, Q15653-2, IKBB_HUMAN, Q96LS1_HUMAN, Q9H014_HUMAN, Q8WXF8-2, DEDD2_HUMAN, KPCD2_HUMAN, Q8NCK8_HUMAN, SF04_HUMAN, Q8N819_HUMAN, Q5UGI7_HUMAN, ZSCA1_HUMAN, Q9BSN5_HUMAN, Q8N8K8_HUMAN, DYN2_HUMAN, Q8N1K8_HUMAN, Q9UPH5_HUMAN, Q8N344_HUMAN, Q9H6N0_HUMAN, Q8N5R7_HUMAN, O00320_HUMAN, JMJ2B_HUMAN, GMIP_HUMAN, TR95_HUMAN, GYS1_HUMAN, Q9BTT9_HUMAN, Q9NXX1_HUMAN, Q6ZUN4_HUMAN, KCNK6_HUMAN, BCL3_HUMAN, Q96FE3_HUMAN, Q8WTQ8_HUMAN, CO5A3_HUMAN, IER2_HUMAN, ST2B1_HUMAN, O00204-2, RS19_HUMAN, SRTD3_HUMAN, Q6ZP66_HUMAN, **Q92947-2**, COMP_HUMAN, CBLC_HUMAN, ELL_HUMAN, Q9BV16_HUMAN, Q96SI1_HUMAN, Q96GS6_HUMAN, Q8WUH9_HUMAN, KCNN4_HUMAN, CLC11_HUMAN, MYADM_HUMAN, B3GN3_HUMAN, PIAS4_HUMAN, A1BG_HUMAN, Q7Z3U3_HUMAN, ARHG1_HUMAN, Q6NX52_HUMAN, Q99786_HUMAN, Q8N1W5_HUMAN, Q659D7_HUMAN, Q96AD9_HUMAN, Y0355_HUMAN, Q9BSS5_HUMAN, Q96005_HUMAN, Q9BQ72_HUMAN, ZN285_HUMAN, Q6ZP71_HUMAN, Q6ZN19_HUMAN, TGFB1_HUMAN, SNPC2_HUMAN, ETHE1_HUMAN, LDLR_HUMAN, GIPC_HUMAN, Q8N214_HUMAN

3.1.2.4 Discussion et perspectives

Plus la contrainte de couverture de notre système de filtrage augmente et plus la sensibilité diminue, alors que la spécificité augmente. À partir d'une couverture de 23 %, le filtrage devient discriminant en éliminant la moitié des données brutes fournies par iBLAST (ReMIX).

Dans ce cas, la sensibilité est de 50 % et cinq gènes de l'ensemble de référence sont détectés. Pour en détecter huit, le maximum possible, la couverture requise doit être fixée au maximum à 11 %. Mais, une telle couverture ne sélectionne pas les résultats (608 245 HSP contigus au lieu de 712 286 pour 0 % de couverture). Une contrainte de 23 % sur la couverture paraît donc dans un premier temps un bon compromis entre, d'une part, la sensibilité exigée, mais surtout, d'autre part, la quantité de données obtenues en sortie des filtres. La spécificité est alors de 79 %. Dans notre étude, nous n'avons pas augmenté la sensibilité, car nous avons préféré réduire le nombre de faux positifs pour lesquels une confirmation expérimentale négative serait dilatoire.

Cette nouvelle approche pour rechercher des protéines mitochondriales n'utilise pas un motif précis pour prédire une protéine mitochondriale PH+, mais plutôt des relations d'homologie. Les résultats obtenus sont encourageants même s'ils sont difficiles à comparer quantitativement avec les autres approches bio-informatiques (Tableau 6). Cependant, bien que les autres approches s'adressent à l'ensemble des protéines mitochondriales (PH+ et PH-) et qu'en conséquence elles sont évaluées par rapport à un ensemble de référence plus complet (656 protéines de MITOP2 [67, 78]), nous pouvons extrapoler les valeurs de sensibilité (Se) et spécificité (Sp) de celles-ci pour les comparer à notre méthode. Les premières approches apparues, basées sur la prédiction du signal d'adressage N-terminal, obtiennent Se = 54,7 %, Sp = 33,1 % pour la méthode ayant la meilleure sensibilité et Se = 43,4 %, Sp = 52,1 % pour la méthode ayant la meilleure spécificité. L'approche intégrative n'obtient pas de meilleurs résultats avec Se = 43,4 %, Sp = 41,4 %, de même que la méthode par homologie avec *Rickettsia prowazekii* qui obtient Se = 42,8 %, Sp = 21,9 %. Par rapport à cette dernière approche, notre méthode est plus sensible (Se = 50 %) car elle est basée sur tous les protéomes procaryotes actuels. En effet, *Rickettsia prowazekii* est le plus proche parent de la mitochondrie actuelle, mais de nombreux transferts horizontaux entre bactéries ont très probablement eu lieu depuis l'endosymbionte originel. Il est donc préférable de considérer tous les procaryotes actuels pour l'étude de la mitochondrie moderne. De plus, notre approche offre malgré tout une spécificité élevée (Sp = 79 %) en raison des contraintes fortes appliquées pour le filtrage. Ce dernier point est indispensable, car la quantité trop importante de données générées est la principale faiblesse des études à haut débit, rendant impossibles les validations expérimentales nécessaires.

L'analyse comparative des résultats obtenus par notre méthode et les autres outils bio-informatiques existants (Tableau 12) montre que notre travail apporte des résultats nouveaux et complémentaires. Les approches existantes reposent sur la présence d'un peptide signal d'adressage N-terminal ou d'un site de clivage dans les séquences, qui font défaut dans la majorité des séquences mitochondriales. MitoProt [83] prend en compte la composition en acides aminés de la partie N-terminale des séquences alors que les autres

outils utilisent des méthodes d'apprentissage automatique supervisé, basées sur des réseaux de neurones artificiels (Predotar [85], TargetP [220]) ou des machines à vecteurs de support (SubLoc [221] et PSORT-II [222]). En particulier, seul notre outil retient la protéine nommée (par son alias) ETFB qui a été localisée expérimentalement dans la matrice mitochondriale [218]. De même, la protéine nommée GPX4 n'est identifiée comme mitochondriale que par une autre méthode (Predotar). Cette protéine est exprimée sous trois formes différentes dans le tissu testiculaire et est ciblée vers le cytosol, la mitochondrie ou le noyau par l'utilisation de codons d'initiation alternatifs [219]. Dans ce dernier cas, la diversité biologique dépasse la modélisation restrictive des outils informatiques existants. Ainsi, notre approche basée sur l'évolution pourrait mieux prendre en compte la complexité biologique.

Tableau 12. Comparaison des prédictions obtenues par notre approche (mitopid) en fixant le paramètre de couverture à 23 % avec celles de cinq outils bio-informatiques de référence : Predotar [85], TargetP [220], SubLoc [221], MitoProt [83] et PSORT-II [222]. Les protéines sont nommées en faisant référence à leur *alias*, c'est-à-dire le nom du gène correspondant donné par l'UCSC (voir Tableau 10). Une protéine est notée « + » si sa localisation est considérée comme mitochondriale, « - » sinon. Les outils ont été utilisés avec leurs paramètres par défaut.

<i>alias</i>	mitopid	Predotar	TargetP	SubLoc	MitoProt	PSORT-II
BCAT2	-	+	+	+	+	+
ECH1	-	+	+	+	+	-
ETFB	+	-	-	-	-	-
GCDH	+	+	+	-	+	+
GPX4	+	+	-	-	-	-
NUKM	+	+	+	+	+	+
ODBA	+	+	+	+	+	+
RM34	-	+	+	-	+	-
SYSM	-	+	+	-	+	+
TIM44	-	+	+	-	+	+

Nous avons montré que notre méthode est complémentaire de celles qui sont existantes. En outre, elle possède la particularité d'identifier de nouvelles protéines candidates par criblage du génome, même si celles-ci ne sont pas caractérisées. Par exemple, un extrait de dix résultats parmi les environ 350 000 du chromosome 19 humain sont représentés dans le Tableau 13. Le premier résultat (sur fond gris) indique le fragment d'un gène candidat pour une protéine mitochondriale PH+ entre les positions 51 753 740 et 51 754 703. Aucun gène ne correspond à cette région génomique dans RefSeq, une banque de données fiable et non redondante de séquences de génomes, produits de transcription et

protéines [223]. Toutefois, une séquence d'ADNc de GenBank [44] (*Accession Number* : AK094504.1) ainsi que des prédictions de gènes (Genscan [224], Geneid [225], N-SCAN [226]) y sont référencées (Figure 22). La séquence de la protéine correspondant à cet ARNm est alors prédite comme mitochondriale par les cinq outils bio-informatiques de référence utilisés précédemment (Tableau 12). Ce gène candidat mériterait donc d'être validé biologiquement. Cependant, notre étude génère beaucoup de résultats qui ne pourraient pas tous être vérifiés expérimentalement. Néanmoins, elle peut s'avérer un outil intéressant dans le cas d'une recherche qui se limiterait à une zone d'intérêt restreinte, par exemple suite à une étude de liaison (*linkage*). Ainsi, notre approche offrirait le double avantage d'identifier un gène codant potentiellement une protéine mitochondriale PH+ qui, de plus, serait préférentiellement impliqué dans une pathologie mitochondriale.

Tableau 13. Extrait de dix résultats parmi les 350 165 HSP contigus retenu par le filtrage en fixant le paramètre de couverture à 23 %. Le contenu du fichier original est présenté sous forme de tableau en ne retenant que certaines colonnes. *query_name* : numéros d'accèsion de la protéine bactérienne requête ; *strand* : sens de l'alignement ; *query_start* : position du début de l'alignement de la séquence requête ; *query_end* : position de la fin de l'alignement de la séquence requête ; *sbjct_start* : position du début de l'alignement de la séquence cible (chromosome 19 humain) ; *sbjct_end* : position de la fin de l'alignement de la séquence cible.

<i>query_name</i>	<i>strand</i>	<i>query_start</i>	<i>query_end</i>	<i>sbjct_start</i>	<i>sbjct_end</i>
gi 15828517 ref NP_325877.1	+	18	340	51753740	51754703
gi 15828677 ref NP_326037.1	+	63	282	47134698	47135388
gi 15828677 ref NP_326037.1	+	96	283	56175973	56176501
gi 15828717 ref NP_326077.1	+	120	331	12531779	12532403
gi 15828877 ref NP_326237.1	+	108	243	1115595	1115997
gi 15828877 ref NP_326237.1	-	96	311	59693392	59694027
gi 15828877 ref NP_326237.1	+	135	273	7711315	7711756
gi 15828877 ref NP_326237.1	+	130	273	7711354	7711768
gi 15828877 ref NP_326237.1	+	100	230	7711390	7711774
gi 15828877 ref NP_326237.1	-	206	389	54388845	54389387

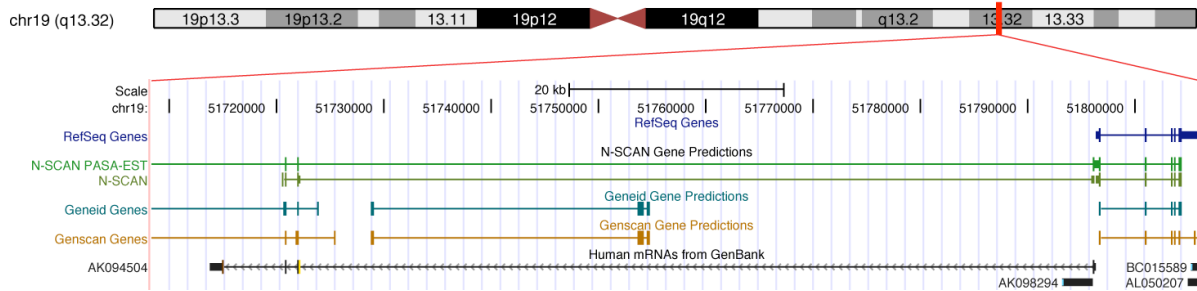


Figure 22. Représentation graphique des résultats obtenus dans la banque de données RefSeq [223], avec différents outils de prédiction de gènes (*Gene Predictions*) et parmi les séquences des ARNm humains de GenBank [44] sur le locus 19q13.32 humain (UCSC Genome Browser, version hg18, mars 2006).

Notre démarche effectuée pour le chromosome 19 doit cependant être validée pour tous les chromosomes afin de pouvoir confirmer notre hypothèse. Le filtre *mitopid3* doit également être intégré au flux de filtrage afin d'augmenter encore la spécificité. Par la suite, nous pouvons attendre de notre approche qu'elle fournisse une quantité limitée de nouvelles protéines mitochondriales possibles avec une fiabilité élevée, candidates préférentielles pour de nouvelles pathologies mitochondriales. Ainsi, notre étude pourrait contribuer aux approches intégratives en apportant une information complémentaire permettant de mieux comprendre la physiopathologie de ces affections.

3.1.3 Synthèse

Nous avons comparé les séquences des 393 protéines mitochondriales humaines répertoriées dans la banque de données Swiss-Prot avec celles de toutes les protéines procaryotes connues à l'époque de l'étude. Seuls 64 % des protéines mitochondriales humaines sont homologues à des protéines procaryotes, ce qui témoigne de leur double origine évolutive procaryote et eucaryote. Nous avons comparé ces deux groupes évolutifs de protéines selon plusieurs critères et avons montré que la structure, la localisation et la fonction des protéines mitochondriales diffèrent fortement selon leur origine. De plus, les protéines impliquées dans les pathologies de la chaîne respiratoire sont majoritairement homologues à des protéines procaryotes. Actuellement, le nombre de protéines mitochondriales connues a presque doublé et les conclusions de cette étude sont confirmées.

Ces résultats montrent que la connaissance de l'évolution des protéines mitochondriales aide à une meilleure compréhension de la dynamique des génomes et des mitochondriopathies. Cependant, celle-ci ne reposait que sur les 393 protéines connues, alors que le protéome mitochondrial est estimé à plus d'un millier de protéines. À ce jour, l'origine génétique de nombreuses pathologies mitochondriales est inconnue et de nombreux gènes restent à analyser. La découverte de nouvelles protéines mitochondriales par criblage bio-informatique à haut débit permettrait ainsi de mieux comprendre la physiopathologie de ces affections.

3.2 ÉTUDE DE LA PROTÉINE MITOCHONDRIALE OPA1 ET DES PATHOLOGIES ASSOCIÉES

Dans la seconde partie de ce travail de thèse, nous avons étudié une protéine mitochondriale en particulier, Opa1, et les pathologies associées à ses dysfonctionnements. Cette partie comporte trois articles publiés, portant d'abord sur la création d'une base de données spécifique des mutations du gène *OPA1* qui a permis, quatre années plus tard, de réaliser une large étude sur les neuropathies optiques héréditaires ; de manière complémentaire nous avons également étudié l'influence du patrimoine génétique mitochondrial sur l'AOAD :

1. ***eOPA1: an online database for OPA1 mutations.***

Ferre M., Amati-Bonneau P., Tourmen Y., Malthiery Y., and Reynier P., Hum Mutat, 2005. **25**(5): p. 423-8.

2. ***Molecular screening of 980 cases of suspected hereditary optic neuropathy with a report on 77 Novel OPA1 mutations.***

Ferre M., Bonneau D., Milea D., Chevrollier A., Verny C., Dollfus H., Ayuso C., Defoort S., Vignal C., Zanlonghi X., Charlin J.F., Kaplan J., Odent S., Hamel C.P., Procaccio V., Reynier P., and Amati-Bonneau P., Hum Mutat, 2009. **30**(7): p. E692-E705.

3. ***OPA1-related dominant optic atrophy is not strongly influenced by mitochondrial DNA background.***

Pierron D., Ferre M., Rocher C., Chevrollier A., Murail P., Thoraval D., Amati-Bonneau P., Reynier P., and Letellier T., BMC Med Genet, 2009. **10**(1): p. 70.

La première association entre une modification au niveau de l'ADN et une maladie fut publiée début 1957 par Ingram [227], confirmant au niveau moléculaire la découverte de la drépanocytose, huit ans plus tôt, par Pauling *et coll.* [228]. Cet événement ouvrit la voie à la recherche des mutations associées aux pathologies humaines, mais ce sont les progrès de la biologie moléculaire dans les années 1970, avec le développement de la technique de séquençage de Sanger [229], qui ont rendu possible la localisation et l'isolement de nombreuses maladies liées à des gènes et la détection directe des lésions pathologiques associées.

Les années 1970 marquèrent également le début de l'organisation de la connaissance dans des bases de données avec la publication de la première édition de l'ouvrage *Mendelian Inheritance in Man* qui a fêté ses 40 ans en 2006 et dont les versions actuelles sont désormais disponibles sur le site Web du NCBI (<http://www.ncbi.nlm.nih.gov/omim/>) [230]. Son auteur, Victor A. McKusick, a été l'un des premiers à comprendre la puissance de

la bio-informatique et, aujourd'hui, des milliers de maladies génétiques ont été décrites et des dizaines de milliers de mutations ont été signalées. Afin de classer, d'interpréter et d'étendre ces connaissances, de nouveaux outils ont été développés : les bases de données de mutations généralistes et les bases de données spécifiques de gènes (*Locus Specific Databases*, LSDB).

La collection de données généraliste la plus connue est certainement *Human Gene Mutation Database* (HGMD, <http://www.hgmd.org/>) qui, au 3 août 2009, enregistre 91 133 mutations répertoriées dans 3 461 gènes différents. Dans les bases de données de ce type, les mutations sont extraites automatiquement par des outils d'exploration de données depuis les publications. En revanche, les LSDB recueillent toutes les mutations publiées et non publiées d'un gène spécifique et leurs administrateurs normalisent les descriptions cliniques et biologiques de chaque mutation. En outre, chaque mutation peut apparaître plus d'une fois à une même position et donc apparaître plusieurs fois dans la base de données, permettant l'identification des points chauds de mutation.

Ainsi, différents utilisateurs ont accès à la même information sous différentes formes. Le chercheur impliqué dans l'annotation du génome humain a besoin de recueillir les variations au niveau des nucléotides et être en mesure de les placer avec précision sur une séquence génomique. Le généticien doit identifier une mutation pathogène dans un gène spécifique alors que le biologiste moléculaire étudiant les mécanismes des mutations a besoin d'obtenir une vue d'ensemble. En outre, comme cela est le cas pour *OPA1*, la découverte de nombreux gènes a été progressive et les premières séquences de référence ont été progressivement corrigées, ne permettant plus de retrouver les mutations décrites auparavant. Afin de résoudre ces problèmes, différents experts ont essayé de définir un langage commun pour le nom des mutations. Différentes tentatives ont été faites pour enfin aboutir aux recommandations d'Antonarakis en 1998 [231].

La nomenclature la plus univoque devrait être basée sur l'ADN génomique, mais seule la séquence de l'ADNc est disponible pour la plupart des gènes. Alors, l'existence de divers sites de transcription, d'exons alternatifs ou de nombre variable de répétitions complique le système de numérotation des ADNc. Puis, un troisième niveau de la nomenclature a été mis en place pour nommer les conséquences protéiques et dans tous les cas, comme la connaissance du génome évolue, la séquence de référence sur laquelle est basé le nom de chaque mutation doit être précisée pour en garantir l'intégrité dans le temps. Les substitutions d'une paire de bases et les délétions sont les plus fréquentes mutations du génome humain ; les autres mutations possibles sont les insertions, les duplications, les inversions, les extensions et les réarrangements complexes. Ces mutations se produisent non seulement dans les séquences codantes, mais également dans les régions promotrices, les jonctions d'épissage des introns et les régions non traduites. Elles peuvent interférer à

chaque étape de la voie d'expression, du gène à la protéine produite. Avec l'évolution de la génétique humaine, de nouveaux types de mutations sont identifiés et les règles de nomenclature doivent être constamment adaptées. Antonarakis et den Dunnen ont joué un rôle majeur dans l'élaboration de ces recommandations, qui sont maintenant largement utilisées [232]. La description complète du système de nomenclature, en évolution continue, est consultable sur le site Internet de la société savante dont le but est de « favoriser la découverte et la caractérisation des variations du génome » (*Human Genome Variation Society*, HGVS ; <http://www.hgvs.org/>).

C'est dans ce contexte, alors que les liens entre les atrophies optiques héréditaires et les mitochondries commençaient à être compris grâce à l'identification du gène *OPA1* et à un spectre de mutations associées, que nous avons débuté en 2002 le développement de la première LSDB de ce gène, soumise pour publication en 2004 (voir Second article : « eOPA1: une base de données pour les mutations du gène *OPA1* », p. 106) [179]. Le recensement de ces données est une condition préalable pour être en mesure d'expliquer les variations interindividuelles et la façon dont une expression phénotypique variable peut être associée à des lésions d'un même gène. Ainsi, après une période de quatre années, nous avons publié une large étude sur près de mille patients adressés auprès de notre laboratoire pour une suspicion de neuropathie optique héréditaire (voir Troisième article : « Dépistage moléculaire de 980 cas suspectés de neuropathie optique héréditaire et publication de 77 nouvelles mutations du gène *OPA1* », p. 114).

En outre, le patrimoine génétique mitochondrial est connu pour influencer l'expression de certaines pathologies. En particulier, la NOHL causée par les mutations mitochondriales m.11778G>A et m.14484T>C serait significativement liée à l'haplogroupe mitochondrial caucasien J [233]. Cet haplogroupe J serait également surreprésenté trois fois chez les patients atteints d'AOAD non liée à des mutations d'*OPA1* [156]. En conséquence, en complément de nos précédents travaux, nous avons étudié l'influence du patrimoine mitochondrial sur les AOAD liées à des mutations d'*OPA1* (voir Quatrième article : « L'atrophie optique dominante liée à *OPA1* n'est pas fortement influencée par le fond génétique mitochondrial », p. 130). Cette étude a été réalisée en collaboration avec l'équipe de Thierry Letellier (Inserm U688, Bordeaux).

3.2.1 Second article : « eOPA1: une base de données pour les mutations du gène *OPA1* »

DATABASES

eOPA1: An Online Database for *OPA1* Mutations

Marc Ferré,* Patrizia Amati-Bonneau, Yves Tourmen, Yves Malthiery, Pascal Reynier

INSERM-E0018, Laboratoire de Biochimie et Biologie Moléculaire, CHU Angers, France

Communicated by Alastair Brown

Autosomal dominant optic atrophy (ADOA), also known as Kjer disease, is characterized by moderate to severe loss of visual acuity with an insidious onset in early childhood, blue-yellow dyschromatopsia, and central scotoma. An optic atrophy gene, called *OPA1*, has been identified in most cases of the disease. A total of 83 *OPA1* mutations, often family-specific, have been reported so far, and the observations support the hypothesis that haploinsufficiency and the functional loss of a single allele may lead to ADOA. We have developed a new locus-specific database (LSDB), *eOPA1* (<http://lbbma.univ-angers.fr/eOPA1/>) aimed at collecting published and unpublished sequence variations in *OPA1*. The database has been designed to incorporate new submissions rapidly and will provide a secured online catalog of *OPA1* mutations and nonpathogenic sequence variants (NPSVs). The LSDB should prove useful for molecular diagnosis, large-scale mutation statistics, and the determination of original genotype–phenotype correlations in studies on ADOA. *Hum Mutat* 25:423–428, 2005. © 2005 Wiley-Liss, Inc.

KEY WORDS: *eOPA1*; optic atrophy 1; *OPA1*; autosomal dominant optic atrophy; ADOA; database

DATABASES:

OPA1 – OMIM: 165500, 605290; GDB: 118848; GenBank: NM_015560, NM_130831, NM_130832, NM_130833, NM_130834, NM_130835, NM_130836, NM_130837; SwissProt: O60313; HGMD: *OPA1*
<http://lbbma.univ-angers.fr/eOPA1/> (*eOPA1*)

INTRODUCTION

Autosomal dominant optic atrophy (ADOA; MIM# 165500), the most frequent form of hereditary optic atrophy, was first described by Kjer in 1959. The worldwide frequency of the disease was estimated at 1:50,000 [Lyle, 1990], and 1:10,000 in Denmark [Eiberg et al., 1994; Kjer et al., 1996]. The disease, which generally appears in childhood, is characterized by a progressive decrease in visual acuity, optic nerve pallor, sometimes by dyschromatopsia that usually predominates in the blue-yellow hues (tritanopia), and loss of sensitivity in the central visual field (central, paracentral, or cecocentral scotomas). A gene located at 3q28–q29 [Eiberg et al., 1994], identified in most cases of ADOA, has been named *OPA1* (optic atrophy 1; MIM# 605290) [Alexander et al., 2000; Delettre et al., 2000]. *OPA1*, which has 30 coding exons (including three alternative exons) [Delettre et al., 2001], encodes a dynamin-related GTPase involved in mitochondrial network dynamics.

A total of 83 *OPA1* gene mutations, mainly family-specific, and one deletion of the entire *OPA1* gene have been described so far [Alexander et al., 2000; Delettre et al., 2000, 2001, 2002; Pesch et al., 2001; Thiselton et al., 2001; Toomes et al., 2001; Marchbank et al., 2002; Shimizu et al., 2002; Baris et al., 2003; Yamada et al., 2003]. It has been hypothesized that haploinsufficiency with null mutations and the functional loss of a single allele play a major role in ADOA.

While it is relatively easy to publish initial reports of novel phenotype–genotype studies, it is rather more difficult to find publishing space for new mutations discovered later unless they present some immediate advantage. Obviously, new phenotypes associated with *OPA1* could help improve the indication of

molecular diagnosis, as in the case of optic atrophy associated with deafness [Amati-Bonneau et al., 2003]. Again, catalogs of benign polymorphisms might appear to be of little interest from the editorial point of view although they could be of considerable practical importance in molecular diagnosis. With these considerations in mind, we have developed *eOPA1*, a web-based locus-specific database (LSDB) for *OPA1* sequence variations with short publication delays. To our knowledge, this is the first LSDB dedicated to *OPA1*. We here describe the construction of the database, the procedure for data submission, and the presentation of the data.

DATABASE STRUCTURE AND CONTENT

The *eOPA1* database has been developed with the use case model taking into account all possible interactions between different types of users and the various functions of the system (Fig. 1).

Received 12 July 2004; accepted revised manuscript 8 November 2004.

Grant sponsors: INSERM; GIS-Institut des Maladies Rares; University Hospital of Angers, University of Angers, Angers, France; Retina France.

*Correspondence to: Marc Ferré, INSERM-E0018, Laboratoire de Biochimie et Biologie Moléculaire, CHU Angers, 4 rue Larrey, F-49033 Angers Cedex 01, France.

E-mail: marc.ferre@med.univ-angers.fr

DOI 10.1002/humu.20161

Published online in Wiley InterScience (www.interscience.wiley.com).

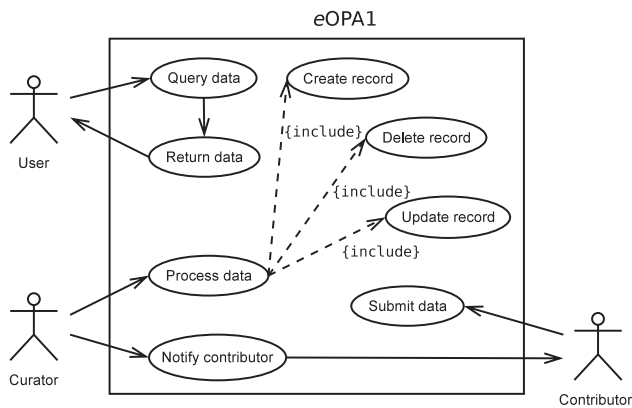


FIGURE 1. The use case diagram for *eOPA1*. The solid frame represents the boundary of the system; the ovals represent the functions of the system; and the ball-and-stick figures represent the roles that interact with the system.

Source of Data

Mutation data was gathered from all the available publications (links shown on the website). All the names of the mutations and the nonpathogenic sequence variants (NPSVs) in DNA, as well as the names of the protein sequence variants, were converted according to the widely accepted system of the Nomenclature Working Group [Antonarakis and the Nomenclature Working Group, 1998; den Dunnen and Antonarakis, 2000; den Dunnen and Paalman, 2003].

Database Structure

The website is constructed with MySQL as the database system (www.mysql.com). MySQL is a free, open-source, relational database management system that uses the Structured Query Language (SQL) to process and present data. As a relational database, MySQL uses tables and columns that can be interrelated by keys. Figure 2 shows the content of the various tables that constitute the database. The entire *eOPA1* database, including the SQL files and the tab-delimited flat files, is available on request from the authors.

Deployment

The system, protected by a firewall, is hosted on two Linux 2.4/dual Pentium III Xeon 1.4-GHz servers (Sun Microsystems, Santa Clara, CA), connected in a master/slave configuration. The server within the firewall stores the master copy of *eOPA1*, supervised by the Curator, and provides Intranet service for local users; whereas the server located outside the firewall stores the slave copy, and offers Internet service for international users. The service is implemented as server-side software written in PHP (PHP: Hypertext Processor) with a hypertext markup language (HTML) and a JavaScript front-end. The database statistics, computed dynamically taking into account the current content, are displayed using the graphical functionality of PHP Version 4.3.2 (www.php.net).

WEBSITE USE

eOPA1 is available at <http://lbbma.univ-angers.fr/eOPA1/>. Data may be accessed: 1) by browsing through the exhaustive list of mutations and NPSVs; 2) by clicking on a graphical representation of the gene; or 3) by using a search engine. Furthermore, each mutation and NPSV is fully described on a dedicated page.

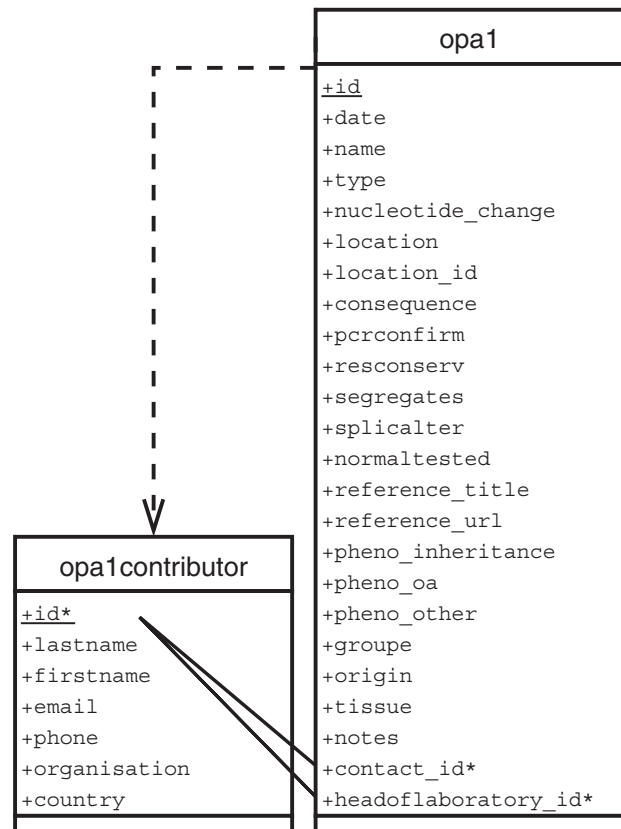


FIGURE 2. The structure of the database, showing the components of the various tables. The item *id* is a unique identifier for the database entry (primary key); the item *date* is the date of submission of the data; the item *location* is the gene location of the mutation, e.g., "Exon 1," "Intron 1," "Exon 2," etc.); the item *location_id* is the location identifier, which allows calling up a list of mutations by full exon/intron number (rather than by the first digit); the item *type* indicates whether the sequence variation has been identified as a mutation or an NPSV, the reason being briefly described in *nucleotide_change*, and the consequence, in terms of the effect on the amino acid sequence, in *consequence*. Publication details include the *reference_title* and the *reference_url*. The item *pcrconfirm* indicates if the mutation has been confirmed by a second PCR; the item *resconserv* indicates if the mutation affects conserved residue; the item *segregates* indicates if the mutation segregates with the phenotype; the item *splicalter* indicates if the splicing alteration has been verified by RT-PCR; and the item *normaltested* indicates the number of normal individuals tested. The item *pheno_inheritance* indicates the phenotype inheritance; the item *pheno_oa* indicates if the classic phenotype (optic atrophy) is present; and the item *pheno_other* indicates if another non-classic phenotype is observed. The items *groupe* and *origin* indicate respectively the ethnic group and the country of origin of the patient; the item *tissue* indicates the tissue tested. There are two links to the *opa1contributor* table (foreign keys): *contact_id*, containing personal information concerning the submitter, and *headoflaboratory_id*, containing information concerning the head of the laboratory. The item *notes* contains any further comments the submitter may wish to include. The database fields denoted by a star (*) are for internal use only.

Query Interface

First, users of the website may call up lists of all the *OPA1* mutations and NPSVs by clicking on the corresponding link on the Left Menu Bar of the *eOPA1* Home Page (Fig. 3a). Fields may be browsed (Fig. 3b): 1) by location (*Loc*); 2) by name (*Name*); 3) by

(a)

(b)

(c)

(d)

FIGURE 3. The eOPA1 web interface. **a**: the Front Page; **b**: the Mutations list page; and **c**: a page dedicated to the mutation named c.6G>A. First, the main data is displayed in a yellow box: name (with the reference transcript variant specified), location, nucleotide change, consequence, phenotype inheritance, presence of optic atrophy, country of origin, and a link to the relative publication. The “Note” field indicates whether the name of the mutation has been modified according to the recommendations of the Nomenclature Working Group. Secondly, six extra elements of clinical information are provided, concerning the criteria of pathogenicity and the tissue tested. Thirdly, information relative to the submission includes the date of submission and contact information. Finally, the permanent eOPA1 identifier, which must be cited with any questions regarding the information displayed, is indicated; and **d**: The Advanced Search page.

function (*Nucleotide Change*); and 4) by the consequence at the proteic level (*Consequence*). Users may then follow the link ([+]) to obtain further details (Fig. 3c). Second, clicking on the graphical representation of OPA1 (Fig. 3a and b) displays a list restricted to a single exon or intron. Third, eOPA1 may be interrogated through an Advanced Search page, linked to the Left

Menu Bar of the Home Page (Fig. 3d), by filling in the name and location of the mutation, the nucleotide change, the consequence at the proteic level, the reference of the publication (name of leading author) and the country or region of origin of the patients described. Up to three fields, containing space-separated terms combined with the Boolean operators [AND] and [OR] may be

used. Finally, searches may be performed either on the mutations alone, or on NPSVs alone, or simultaneously on mutations and NPSVs. In addition, the Search Bar, at the top of each page, gives quick access to the *eOPAI* database.

Data Submission and Control

Researchers may submit published or unpublished data to *eOPAI* by means of a web form adapted from the HUGO Mutation Database Initiative/Human Genome Variation Society web form (www.genomic.unimelb.edu.au/mdi/entry.html or <http://ariel.ucsf.unimelb.edu.au/~cotton/entry.htm>). This has a small number of fields, designed to be filled conveniently and quickly. Moreover, each contributor may create a personal account by clicking on the identification box at the top right of the page. Thus, in the event of a new submission, personal information will be immediately recalled and the first section of the submission

form automatically completed so that it will be necessary to fill in only three fields to enter the new submission. However, there will be provision for the optional addition of further details, such as phenotype information, e.g., indication of the original phenotype, or other comments.

Mutations should be described by contributors in natural language. The Curator will check the conformity of the data submitted and then name it according to the standard nomenclature. On submission, a temporary identifier will be affected to each new mutation (TA_###). After validation, this identifier will be made permanent (OA_###). Obviously, permanent identifiers will act as the unique keys to the *eOPAI* database. Finally, two automatically generated e-mail messages will be sent to contributors to confirm: 1) the submission of data, indicating the temporary identifier; and 2) the validation of data, indicating the permanent identifier.

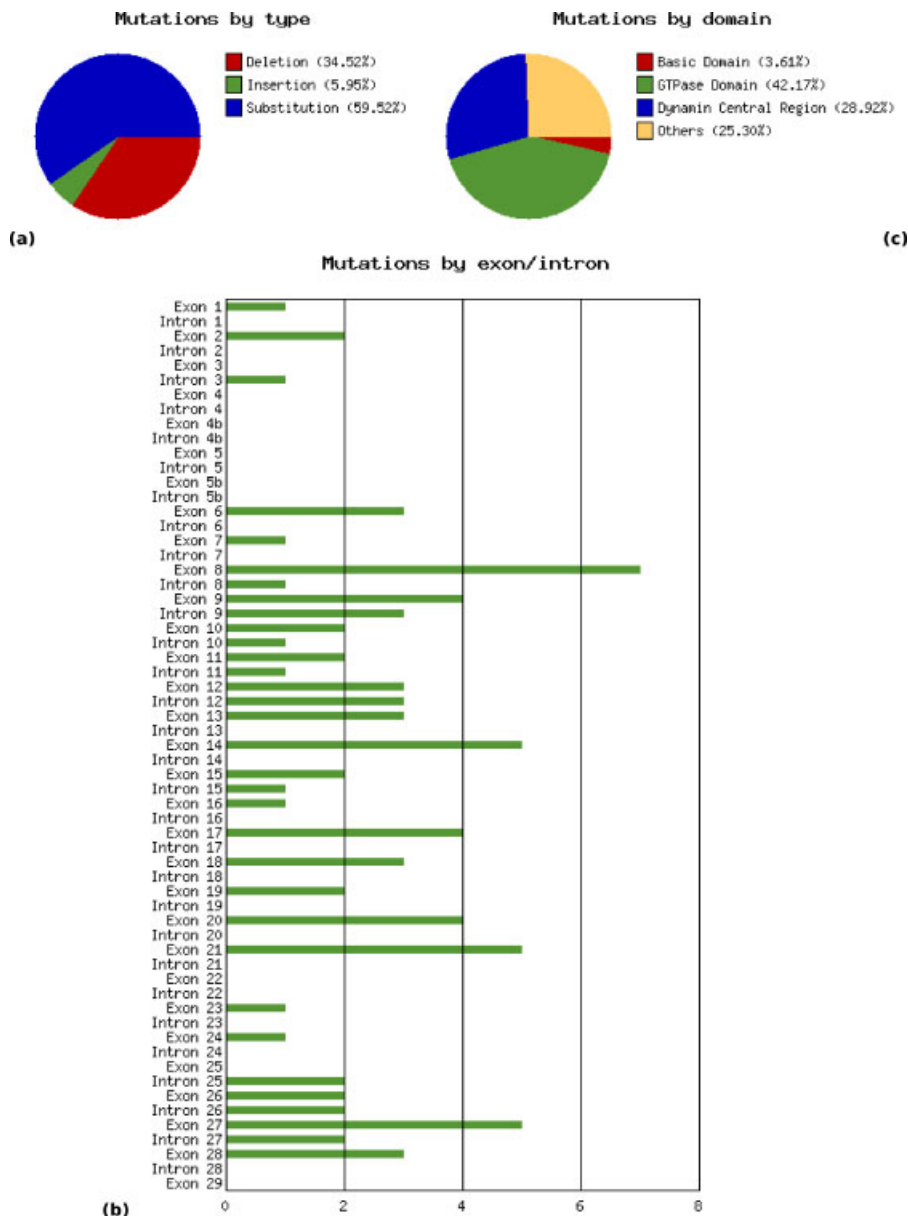


FIGURE 4. Database statistics automatically generated on the website, as of June 2004. Distribution of the 83 mutations according to their: **a:** type; **b:** location; and **c:** domain.

RESULTS AND DISCUSSION

Database Statistics

The website includes a page displaying statistics; these are automatically recalculated after each new entry and thus always up to date. The database currently contains 83 mutations and 34 NPSVs. The *OPA1* mutations are mainly substitutions (Fig. 4a). They are spread throughout the gene coding sequence (Fig. 4b), 35 (41%) are localized in GTPase domain (Fig. 4c) and 16 (19%) in the 3' end of the coding region (exons 25–30). Almost 50% of the mutations cause premature truncation of the *OPA1* protein, and 20% of the mutations are intronic.

Nomenclature

While most variants can be described in the latest update of the standard nomenclature (www.hgvs.org/mutnomen) [see also den Dunnen and Antonarakis, 2000; den Dunnen and Paalman, 2003], some inaccuracies may persist because gene anomalies discovered earlier might have been named according to a convention now out of use. Eventually, the “Notes” field of the page dedicated to each entry (Fig. 3c) indicates whether the name of the mutation has been modified by the Curator. The *eOPA1* website will require absolute compliance with the rules set out above to describe sequence variants in order to provide uniform and comparable data.

The naming *OPA1* gene mutations is a complex task because: 1) there are three alternative exons (exons 4, 4b, and 5b) and 8 isoforms [Delettre et al., 2001], and thus eight different potential classifications; 2) the published mutations and NPSVs were named according to the *OPA1* transcript variant 1. This variant (1) is the original transcript identified. It contains 29 exons and encodes an isoform (1) of 960 amino acids (aa). So, in order to maintain compatibility and homogeneity, mutations and NPSVs in *eOPA1* are named according to *OPA1* transcript variant 1. But, if a mutation or a NPSV affects alternative exons 4b or 5b, the *OPA1* transcript variant 8, which is for present the most complete cDNA, is used as reference. This variant (8) contains two additional exons, 4b and 5b, compared to transcript variant 1. However, it maintains the same reading frame and encodes an isoform (8) of 1015 aa.

Data Security

The main advantage of the master/slave deployment is data integrity. The firewall systems that protect the servers are configured to allow data to be transferred from the master to the slave but not contrariwise. If the data in the Internet server is corrupted or modified accidentally, it will be replaced by a copy duplicated from the master server. All submissions from contributors will be stored in a temporary database. These records will be retrieved by the master server and stored while awaiting validation. Each new submission will be validated by the Curator on the Intranet server before being made available on the two servers of the system. This network architecture induces an offset between the data supported by the master and the slave servers. The frequency of the updates will depend on the server load but ideally the delay should not exceed 24 hr.

FUTURE PROSPECTS

In the future, we hope to improve the molecular diagnosis of hereditary optic neuropathy by adapting our website to support new genes. For instance, we have recently shown that the *OPA3* (optic atrophy 3) gene is involved in ADOA associated with

cataract [Reynier et al., 2004] and this finding will soon be incorporated into a twin database.

When citing the present article in their publications, users of this website are requested to include the website address (universal resource locator, URL) <http://Ibbma.univ-angers.fr/eOPA1/>. We invite researchers and clinicians to submit not only their mutation data but also NPSVs, which can be very useful in molecular diagnosis. We welcome all suggestions and comments that would help keep this website relevant to the *OPA1* research community. Finally, we would be grateful for indications of any omissions or errors in the current version of the database.

ACKNOWLEDGMENTS

We thank the technicians of the Molecular Biology Laboratory of the University Hospital of Angers, France, for their excellent technical assistance, research students Thomas Raimbault and Virginie Guillet, and Kanaya Malkani for critical reading and comments on the manuscript.

REFERENCES

- Alexander C, Votruba M, Pesch UE, Thiselton DL, Mayer S, Moore A, Rodriguez M, Kellner U, Leo-Kottler B, Auburger G, Bhattacharya SS, Wissinger B. 2000. *OPA1*, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28. *Nat Genet* 26: 211–215.
- Amati-Bonneau P, Odent S, Derrien C, Pasquier L, Malthiery Y, Reynier P, Bonneau D. 2003. The association of autosomal dominant optic atrophy and moderate deafness may be due to the R445H mutation in the *OPA1* gene. *Am J Ophthalmol* 136:1170–1171.
- Antonarakis SE, the Nomenclature Working Group. 1998. Recommendations for a nomenclature system for human gene mutations. *Hum Mutat* 11:1–3.
- Baris O, Delettre C, Amati-Bonneau P, Surget MO, Charlin JF, Catier A, Derieux L, Guyomard JL, Dollfus H, Jonveaux P, Ayuso C, Maumenee I, Lorenz B, Mohammed S, Tourmen Y, Bonneau D, Malthiery Y, Hamel C, Reynier P. 2003. Fourteen novel *OPA1* mutations in autosomal dominant optic atrophy including two de novo mutations in sporadic optic atrophy. *Hum Mutat* 21:656.
- Delettre C, Lenaers G, Griffoin J-M, Gigarel N, Lorenzo C, Belenguer P, Pelloquin L, Grosgeorge J, Turc-Carel C, Perret E, Astarie-Dequeker C, Lasquelles L, Arnaud B, Ducommun B, Kaplan J, Hamel C. 2000. Nuclear gene *OPA1*, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy. *Nat Genet* 26:207–210.
- Delettre C, Griffoin JM, Kaplan J, Dollfus H, Lorenz B, Faivre L, Lenaers G, Belenguer P, Hamel C. 2001. Mutation spectrum and splicing variants in the *OPA1* gene. *Hum Genet* 109:584–591.
- Delettre C, Lenaers G, Pelloquin L, Belenguer P, Hamel CP. 2002. *OPA1* (Kjer type) dominant optic atrophy: a novel mitochondrial disease. *Mol Genet Metab* 75:97–107.
- den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat*. 15:7–12.
- den Dunnen JT, Paalman MH. 2003. Standardizing mutation nomenclature: why bother? *Hum Mutat* 22:181–2.
- Eiberg H, Kjer B, Kjer P, Rosemberg T. 1994. Dominant optic atrophy (*OPA1*) mapped to chromosome 3q region. Linkage analysis. *Hum Mol Genet* 3:977–980.

- Kjer B, Eiberg H, Kjer P, Rosemberg T. 1996. Dominant optic atrophy mapped to chromosome 3q region II. Clinical and epidemiological aspects. *Acta Ophthalmol Scand* 74:3–7.
- Lyle WM. 1990. Genetic risks. Waterloo, Ontario: University of Waterloo Press.
- Marchbank NJ, Craig JF, Leek JP, Toohey M, Churchill AJ, Markham AF, Mackey DA, Toomes C, Inglehearn CF. 2002. Deletion of the *OPA1* gene in a dominant optic atrophy family: evidence that haploinsufficiency is the cause of disease. *J Med Genet* 39:e47.
- Pesch UE, Leo-Kottler B, Mayer S, Jurklics B, Kellner U, Apfelstedt-Sylla E, Zrenner E, Alexander C, Wissinger B. 2001. *OPA1* mutations in patients with autosomal dominant optic atrophy and evidence for semi-dominant inheritance. *Hum Mol Genet* 10:1359–1368.
- Reynier P, Amati-Bonneau P, Verny C, Olichon A, Simard G, Guichet A, Bonnemains C, Malecaze F, Malinge MC, Pelletier JB, Calvas P, Dollfus H, Belenguer P, Malthiery Y, Lenaers G, Bonneau D. 2004. *OPA3* gene mutations responsible for autosomal dominant optic atrophy and cataract. *J Med Genet* 41:e110.
- Shimizu S, Mori N, Kishi M, Sugata H, Tsuda A, Kubota N. 2002. A novel mutation of the *OPA1* gene in a Japanese family with optic atrophy type I. *Jpn J Ophthalmol* 46:336–340.
- Thiselton DL, Alexander C, Morris A, Broooks S, Rosenberg T, Eiberg H, Kjer B, Kjer P, Bhattacharya SS, Votruba M. 2001. A frameshift mutation in exon 28 of the *OPA1* gene explains the high prevalence of dominant optic atrophy in the Danish population: evidence for a founder effect. *Hum Genet* 109:498–502.
- Toomes C, Marchbank NJ, Mackey DA, Craig JE, Newbury-Ecob R, Bennet CP, Vize CJ, Desai SP, Black GCM, Patel N, Teimory M, Markham AF, Inglehearn CF, Churchill AJ. 2001. Spectrum, frequency and penetrance of *OPA1* mutations in dominant optic atrophy. *Hum Mol Genet* 10:1369–1378.
- Yamada T, Hayasaka S, Matsumoto M, Budu, Esa T, Hayasaka Y, Endo M, Nagaki Y, Fujiki K, Murakami A, Kanai A. 2003. *OPA1* gene mutations in Japanese patients with bilateral optic atrophy unassociated with mitochondrial DNA mutations at nt 11778, 3460, and 14484. *Jpn J Ophthalmol* 47:409–11.

3.2.2 Troisième article : « Dépistage moléculaire de 980 cas suspectés de neuropathie optique héréditaire et publication de 77 nouvelles mutations du gène *OPA1* »

Molecular Screening of 980 Cases of Suspected Hereditary Optic Neuropathy with a Report on 77 Novel *OPA1* Mutations



Marc Ferré,^{1,2,3*} Dominique Bonneau,^{1,2,3} Dan Milea,^{4,5} Arnaud Chevrollier,^{1,3} Christophe Verny,^{2,6} Hélène Dollfus,^{7,8,9} Carmen Ayuso,¹⁰ Sabine Defoort,^{11,12,13} Catherine Vignal,¹⁴ Xavier Zanolonghi,^{14,15} Jean-Francois Charlin,^{16,17} Josseline Kaplan,^{18,19,20} Sylvie Odent,^{16,21} Christian P. Hamel,^{22,23} Vincent Procaccio,^{2,3,24,25} Pascal Reynier,^{1,2,3} and Patrizia Amati-Bonneau^{1,3}

¹INSERM, U694, Angers, F-49000, France; ²Université d'Angers, Faculté de Médecine, Angers, F-49000, France; ³CHU d'Angers, Département de Biochimie et Génétique, Angers, F-49000, France; ⁴Glostrup Hospital, Department of Ophthalmology, Glostrup, DK-2600, Denmark; ⁵University of Copenhagen, Copenhagen, DK-1165, Denmark; ⁶CHU d'Angers, Département de Neurologie, Angers, F-49000, France; ⁷INSERM, Equipe Avenir 3439, Strasbourg, F-67000, France; ⁸Université Louis Pasteur-Strasbourg, Faculté de Médecine, Laboratoire de Génétique Médicale, Strasbourg, F-67000, France; ⁹CHRU de Strasbourg, Service de Génétique Médicale, Strasbourg, F-67000, France; ¹⁰Fundación Jiménez Díaz, Servicio de Genética, CIBERER, Madrid, Spain; ¹¹CNRS, UMR 8160, Lille, F-59000, France; ¹²Université de Lille 2, Lille, F-59000, France; ¹³CHRU de Lille, Hôpital Roger Salengro, Service d'Explorations Fonctionnelles de la Vision, Lille, F-59000, France; ¹⁴Fondation Rothschild, Département d'Ophtalmologie, Paris, F-75019, France; ¹⁵Clinique Sourdirille, Laboratoire d'Explorations Fonctionnelles de la Vision, Nantes, F-44000, France; ¹⁶Université de Rennes 1, Faculté de Médecine, Rennes, F-35000, France; ¹⁷CHU de Rennes, Service d'Ophtalmologie, Rennes, F-35000, France; ¹⁸INSERM, U781, Unité de Recherches Génétique et Epigénétique des Maladies Métaboliques, Neurosensorielles et du Développement, Paris, F-75014, France; ¹⁹Université Paris Descartes, Faculté de Médecine, Paris, F-75014, France; ²⁰AP-HP, Groupe Hospitalier Necker, Service de Génétique Médicale, Paris, F-75014, France; ²¹CHU de Rennes, Département de Médecine de l'Enfant et de l'Adolescent, Rennes, F-35000, France; ²²CHRU de Montpellier, Montpellier, F-34000, France; ²³Université Montpellier1 et Montpellier2, Institut des Neurosciences, Montpellier, F-34000, France; ²⁴CNRS, UMR6214, F-49000 Angers, France; ²⁵INSERM, U771, F-49000 Angers, France.

*Correspondence to Marc Ferré, Laboratoire de Biochimie et Biologie Moléculaire, CHU d'Angers, 4 rue Larrey, F-49933 Angers Cedex 9, France. Phone: +33 241 355 886; Fax: +33 241 354 017; E-mail: MaFerre@chu-angers.fr

Contract grant sponsor: Institut National de la Santé et de la Recherche Médicale (INSERM); GIS-Institut des Maladies Rares; Centre Hospitalier Universitaire d'Angers; Université d'Angers, Forskningsrådet for Sundhed og Sygdom, Denmark; and the following patients' associations: "Retina France", "Ouvrir les Yeux", and "Union Nationale des Aveugles et Déficiants Visuels".

Communicated by Peter Humphries

ABSTRACT: We report the results of molecular screening in 980 patients carried out as part of their work-up for suspected hereditary optic neuropathies. All the patients were investigated for Leber's hereditary optic neuropathy (LHON) and autosomal dominant optic atrophy (ADOA), by searching for the ten primary LHON-causing mtDNA mutations and examining the entire coding sequences of the *OPA1* and *OPA3* genes, the two genes currently identified in ADOA. Molecular defects were identified in 440 patients (45% of screened patients). Among these, 295 patients (67%) had

Received 9 December 2008; accepted revised manuscript 16 March 2009.

an *OPA1* mutation, 131 patients (30%) had an mtDNA mutation, and 14 patients (3%), belonging to three unrelated families, had an *OPA3* mutation. Interestingly, *OPA1* mutations were found in 157 (40%) of the 392 apparently sporadic cases of optic atrophy. The *eOPA1* locus-specific database now contains a total of 204 *OPA1* mutations, including 77 novel *OPA1* mutations reported here. The statistical analysis of this large set of mutations has led us to propose a diagnostic strategy that should help with the molecular work-up of optic neuropathies. Our results highlight the importance of investigating LHON-causing mtDNA mutations as well as *OPA1* and *OPA3* mutations in cases of suspected hereditary optic neuropathy, even in absence of a family history of the disease. © 2009 Wiley-Liss, Inc.

KEY WORDS: hereditary optic atrophy, mitochondria, autosomal dominant optic atrophy, ADOA, optic atrophy 1, *OPA1*, Leber's hereditary optic atrophy, LHON, optic atrophy 3, *OPA3*

INTRODUCTION

Hereditary optic atrophy is a generic term referring to a heterogeneous group of genetic disorders that affect retinal ganglion cells and the optic nerve, leading to impaired vision. The commonest forms of these disorders are autosomal dominant optic atrophy (ADOA; MIM# 165500) and Leber's hereditary optic atrophy (LHON; MIM# 53500).

ADOA, also known as Kjer's disease (Kjer, 1959), is the most common hereditary optic neuropathy, with a prevalence of 1/12,000–1/50,000 (Eiberg et al., 1994; Kjer et al., 1996; Lyle, 1990). The disease is generally diagnosed in early childhood and is characterized by a progressive bilateral decrease of visual acuity, blue-yellow dyschromatopsia or generalized color vision deficits, variable centrocecal, central or paracentral visual field defects, and temporal or diffuse optic nerve pallor with optic disc excavation (Kerrison, 2001). ADOA is associated with a marked intra- and interfamilial clinical variability and an incomplete penetrance, estimated at about 90% in the familial forms of the disease (Cohn et al., 2007). Mutations in the optic atrophy 1 gene (*OPA1*; MIM# 605290), located on chromosome 3q28–q29, are responsible for about 60–80% of the cases of ADOA (Alexander et al., 2000; Amati-Bonneau et al., 2005; Delettre et al., 2000; Kjer et al., 1983). *OPA1* encodes a mitochondrial dynamin-related GTPase, an ubiquitously expressed protein (Delettre et al., 2002; Olichon et al., 2006), which is anchored to the mitochondrial inner membrane. The *OPA1* protein is involved in multiple functions, playing a key role in the fusion of mitochondria and thus in the organization of the mitochondrial network (Delettre et al., 2002; Olichon et al., 2006). The other functions of the *OPA1* protein are related to oxidative phosphorylation and maintenance of membrane potential (Amati-Bonneau et al., 2005; Lodi et al., 2004; Olichon et al., 2003), maintenance of mtDNA (Amati-Bonneau et al., 2007; Hudson et al., 2008), and organization of cristae and control of mitochondrial apoptosis through the compartmentalization of cytochrome c (Frezza et al., 2006; Olichon et al., 2003).

Phenotype-genotype studies of optic atrophies have led to the identification of severe phenotypes, the so-called "ADOA plus" phenotypes, which associate *OPA1* mutations with syndromic forms of optic atrophy including sensorineural deafness (ADOAD; MIM# 125250) (Amati-Bonneau et al., 2003; Shimizu et al., 2003), and ptosis and myopathy (Amati-Bonneau et al., 2007; Hudson et al., 2008; Meire et al., 1985; Payne et al., 2004; Treft et al., 1984).

More than a hundred *OPA1* mutations, often family-specific, have been reported (see <http://lbbma.univ-angers.fr/eOPA1>) (Ferre et al., 2005). Most of these mutations result in the loss of function of the mutated allele involved in most cases of ADOA, supporting the notion that haploinsufficiency is the most likely pathomechanism of the disease (Pesch et al., 2001).

LHON, caused by specific mtDNA mutations, is a maternally inherited disorder (Giles et al., 1980; Wallace et al., 1988) characterized by acute or subacute, often sequential, visual loss, usually occurring between the ages of 18 and 35, and preferentially affecting men. The prevalence of LHON is estimated at 1/50,000 and, like ADOA, the disease is associated with incomplete penetrance and variable expressivity. In addition to optic atrophy, some patients affected with LHON may have other neurological symptoms related to lesions of the white matter of the central nervous system (Nikoskelainen et al., 1995). More than 95% of the patients carry one of the three primary LHON-causing mtDNA mutations at nucleotide positions 11778, 3460 and 14484 in genes encoding subunits of

the respiratory chain complex I. Incidentally, these mutations have been shown to seriously impair ATP synthesis, which depends on complex I activity, in trans-mitochondrial cybrid models (Carelli et al., 2004).

Other rarer forms of hereditary optic atrophy include autosomal dominant optic atrophies: *OPA4* (MIM# 605293) and *OPA5* (MIM# 610708), mapped to chromosome 18q12.2–q12.3 (Kerrison et al., 1998) and chromosome 22q12–q13 (Barbet et al., 2005) respectively, X-linked optic atrophy (XLAO), mapped to locus Xp11.4–p11.21 (*OPA2*; MIM# 311050) (Assink et al., 1997), and autosomal recessive optic atrophy (AROA), for which a first locus (*OPA6*; MIM# 258500) has been mapped to chromosome 8q21–q22 (Barbet et al., 2003). However, none of the genes responsible for *OPA2*, *OPA4*, *OPA5*, and *OPA6* optic atrophies has been identified so far.

Finally, more than 15 disorders, mostly inherited in the autosomal recessive mode, combine optic atrophy and extraocular anomalies. Among these syndromic optic atrophies, type III 3-methylglutaconic aciduria (MGA; MIM# 258501), also known as the Costeff syndrome (Costeff et al., 1989), is a form of early-onset bilateral optic atrophy associated with late-onset spasticity, extrapyramidal signs, and cognitive deficit. Urinary excretions of 3-methylglutaconic acid and increased plasma 3-methylglutaric acid level are the hallmarks of MGA (Costeff et al., 1993). The gene responsible for type III MGA, namely optic atrophy 3 (*OPA3*; MIM# 606580) is located in 19q13.2–q13.3 (Anikster et al., 2001). *OPA3* encodes a mitochondrial protein of unknown function located on the inner membrane (Da Cruz et al., 2003). We have found that dominant heterozygous mutations in *OPA3* may be also responsible for an autosomal dominant form of optic atrophy associated with cataract (ADOAC) but without the presence of organic aciduria and neurological involvement (Reynier et al., 2004).

OPA1 and *OPA3* genes, responsible for ADOA and ADOAC, respectively, are currently the only nuclear genes identified in dominantly inherited optic atrophy. Interestingly, three forms of hereditary optic atrophies, for which the molecular bases are known, involve either the mtDNA, as in LHON, or nuclear genes coding for mitochondrial proteins, as in *OPA1* and *OPA3* optic atrophies. This indicates that altered mitochondrial functions play an essential role in the pathogenesis of optic neuropathies. Although the precise molecular mechanisms involved in the pathomechanism of hereditary optic atrophies are still not well understood, we have recently demonstrated that ADOA, ADOAC and LHON share a common coupling defect of oxidative phosphorylation (Chevrollier et al., 2008).

Here we report the results of the molecular screening of 980 patients referred over a four-year period to our laboratory for the diagnosis of hereditary optic neuropathies and propose a general strategy for the molecular screening of these disorders.

PATIENTS AND METHODS

Patients

Nine hundred and eighty unrelated patients undergoing a diagnostic evaluation of suspected hereditary optic neuropathy were recruited from different medical centers in France and Spain during the period 2003–2007. Half of the patients were referred by geneticists, the other half being enrolled by experienced ophthalmologists and neuro-ophthalmologists, as part of the work-up for unexplained bilateral optic neuropathy. Among these patients, 588 (60%) had a family history of optic neuropathy, while 392 patients (40%) had no obvious family history of hereditary optic neuropathy. In this latter group, an extensive initial work-up ruled out the usual compressive, glaucomatous, inflammatory, ischemic, toxic, nutritional causes of bilateral optic neuropathy.

Methods

Blood samples were taken from patients and members of their families after obtaining informed consent. Genomic DNA was extracted from the blood samples using the High Pure PCR Template Preparation Kit (Roche Diagnosis, Mannheim, Germany).

Screening strategy

Familial cases. In the familial cases, *OPA1* was tested first, whenever an autosomal dominant inheritance was obvious, e.g. father-to-son transmission. When no *OPA1* mutation was found in these patients, the coding regions of *OPA3* were subsequently sequenced. In the absence of evidence of father-to-son transmission, the 10 primary

LHON-causing mtDNA mutations were first screened; this was followed up by the sequencing of *OPA1* and *OPA3*.

Sporadic cases. In patients without any family history of optic neuropathy, the screening of the LHON mutations and the sequencing of *OPA1* were performed concomitantly. All the negative cases were then analyzed by the sequencing of *OPA3*.

Screening the *OPA1* and *OPA3* genes

Thirty primer sets (available on request from the authors) were used for amplifying the 30 coding exons of the *OPA1* gene, including the exon-intron junctions. Four primer sets were used for amplifying the two coding exons of the *OPA3* gene, and fourteen PCR reactions were carried out.

PCR reactions were carried out under standard conditions with 100 ng of genomic DNA in a 50 μ l volume: 1.5 mM MgCl₂, 75 mM Tris-HCl (pH 9 at 25°C), 20 mM (NH₄)₂SO₄, 0.01% Tween 20, 50 pmol of each primer, 200 μ M of each dNTP and 2 units of Hot GoldStar (Eurogentec, Seraing, Belgium) as follows: one cycle for 4 min at 94°C followed by 30 cycles at 94°C for 30 s, 58°C for 30 s, 72°C for 1 min, and one last cycle at 72°C for 5 min. The purified PCR products were sequenced using a Ceq2000/8000 DNA sequencer (CEQ DTCS-Quick Start Kit, Beckman Coulter, Fullerton, CA, USA).

In order to analyze splicing mutations, total RNA was extracted from blood samples (Trizol, Invitrogen Life Technologies, Groningen, The Netherlands). cDNA was obtained using poly-T as a primer and SuperScript RNA polymerase as described by the manufacturer (Invitrogen, Cergy Pontoise, France).

***OPA1* Locus-Specific Database**

We have set up *eOPA1*, a locus-specific database of *OPA1* gene mutations and nonpathogenic sequence variants (<http://lbbma.univ-angers.fr/eOPA1>) (Ferre et al., 2005). Following the general policy of *eOPA1*, each *OPA1* mutation discovered is referenced only once, with a specification of the first report.

Mitochondrial DNA screening

The ten primary mtDNA mutations, most commonly involved in LHON as reported in MITOMAP (<http://www.mitomap.org>) (Ruiz-Pesini et al., 2007), were screened.

Nomenclature of mutations

The nomenclature of the mutations described follows the recommendations of the Human Genome Variation Society: <http://www.hgvs.org/mutnomen> (den Dunnen and Antonarakis, 2000; den Dunnen and Paalman, 2003).

The mtDNA mutations are described according to the Revised Cambridge Reference Sequence (rCRS) of the Human mtDNA (RefSeq: AC_000021.2). The *OPA1* mutations are described according to the *OPA1* transcript variant 1 (exon 4/not 4b and 5b; RefSeq: NM_015560.1). Variant 1, the original transcript identified, contains 29 exons and encodes an isoform (1) of 960 amino acids (aa). *OPA3* mutations are described according to *OPA3* transcript variant 2 (RefSeq: NM_025136.2). Variant 2, the original transcript identified, contains 2 exons and encodes an isoform (b) of 179 aa. Nucleotide numbering of nuclear genes reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, according to journal guidelines. The initiation codon is codon 1.

Statistical analysis

A total of 204 *OPA1* mutations, including the 77 new mutations reported here, were subjected to in-depth analysis of distribution and density. A mutation was considered intronic if at least one mutated base was located in the intron. The mutation density was calculated by sliding a 289-base window, corresponding to 10% of the size of the coding sequence (CDS), along the entire *OPA1* coding sequence, and considering only single base-pair substitutions (excluding all deletions and insertions).

RESULTS

The strategy of molecular screening allowed us to detect mutations in 440 patients (45% of screened patients). More particularly, an *OPA1* mutation was identified in 295 patients, an mtDNA mutation in 131 patients, and an *OPA3* mutation in 14 patients belonging to three unrelated families (Fig. 1).

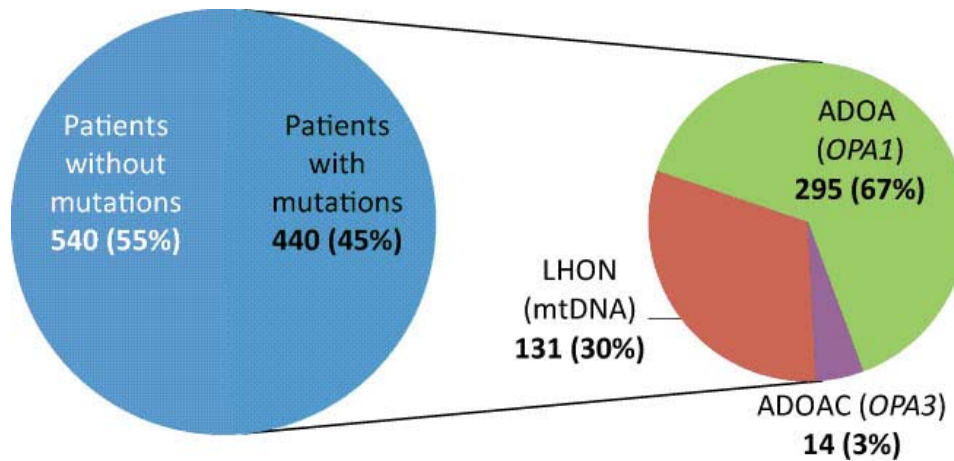


Figure 1. Results of the molecular screening of 980 patients with hereditary optic neuropathies. ADOA: autosomal dominant optic atrophy; ADOAC: optic atrophy associated with cataract; LHON: Leber's hereditary optic neuropathy; mtDNA: mitochondrial DNA.

MtDNA mutations

A primary LHON mtDNA mutation was found in 131 patients (13% of the optic atrophy cases), of which 13 were apparently sporadic cases (10% of the LHON cases). The m.11778G>A mutation was found in 94 patients (72% of the LHON cases), the m.3460A>G mutation in 16 patients (12% of the LHON cases), and the m.14484T>C mutation in 15 patients (11% of the LHON cases). Taken together, these three mutations account for 95% of the patients with LHON mutations in this study. Six patients had other rare primary LHON mutations: two cases each of the m.14568C>T and the m.14482C>A mutations, and one case each of the m.4171C>A and the m.14459G>A mutations (Fig. 2a). The sex ratio of LHON mtDNA mutations was significantly higher for males than for females: 94 males (72 %) vs. 37 females (28%). Six patients presented the so-called "LHON plus" phenotype (Harding et al., 1992), i.e. optic atrophy associated with additional neurological features. Three patients presented with optical atrophy plus a multiple-sclerosis like syndrome, two had cerebellar ataxia and ophthalmoplegia and one suffered from spastic paraparesis (Chevrollier et al., 2008).

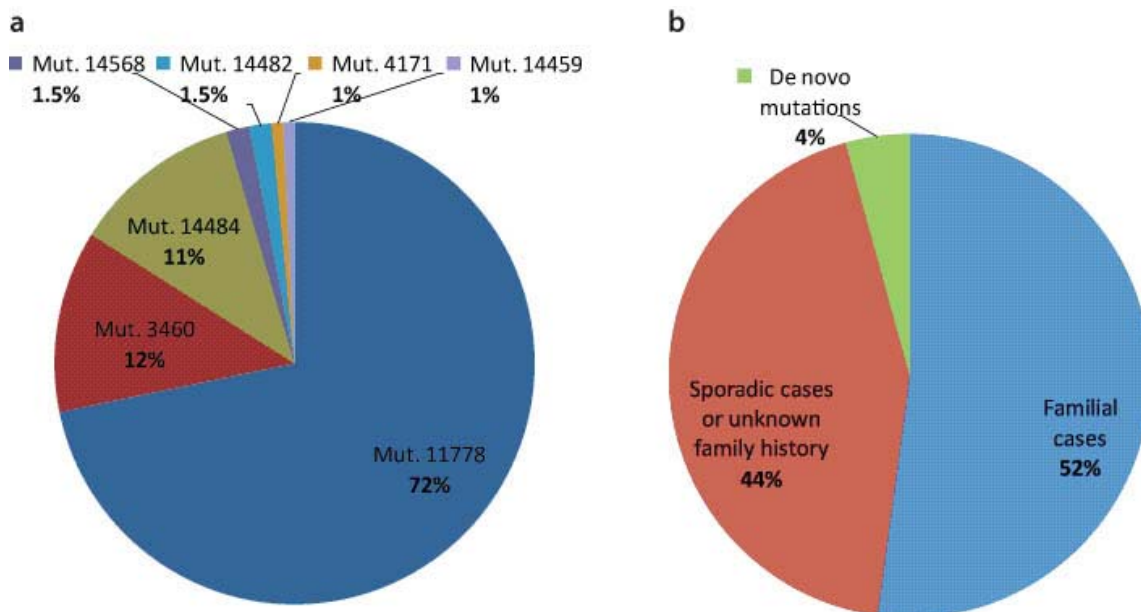


Figure 2. Proportion of primary Leber's hereditary optic neuropathy (LHON)-causing mtDNA mutations among 131 patients. **b:** Proportion of patients bearing *OPA1* mutations with regard to familial data.

OPA1 mutations

In the population studied, an *OPA1* mutation was found in 295 patients (30% of the optic atrophy cases). More specifically, 153 patients (52%) had a familial history, and 142 patients (48%) were apparently sporadic cases. However, we were able to demonstrate that the mutation was *de novo* in only 12 of the apparently sporadic cases (4%) (Fig. 2b). Parental samples were analyzed and paternity was tested in all these families to confirm these *de novo* mutations. In the remaining 130 sporadic cases, DNA samples of none of the parents were available and it was impossible to perform further familial analyses. Thirty patients (10% of the cases with *OPA1* mutations) presented a so-called “OPA1 plus” phenotype. Ten patients from six families bore the p.R445H mutation in exon 14 and were all affected by optic atrophy and sensorineural deafness. The sex ratio of patients with *OPA1* mutations did not show a significant gender difference: 133 females (45%) vs. 162 males (55%).

We have identified 77 novel *OPA1* mutations (Table 1) including 29 splice variants (37.5%), 22 missense mutations (28.5%), 13 nonsense mutations (17%), 10 frameshifts causing premature truncation of the *OPA1* protein (13%), two deletions (2.5%) and one duplication (1.5%). No such mutations were found in a panel of 1,000 control chromosomes. All the missense mutations were shown to affect a highly conserved amino acid. These mutations were scattered along the gene: five mutations in the basic domain (mitochondrial target peptide, exons 1–3) (6.5%), three mutations in the coiled coil domain (exons 5–7) (4%), 25 mutations in the GTPase domain (exons 8–15) (32.5%), 33 mutations in the dynamin domain (exons 16–24) (43%), 7 mutations in exons 25–26 (9%), and four mutations in the 3' end of the coding region probably corresponding to a GTPase effector domain (GED, exons 27–28) (5%).

Table 1. Seventy-seven novel mutations of the *OPA1* gene, grouped according to the exons involved (grey/white lines).

Name	Consequence	Location
c.190delT	p.S64LfsX2	Exon 2
c.284C>T	p.T95M	Exon 2
c.305A>G	p.Y102C	Exon 2
c.361C>T	p.Q121X	Exon 3
c.448+2T>G	Splicing defect	Intron 3
c.665T>C	p.L222P	Exon 6
c.728T>A	p.L243X	Exon 6
c.784-1G>A	Splicing defect	Intron 7
c.870+1G>T	Splicing defect	Intron 8
c.871-1G>A	Splicing defect	Intron 8
c.877_882delGTGGTT	p.V293_V294del	Exon 9
c.929A>G	p.Q310R	Exon 9
c.983_984+3delAGGTA	p.K328SfsX4	Exon 9-Intron 9
c.1069G>A	p.A357T	Exon 11
c.1140+5G>C	Splicing defect	Intron 11
c.1146A>G	p.I382M	Exon 12
c.1187T>C	p.L396P	Exon 12
c.1212+2T>G	Splicing defect	Intron 12
c.1212+5G>C	Splicing defect	Intron 12
c.1285_1290delCCTAAT	p.P429_N430del	Exon 13
c.1288A>G	p.N430D	Exon 13
c.1213-2A>G	Splicing defect	Intron 13
c.1313-1G>A	Splicing defect	Intron 13
c.1315_1318delGGAT	p.G439LfsX27	Exon 14

Name	Consequence	Location
c.1346C>G	p.T449R	Exon 14
c.1387_1392dupATATTC	p.I463_F464dup	Exon 14
c.1444-1G>C	Loss of 24 aa (splicing defect)	Intron 14
c.1444-1G>T	Splicing defect	Intron 14
c.1447C>T	p.Q483X	Exon 15
c.1459G>A	p.E487K	Exon 15
c.1474_1475delCCinsTA	p.P492X	Exon 15
c.1497delT	p.F499LfsX4	Exon 15
c.1515A>G	p.(=) (splicing defect)	Exon 15
c.1558_1560delGAA	p.E521del (splicing defect)	Exon 16
c.1571delA	p.Q524RfsX6	Exon 16
c.1635C>G	p.S545R	Exon 17
c.1652G>A	p.C551Y	Exon 17
c.1667_1668dupTA	p.R557YfsX53	Exon 17
c.1706-1G>A	Splicing defect	Intron 17
c.1769G>A	p.R590Q	Exon 18
c.1770delG	p.N591MfsX18	Exon 18
c.1770G>T	p.(=) (splicing defect)	Exon 18
c.1770+2T>C	Splicing defect	Intron 18
c.1771-2A>G	Splicing defect	Intron 18
c.1778T>C	p.L593P	Exon 19
c.1833_1836delTACA	p.T612QfsX20	Exon 19
c.1847G>A	p.W616X	Exon 19
c.1847_1847+4delGGTAA	Splicing defect	Exon 19-Intron 19
c.1847+1G>T	Splicing defect	Intron 19
c.1847+1_1847+4delGTAA	Splicing defect	Intron 19
c.1848-1G>A	Splicing defect	Intron 19
c.1848G>A	p.W616X	Exon 20
c.1879A>T	p.R627X	Exon 20
c.1891_1892ins41	Splicing defect	Exon 20
c.1892_1893delAT	p.H631RfsX3	Exon 20
c.1937C>T	p.S646L	Exon 20
c.2013+1G>T	Splicing defect	Intron 20
c.2059_2060delGA	p.E687TfsX7	Exon 21
c.2197C>T	p.R733X	Exon 22
c.2303G>A	p.G768D	Exon 23
c.2341C>T	p.R781W	Exon 23
c.2468C>A	p.S823Y	Exon 24
c.2470C>T	p.R824X	Exon 24
c.2496+1G>T	Splicing defect	Intron 24
c.2496+2T>C	Splicing defect	Intron 24
c.2496+4A>G	Splicing defect	Intron 24
c.2569C>T	p.R857X	Exon 25
c.2590C>T	p.Q864X	Exon 25
c.2614-1G>A	Splicing defect	Intron 25
c.2645G>T	p.R882L	Exon 26
c.2650C>T	p.Q884X	Exon 26
c.2660T>C	p.L887P	Exon 26
c.2707+2T>G	Splicing defect	Intron 26

Name	Consequence	Location
c.2725_2726delAA	p.N909CfsX2	Exon 27
c.2794C>T	p.R932C	Exon 27
c.2818+2T>C	Splicing defect	Intron 27
c.2846T>C	p.L949P	Exon 28

Mutational data are described using the nomenclature of the Human Genome Variation Society (<http://www.hgvs.org/mutnomen>). Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence (human OPA1, RefSeq: NM_015560.1), according to journal guidelines. The initiation codon is codon 1.

Statistical analysis of the entire *OPA1* locus-specific database

Since *eOPA1*, the locus-specific database for *OPA1* mutations, was published (Ferre et al., 2005), the number of registered mutations has more than doubled. Currently, the database contains a total of 204 *OPA1* mutations, including the 77 novel mutations listed in Table 1. The *OPA1* mutations are spread all through the gene-coding sequence (Fig. 3a):

- Fourteen of the *OPA1* mutations (7%) are localized in the basic domain (mitochondrial target peptide), 10 (5%) in the coiled coil domain, 76 (37%) in the GTPase domain, 67 (33%) in the dynamin central region, 18 (9%) in exons 25–26, and 19 (9%) in the 3' end of the coding region (putative GED domain);
- Fifty-five of the *OPA1* mutations (27%) are missense mutations, 55 (27%) are splice variants, 44 (21.5%) are frameshifts causing premature truncation of the OPA1 protein, 34 (16.5%) are nonsense mutations, 10 (5%) are deletions, 4 (2%) are frameshifts without premature truncation of the OPA1 protein, and 2 (1%) are duplications.

The average density of substitutions for the whole gene is 9 substitutions per 289 CDS bases. Interestingly, the highest density of substitutions is found in the two main functional domains (the GTPase domain and the dynamin central region), mainly between exon 8 and exon 19, with a maximum of 17 substitutions per 289 CDS bases.

The *eOPA1* locus-specific database contains 55 intronic mutations (27% of the *OPA1* mutations), including 48 intronic point mutations (Fig. 3b): the most extreme upstream point mutation position registered is -10 (mutation c.2708-10C>G); the most extreme downstream point mutation position registered is +6 (mutation c.2818+6T>C).

OPA3 mutations

We studied three independent families in which 14 affected patients (12 women and two men) presented a bilateral optic atrophy and posterior cataract inherited as an autosomal dominant trait. In addition, we examined 13 unaffected subjects in these families.

Two different *OPA3* mutations were found in exon 2: the c.277G>A (p.G93S) mutation and the c.313C>G (p.Q105E) mutation (Reynier et al., 2004). Both mutations segregated with the disease in each family and were absent in healthy relatives and in 400 control chromosomes.

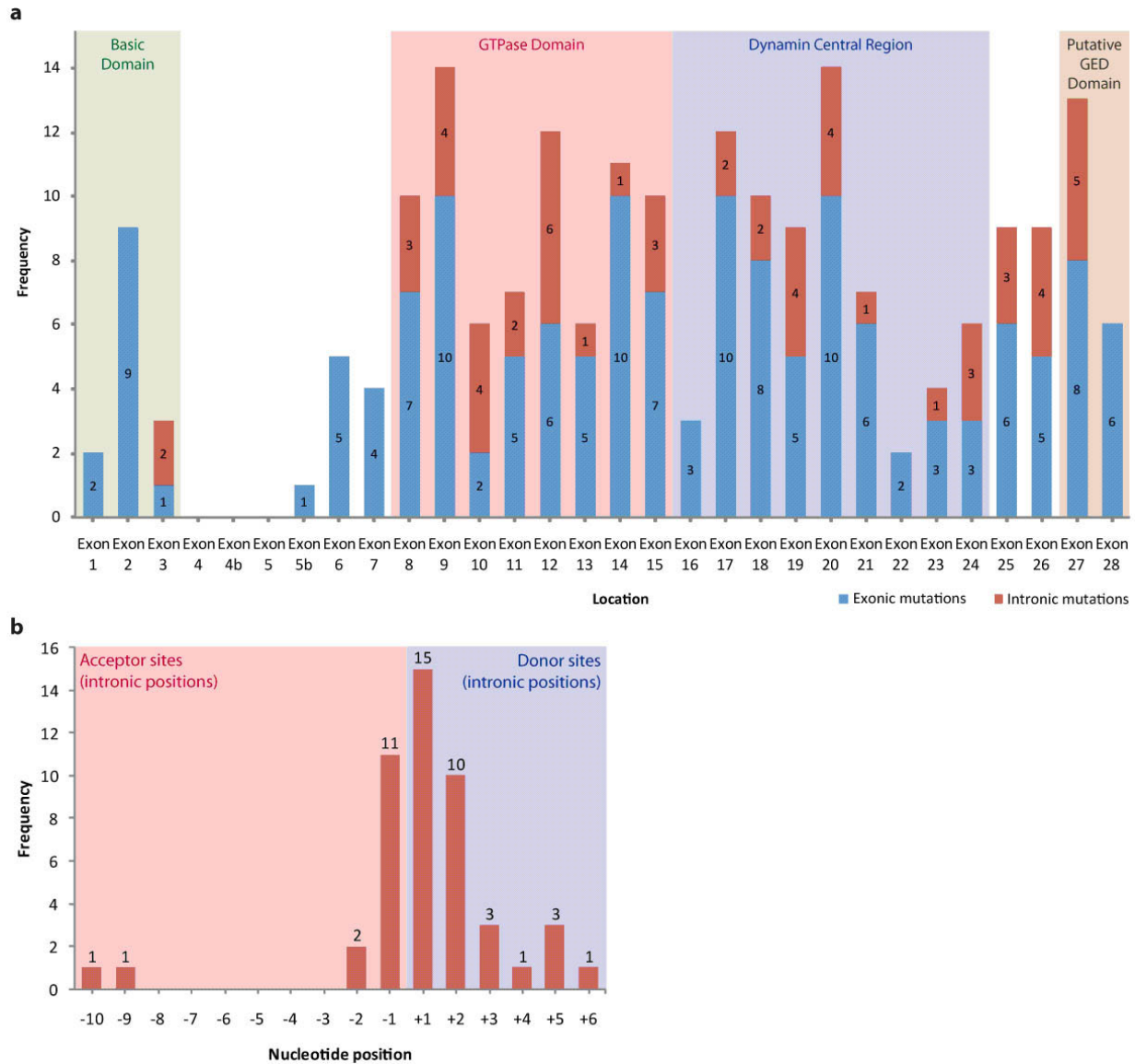


Figure 3. Data obtained from the *eOPA1* locus-specific database (Ferre et al., 2005), updated in December 2008. **a:** Distribution of the mutations in the *OPA1* gene according to the exons involved (blue bars); the mutations in the intronic neighbourhood of the exons are indicated as red bars. **b:** Distribution of disease-causing single base-pair substitutions in splice-sites (intronic positions only) of the *OPA1* gene. GED: GTPase effector domain.

DISCUSSION

We report the results of molecular analyses performed in 980 patients referred to our laboratory over the last four years with either the clinical diagnosis or the suspicion of hereditary optic atrophy. To our knowledge, this the largest series of patients screened for the currently admitted genetic causes of optic atrophy, i.e. mutations in the *OPA1* and *OPA3* genes, and the LHON-causing mtDNA mutations.

Information obtained from patients' pedigrees might be expected to help in determining the appropriate molecular screening strategy. For instance, the compatibility of the pedigree with maternal inheritance would suggest that the mtDNA be tested first whereas in the presence of a father-to-son transmission, the *OPA1* gene analysis would be given priority. However, this approach proved of little help in our cohort of patients since about half of the patients (48%) carrying an *OPA1* mutation (Fig. 2b) and 10% of the patients carrying a primary LHON-causing mtDNA mutation, were either definitely sporadic cases or had an unknown family history. This highlights

the usefulness of performing molecular genetic analyses in patients with bilateral optic atrophy even in absence of a known family history.

ADOA was initially described (Kjer, 1959) as a relatively mild ophthalmologic disorder starting during childhood and slowly progressing thereafter. The sequencing of *OPAI* has now brought to light a much wider clinical spectrum including severe forms of neonatal onset (Baris et al., 2003), acute onset identical to LHON (Bonneau, et al., manuscript in preparation), late onset (Johnston et al., 1999), association of optic atrophy with extra-ocular symptoms (Amati-Bonneau et al., 2007), and optic atrophy with spontaneous visual recovery (Cornille et al., 2008). However, the phenotypical expression of *OPAI* mutations is highly variable, even within a given family, and genotype-phenotype correlations are very scarce. The only strong correlation revealed by our study is that of optic atrophy associated with sensorineural deafness and the presence of the p.R445H mutation, which was recurrently found in six families. However, another mutation, namely c.2848_2849delGA, has recently been found in patients with optic atrophy associated with deafness (Chen et al., 2007), minimizing the specificity of the p.R445H mutation.

The best indication for a strategy of *OPAI* screening is given by the analysis of the frequency of mutations, exon by exon (Fig. 4). The data obtained from the 295 probands in this study shows that 28% of the mutations are located in exon 27. More specifically, the c.2708_2711delTTAG mutation (exon 27), present in 17% of the probands, is by far the most frequent mutation. Exons 8 and 15, which code for the GTPase domain and contain 17.5% and 11% of the *OPAI* mutations respectively, are the two other exons most frequently involved. These observations suggest that exons 27, 8 and 15 may be regarded as mutational hot spots that should be given priority during *OPAI* screening.

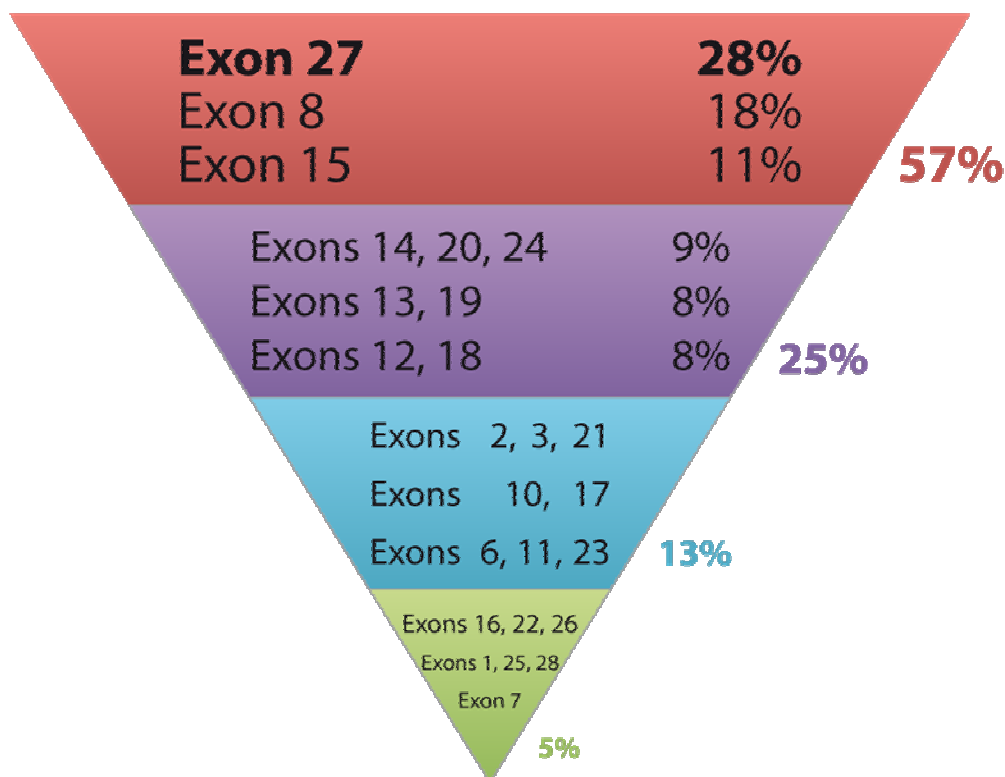


Figure 4. Frequencies of *OPAI* mutations in 295 patients with hereditary optic atrophy, determined exon by exon and arranged in decreasing order of importance from the apex to the base of the triangle. Interestingly, more than 57% of the mutations are concentrated in exons 27, 8 and 15.

The analysis of the 204 mutations, now registered in the *eOPAI* locus-specific database, provides further interesting information regarding *OPAI* screening. Firstly, contrary to earlier reports (Delettre et al., 2002; Thiselton et al., 2002), *OPAI* mutations are found not only in the GTPase domain of *OPAI* but are spread over the

entire coding sequence. Secondly, the intronic regions, which contain about a quarter of the mutations, need to be systematically explored, including at least ten nucleotides downstream from each exon. Thirdly, the alternate splice exons (exons 4, 4b and 5b) will also require to be analyzed since we have recently found that a mutation in exon 5b could be responsible for bilateral optic neuropathy, with spontaneous visual recovery six months after the appearance of the first symptoms (Cornille et al., 2008).

Although our protocol included the extensive analysis of the *OPA1* and *OPA3* genes as well as the search for the ten primary LHON-causing mtDNA mutations, a molecular defect was identified in only about half of the patients. We are currently trying to improve the efficiency of the screening in negative cases by searching for large-scale *OPA1* deletions and novel LHON mtDNA mutations by means of the more sophisticated methods now available, such as the Surveyor™ Nuclease assay (Bannwarth et al., 2008). However, the large proportion of negative results probably reflects the fact that hereditary optic atrophies are genetically heterogeneous disorders with at least four other loci identified to date (Assink et al., 1997; Barbet et al., 2003; Barbet et al., 2005; Kerrison et al., 1998). Moreover, the fact that certain sporadic optic atrophies may be phenocopies caused by non-genetic factors should be borne in mind.

Finally, on the bases of the findings reported here, we propose a flowchart (Fig. 5) to guide the molecular diagnostic work-up of patients suspected of suffering from hereditary optic neuropathy.

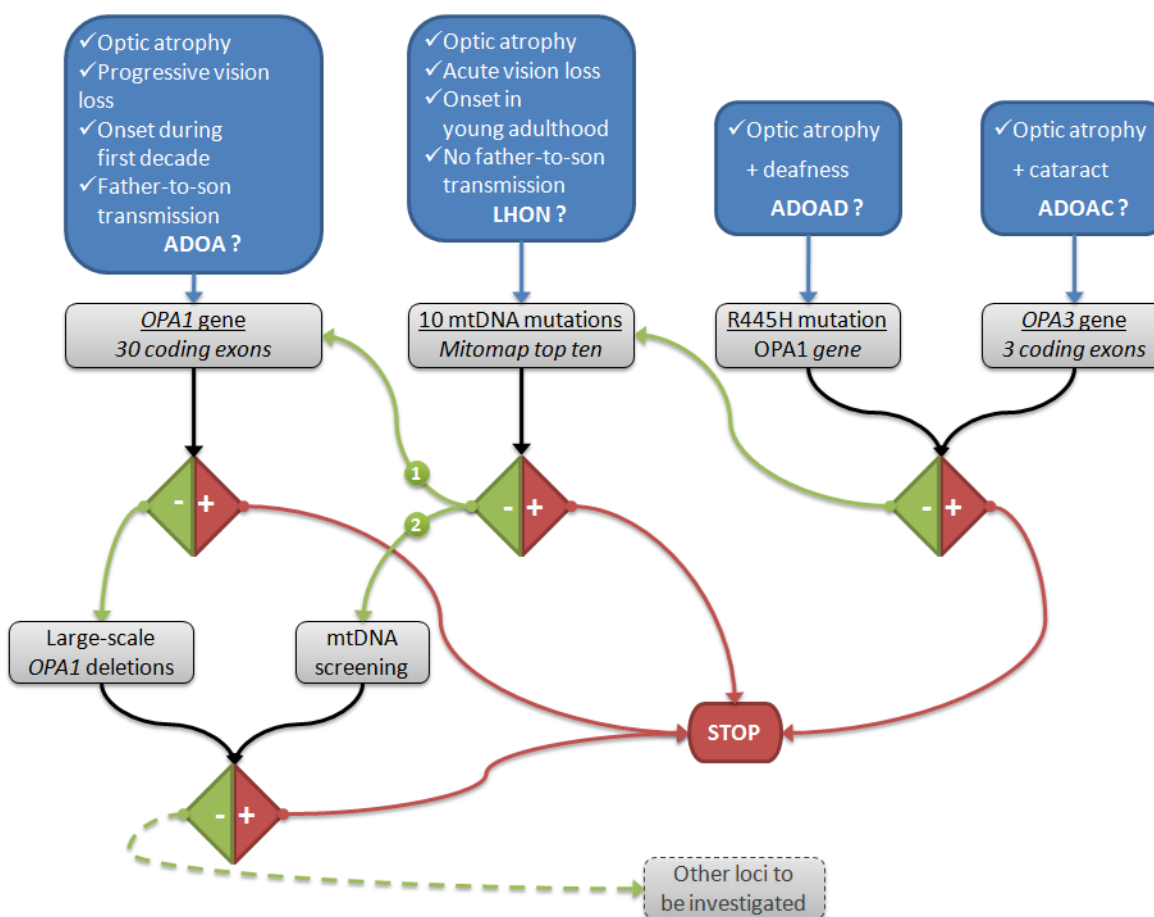


Figure 5. Flowchart for the molecular diagnosis of hereditary optic neuropathies. Blue boxes indicate the clinical signs of various hereditary optic atrophies; grey boxes indicate the suggested analysis. In the decision box, the red “+” sign indicates a positive result; and the green “-” sign, a negative result. The red “STOP” box corresponds to the end point of the diagnostic process. The arrow labeled “1” indicates the analysis to be carried out first, followed by the analysis indicated by the arrow labeled “2” in case of a negative result. The dashed arrow corresponds to future investigations of as yet undiscovered optic atrophy genes. ADOA: autosomal dominant optic atrophy; LHON: Leber’s hereditary optic atrophy; ADOAD: optic atrophy associated with sensorineural deafness; ADOAC: optic atrophy associated with cataract.

ACKNOWLEDGMENTS

We are grateful to all the technicians of our laboratory for their excellent technical assistance, and to Kanaya Malkani for critical reading and comments on the manuscript. We also thank all the clinicians who collaborate with our center. This work was partly supported by the INSERM, the GIS-Institut des Maladies Rares, the University Hospital of Angers, the University of Angers, and the following patients' associations: "*Retina France*", "*Ouvrir les Yeux*" and "*Union Nationale des Aveugles et Déficients Visuels*".

REFERENCES

- Alexander C, Votruba M, Pesch UE, Thiselton DL, Mayer S, Moore A, Rodriguez M, Kellner U, Leo-Kottler B, Auburger G and others. 2000. OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28. *Nat Genet* 26(2):211-5.
- Amati-Bonneau P, Guichet A, Olichon A, Chevrollier A, Viala F, Miot S, Ayuso C, Odent S, Arrouet C, Verny C and others. 2005. OPA1 R445H mutation in optic atrophy associated with sensorineural deafness. *Ann Neurol* 58(6):958-63.
- Amati-Bonneau P, Odent S, Derrien C, Pasquier L, Malthiery Y, Reynier P, Bonneau D. 2003. The association of autosomal dominant optic atrophy and moderate deafness may be due to the R445H mutation in the OPA1 gene. *Am J Ophthalmol* 136(6):1170-1.
- Amati-Bonneau P, Valentino ML, Reynier P, Gallardo ME, Bornstein B, Boissiere A, Campos Y, Rivera H, de la Aleja JG, Carroccia R and others. 2007. OPA1 mutations induce mitochondrial DNA instability and optic atrophy 'plus' phenotypes. *Brain*.
- Anikster Y, Kleta R, Shaag A, Gahl WA, Elpeleg O. 2001. Type III 3-methylglutaconic aciduria (optic atrophy plus syndrome, or Costeff optic atrophy syndrome): identification of the OPA3 gene and its founder mutation in Iraqi Jews. *Am J Hum Genet* 69(6):1218-24.
- Assink JJ, Tijmes NT, ten Brink JB, Oostra RJ, Riemsdag FC, de Jong PT, Bergen AA. 1997. A gene for X-linked optic atrophy is closely linked to the Xp11.4-Xp11.2 region of the X chromosome. *Am J Hum Genet* 61(4):934-9.
- Bannwarth S, Procaccio V, Rouzier C, Fragaki K, Poole J, Chabrol B, Desnuelle C, Pouget J, Azulay JP, Attarian S and others. 2008. Rapid identification of mitochondrial DNA (mtDNA) mutations in neuromuscular disorders by using surveyor strategy. *Mitochondrion* 8(2):136-45.
- Barbet F, Gerber S, Hakiki S, Perrault I, Hanein S, Ducroq D, Tanguy G, Dufier JL, Munnich A, Rozet JM and others. 2003. A first locus for isolated autosomal recessive optic atrophy (ROA1) maps to chromosome 8q. *Eur J Hum Genet* 11(12):966-71.
- Barbet F, Hakiki S, Orssaud C, Gerber S, Perrault I, Hanein S, Ducroq D, Dufier JL, Munnich A, Kaplan J and others. 2005. A third locus for dominant optic atrophy on chromosome 22q. *J Med Genet* 42(1):e1.
- Baris O, Delettre C, Amati-Bonneau P, Surget MO, Charlin JF, Catier A, Derieux L, Guyomard JL, Dollfus H, Jonveaux P and others. 2003. Fourteen novel OPA1 mutations in autosomal dominant optic atrophy including two de novo mutations in sporadic optic atrophy. *Hum Mutat* 21(6):656.
- Carelli V, Rugolo M, Sgarbi G, Ghelli A, Zanna C, Baracca A, Lenaz G, Napoli E, Martinuzzi A, Solaini G. 2004. Bioenergetics shapes cellular death pathways in Leber's hereditary optic neuropathy: a model of mitochondrial neurodegeneration. *Biochim Biophys Acta* 1658(1-2):172-9.
- Chen S, Zhang Y, Wang Y, Li W, Huang S, Chu X, Wang L, Zhang M, Liu Z. 2007. A novel OPA1 mutation responsible for autosomal dominant optic atrophy with high frequency hearing loss in a Chinese family. *Am J Ophthalmol* 143(1):186-188.
- Chevrollier A, Guillet V, Loiseau D, Gueguen N, de Crescenzo MA, Verny C, Ferre M, Dollfus H, Odent S, Milea D and others. 2008. Hereditary optic neuropathies share a common mitochondrial coupling defect. *Ann Neurol* 63(6):794-8.
- Cohn AC, Toomes C, Potter C, Towns KV, Hewitt AW, Inglehearn CF, Craig JE, Mackey DA. 2007. Autosomal dominant optic atrophy: penetrance and expressivity in patients with OPA1 mutations. *Am J Ophthalmol* 143(4):656-62.

- Cornille K, Milea D, Amati-Bonneau P, Procaccio V, Zazoun L, Guillet V, El Achouri G, Delettre C, Gueguen N, Loiseau D and others. 2008. Reversible optic neuropathy with OPA1 exon 5b mutation. *Ann Neurol* 63(5):667-71.
- Costeff H, Elpeleg O, Apter N, Divry P, Gadoth N. 1993. 3-Methylglutaconic aciduria in "optic atrophy plus". *Ann Neurol* 33(1):103-4.
- Costeff H, Gadoth N, Apter N, Prialnic M, Savir H. 1989. A familial syndrome of infantile optic atrophy, movement disorder, and spastic paraplegia. *Neurology* 39(4):595-7.
- Da Cruz S, Xenarios I, Langridge J, Vilbois F, Parone PA, Martinou JC. 2003. Proteomic analysis of the mouse liver mitochondrial inner membrane. *J Biol Chem* 278(42):41566-71.
- Delettre C, Lenaers G, Griffoin JM, Gigarel N, Lorenzo C, Belenguer P, Pelloquin L, Grosgeorge J, Turc-Carel C, Perret E and others. 2000. Nuclear gene OPA1, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy. *Nat Genet* 26(2):207-10.
- Delettre C, Lenaers G, Pelloquin L, Belenguer P, Hamel CP. 2002. OPA1 (Kjer type) dominant optic atrophy: a novel mitochondrial disease. *Mol Genet Metab* 75(2):97-107.
- den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 15(1):7-12.
- den Dunnen JT, Paalman MH. 2003. Standardizing mutation nomenclature: why bother? *Hum Mutat* 22(3):181-2.
- Eiberg H, Kjer B, Kjer P, Rosenberg T. 1994. Dominant optic atrophy (OPA1) mapped to chromosome 3q region. I. Linkage analysis. *Hum Mol Genet* 3(6):977-80.
- Ferre M, Amati-Bonneau P, Tourmen Y, Malthiery Y, Reynier P. 2005. eOPA1: an online database for OPA1 mutations. *Hum Mutat* 25(5):423-8.
- Frezza C, Cipolat S, Martins de Brito O, Micaroni M, Beznoussenko GV, Rudka T, Bartoli D, Polishuck RS, Danial NN, De Strooper B and others. 2006. OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion. *Cell* 126(1):177-89.
- Giles RE, Blanc H, Cann HM, Wallace DC. 1980. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 77(11):6715-9.
- Harding AE, Sweeney MG, Miller DH, Mumford CJ, Kellar-Wood H, Menard D, McDonald WI, Compston DA. 1992. Occurrence of a multiple sclerosis-like illness in women who have a Leber's hereditary optic neuropathy mitochondrial DNA mutation. *Brain* 115 (Pt 4):979-89.
- Hudson G, Amati-Bonneau P, Blakely EL, Stewart JD, He L, Schaefer AM, Griffiths PG, Ahlqvist K, Suomalainen A, Reynier P and others. 2008. Mutation of OPA1 causes dominant optic atrophy with external ophthalmoplegia, ataxia, deafness and multiple mitochondrial DNA deletions: a novel disorder of mtDNA maintenance. *Brain* 131(Pt 2):329-37.
- Johnston RL, Sellar MJ, Behnam JT, Burdon MA, Spalton DJ. 1999. Dominant optic atrophy. Refining the clinical diagnostic criteria in light of genetic linkage studies. *Ophthalmology* 106(1):123-8.
- Kerrison JB. 2001. Hereditary optic neuropathies. *Ophthalmol Clin North Am* 14(1):99-107.
- Kerrison JB, Koenekoop RK, Arnould VJ, Zee D, Maumenee IH. 1998. Clinical features of autosomal dominant congenital nystagmus linked to chromosome 6p12. *Am J Ophthalmol* 125(1):64-70.
- Kjer B, Eiberg H, Kjer P, Rosenberg T. 1996. Dominant optic atrophy mapped to chromosome 3q region. II. Clinical and epidemiological aspects. *Acta Ophthalmol Scand* 74(1):3-7.
- Kjer P. 1959. Infantile optic atrophy with dominant mode of inheritance: a clinical and genetic study of 19 Danish families. *Acta Ophthalmol Suppl* 164(Supp 54):1-147.
- Kjer P, Jensen OA, Klinken L. 1983. Histopathology of eye, optic nerve and brain in a case of dominant optic atrophy. *Acta Ophthalmol (Copenh)* 61(2):300-12.
- Lodi R, Tonon C, Valentino ML, Iotti S, Clementi V, Malucelli E, Barboni P, Longanesi L, Schimpf S, Wissinger B and others. 2004. Deficit of in vivo mitochondrial ATP production in OPA1-related dominant optic atrophy. *Ann Neurol* 56(5):719-23.

- Lyle WM. 1990. Genetic risks : a reference for eye care practitioners. Waterloo, Ont., Canada: University of Waterloo Press.
- Meire F, De Laey JJ, de Bie S, van Staey M, Matton MT. 1985. Dominant optic nerve atrophy with progressive hearing loss and chronic progressive external ophthalmoplegia (CPEO). *Ophthalmic Paediatr Genet* 5(1-2):91-7.
- Nikoskelainen EK, Marttila RJ, Huoponen K, Juvonen V, Lamminen T, Sonninen P, Savontaus ML. 1995. Leber's "plus": neurological abnormalities in patients with Leber's hereditary optic neuropathy. *J Neurol Neurosurg Psychiatry* 59(2):160-4.
- Olichon A, Baricault L, Gas N, Guillou E, Valette A, Belenguer P, Lenaers G. 2003. Loss of OPA1 perturbs the mitochondrial inner membrane structure and integrity, leading to cytochrome c release and apoptosis. *J Biol Chem* 278(10):7743-6.
- Olichon A, Guillou E, Delettre C, Landes T, Arnaune-Pelloquin L, Emorine LJ, Mils V, Daloyau M, Hamel C, Amati-Bonneau P and others. 2006. Mitochondrial dynamics and disease, OPA1. *Biochim Biophys Acta* 1763(5-6):500-9.
- Payne M, Yang Z, Katz BJ, Warner JE, Weight CJ, Zhao Y, Pearson ED, Treft RL, Hillman T, Kennedy RJ and others. 2004. Dominant optic atrophy, sensorineural hearing loss, ptosis, and ophthalmoplegia: a syndrome caused by a missense mutation in OPA1. *Am J Ophthalmol* 138(5):749-55.
- Pesch UE, Leo-Kottler B, Mayer S, Jurklics B, Kellner U, Apfelstedt-Sylla E, Zrenner E, Alexander C, Wissinger B. 2001. OPA1 mutations in patients with autosomal dominant optic atrophy and evidence for semi-dominant inheritance. *Hum Mol Genet* 10(13):1359-68.
- Reynier P, Amati-Bonneau P, Verny C, Olichon A, Simard G, Guichet A, Bonnemains C, Malecaze F, Malinge MC, Pelletier JB and others. 2004. OPA3 gene mutations responsible for autosomal dominant optic atrophy and cataract. *J Med Genet* 41(9):e110.
- Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, Yi C, Kreuziger J, Baldi P, Wallace DC. 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res* 35(Database issue):D823-8.
- Shimizu S, Mori N, Kishi M, Sugata H, Tsuda A, Kubota N. 2003. A novel mutation in the OPA1 gene in a Japanese patient with optic atrophy. *Am J Ophthalmol* 135(2):256-7.
- Thiselton DL, Alexander C, Taanman JW, Brooks S, Rosenberg T, Eiberg H, Andreasson S, Van Regemorter N, Munier FL, Moore AT and others. 2002. A comprehensive survey of mutations in the OPA1 gene in patients with autosomal dominant optic atrophy. *Invest Ophthalmol Vis Sci* 43(6):1715-24.
- Treft RL, Sanborn GE, Carey J, Swartz M, Crisp D, Wester DC, Creel D. 1984. Dominant optic atrophy, deafness, ptosis, ophthalmoplegia, dystaxia, and myopathy. A new syndrome. *Ophthalmology* 91(8):908-15.
- Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, Elsas LJ, 2nd, Nikoskelainen EK. 1988. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* 242(4884):1427-30.

3.2.3 Quatrième article : « L'atrophie optique dominante liée à *OPA1* n'est pas fortement influencée par le fond génétique mitochondrial »

Research article

Open Access

OPA1-related dominant optic atrophy is not strongly influenced by mitochondrial DNA background

Denis Pierron^{1,2,3}, Marc Ferré⁴, Christophe Rocher³, Arnaud Chevrollier⁴, Pascal Murail¹, Didier Thoraval², Patrizia Amati-Bonneau⁴, Pascal Reynier⁴ and Thierry Letellier*³

Address: ¹Université Bordeaux 1, Laboratoire d'Anthropologie des Populations du Passé, UMR 5199 PACEA, 33400 Talence, France, ²Institut de Biochimie et Génétique Cellulaires, UMR 5095, CNRS, Université Victor Segalen-Bordeaux 2, 33076 Bordeaux, France, ³INSERM, U688 Laboratoire de Physiopathologie Mitochondriale; Université Victor Segalen-Bordeaux 2, 33076 Bordeaux, France and ⁴INSERM, U694, Angers, F-49933 France; Université d'Angers, Angers, F-49035 France; Département de Biochimie et Génétique, Centre Hospitalier Universitaire, Angers, F-49933 France

Email: Denis Pierron - d.pierron@anthropologie.u-bordeaux1.fr; Marc Ferré - MaFerre@chu-angers.fr; Christophe Rocher - crocher@u-bordeaux2.fr; Arnaud Chevrollier - archevrollier@chu-angers.fr; Pascal Murail - p.murail@anthropologie.u-bordeaux1.fr; Didier Thoraval - didier.thoraval@lbms.u-bordeaux2.fr; Patrizia Amati-Bonneau - PaBonneau@chu-angers.fr; Pascal Reynier - PaReynier@chu-angers.fr; Thierry Letellier* - thierry.letellier@phys-mito.u-bordeaux2.fr

* Corresponding author

Published: 20 July 2009

Received: 30 September 2008

BMC Medical Genetics 2009, 10:70 doi:10.1186/1471-2350-10-70

Accepted: 20 July 2009

This article is available from: <http://www.biomedcentral.com/1471-2350/10/70>

© 2009 Pierron et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Leber's hereditary optic neuropathy (LHON) and autosomal dominant optic atrophy (ADOA) are the most frequent forms of hereditary optic neuropathies. LHON is associated with mitochondrial DNA (mtDNA) mutations whereas ADOA is mainly due to mutations in the OPA1 gene that encodes a mitochondrial protein involved in the mitochondrial inner membrane remodeling. A striking influence of mtDNA haplogroup J on LHON expression has been demonstrated and it has been recently suggested that this haplogroup could also influence ADOA expression. In this study, we have tested the influence of mtDNA backgrounds on OPA1 mutations.

Methods: To define the relationships between OPA1 mutations and mtDNA backgrounds, we determined the haplogroup affiliation of 41 French patients affected by OPA1-related ADOA by control-region sequencing and RFLP survey of their mtDNAs.

Results: The comparison between patient and reference populations did not reveal any significant difference.

Conclusion: Our results argue against a strong influence of mtDNA background on ADOA expression. These data allow to conclude that OPA1 could be considered as a "severe mutation", directly responsible of the optic atrophy, whereas OPA1-negative ADOA and LHON mutations need an external factor(s) to express the pathology (i.e. synergistic interaction with mitochondrial background).

Background

Hereditary optic atrophies are frequently related to mitochondrial dysfunction. The maternally-transmitted Leber's hereditary optic neuropathy (MIM535000) is associated with one of the three main primary point mutations of the mitochondrial genome. With incomplete penetrance, these mutations induce, between 15 and 35 years old, an acute and bilateral loss of visual acuity with a caecocentral scotoma and a dyschromatopsia. A "sister disease" of LHON, autosomal dominant optic atrophy (MIM #165500) is the second most common form of inherited optic neuropathy, with a frequency of 1:12 000 to 1:50 000 [1,2]. This disease is characterized by an insidious onset of bilateral visual impairment in early childhood with moderate to severe loss of visual acuity, temporal optic disc pallor, abnormalities of color vision, and caecocentral visual field scotoma. Electrophysiological and histopathological studies have suggested that the underlying defect is the retinal ganglion cell (RGC) degeneration leading to atrophy of the optic nerve. Until now, four loci have been designated for ADOA [2]. Among these, the most commonly mutated gene is OPA1 (3q28-29) [3,4]. One hundred and seventeen OPA1 gene mutations, mainly family-specific [5], were described with substitutions, deletions and insertions respectively spread throughout the coding sequence of the gene. Most are localized in the GTPase domain and the N-terminus of the protein, whereas the C-terminus is largely spared.

The mitochondrial genetic background is known to influence the expression of LHON [6] since mitochondrial primary mutations G11778A and T14484C showed significant clustering on Caucasian mtDNA haplogroup J. Haplogroup J was also found to be three-fold over-represented in patients with ADOA not related to OPA1 mutations [7].

In order to test the influence of the mitochondrial genetic background on the penetrance of the mutations of the OPA1 gene, we compared the distribution of mtDNA haplogroups between 41 French patients and 1385 individuals of the French population described by Richard and al. [8].

Methods

Patients

Screening for ADOA French patients carrying OPA1 mutations was done by the Department of Biochemistry and genetics in the University Hospital of Angers. Genetic analyses were carried out with the appropriate consent of the patients and in compliance with the Helsinki Declaration. Experimental research has been performed with the ethics approval of the "centre hospitalier universitaire d'Angers" and ethics policy of the French mitochondrial disease network [9]. This laboratory collects samples from all French patients suspected to be affected by ADOA. All

the patients presented typical history of ADOA with insidious, painless and progressive bilateral visual loss, impairment of visual acuity, central scotoma and optic nerve pallor. OPA1 analysis was carried out by sequencing with the appropriate consent of the patients. Among the cohort of individuals diagnosed as carriers of the OPA1 mutations, we have listed 41 distinct maternal lineages.

Analysis of the haplogroups

Each patient's DNA was extracted from blood using standard procedures (phenol/chloroform or Qiagen® QIAamp DNA kit). For each patient, the mitochondrial DNA control region was then amplified using the following primers: L15832 (light chain, nps 15838-15858) and H408 (heavy chain nps 408-429). The results of the amplification were then purified using ExoSAP-IT® technology. Finally, starting from position 16050, at least 725 pb of the mitochondrial DNA control region were double-strand sequenced using an ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer®). The same primers were used for amplification and sequencing, L15832 and H408, together with 2 additional primers: L16200 (light chain, nps 16194-16217) and H263 (heavy chain nps 263-285). Each individual was also tested for the presence of a polymorphism in position 7028 by digestion with the AluI enzyme of an amplicon obtained with primers L6909 (light chain, nps 6890-6909) and H7115 (heavy chain nps 7115-7131).

The analysis made it possible to identify each individual's haplotype and haplogroup markers. The individuals' haplogroup affiliation was determined on the basis of recent studies involving total sequencing of mitochondrial DNA [10-13]. Affiliations to the main European haplogroups were confirmed by standard RFLP tests (UK: 12308HinfI; JT 4216NlaIII; J: 13704BstI; T: 15606AluI; H7025AluI V 4577NlaIII) on diagnostic positions for the haplogroups on the mtDNA coding segment [14,15].

Control Population and Statistics

The distribution of carriers of mutation on OPA1 gene among the haplogroups was compared with frequencies observed in a sample of the French population consisting of 1385 individuals [8]. Each haplogroup was tested as a possible risk factor for developing the pathology against the rest of the population. In turn statistical comparisons were carried out on the largest cluster, using standard "Chi 2" methods without correction; the haplogroup and sub haplogroup were tested by Fisher's exact test. The relative risk estimated for each haplogroup was defined by the "odds ratio" method, with a confidence interval calculated by the Miettinen method [16].

Results

MtDNA sequencing D-loop and RFLP analysis of the coding segment revealed 37 different haplotypes, among the

Table 1: RFLP and Control-Region mtDNA Haplotypes from the French OPA1 mutation carrier

patient	super haplogroup	haplogroup	POLYMORPHISM OBSERVED BETWEEN POSITION 16050–16569 AND 1–205	RFLP
I421	R0	H	16519T>C, 195T>C	-7025Alul
I043	R0	H	16519T>C, 146T>C	-7025Alul
I659	R0	H	16519T>C	-7025Alul
I148	R0	H	16519T>C	-7025Alul
I284	R0	H	16311T>C, 16519T>C	-7025Alul
I090	R0	H	16235A>G, 16291C>T, 16293A>G, 16400C>T	-7025Alul
I428	R0	H	16189T>C, 16194-del, 16519T>C	-7025Alul
I568	R0	H	16188C>T, 16295C>T, 16519T>C, 150C>T	-7025Alul
I344	R0	H	16188C>G, 16189T>C, 16264C>T, 16311T>C, 16519T>C	-7025Alul
I267	R0	H	16129G>A, 16264C>T, 16316A>G, 16519T>C, 195T>C	-7025Alul
I366	R0	H	16086T>C, 16189T>C, 16519 T>C	-7025Alul
I091	R0	HIa	16162A>G, 16274G>A, 16519T>C, 73A>G	-7025Alul
I163	R0	HIb	16183A>C, 16189T>C, 16291C>T, 16356T>C, 16519T>C, 152T>C	-7025Alul
I531	R0	H5	16184C>T, 16304T>C, 16519T>C	-7025Alul
I686	R0	H5	16243T>C, 16304T>C, 152T>C	-7025Alul
I505	R0	H5	16304T>C, 16519T>C, 146T>C	-7025Alul
I054	R0	HI I	16293A>G, 16519T>C	-7025Alul
I032	R0	preV (HV0)	16298T>C, 72T>C	+4577NlaIII, +7025Alul
I620	R0	V (HV0)	16298T>C, 72T>C, 195T>C, 198C>T	+4577NlaIII, +7025Alul
I697	R0	V (HV0)	16298T>C, 16311T>C, 72T>C, 195T>C	-4577NlaIII, +7025Alul
I332	JT	J	16069C>T, 16126T>C, 73A>G, 185G>A	+4216NlaIII; -13704BstOI, +7025Alul
I743	JT	J	16069C>T, 16126T>C, 73A>G, 185G>A	+4216NlaIII; -13704BstOI, +7025Alul
I779	JT	J	16069C>T, 16126T>C, 73A>G, 185G>A	+4216NlaIII; -13704BstOI, +7025Alul
I464	JT	T	16126T>C, 16153G>A, 16294C>T, 16519T>C, 41C>T, 73A>G, 150C>T, 200A>G	+4216NlaIII; +13704BstOI, +7025Alul
I099	JT	T1	16126T>C, 16163A>G, 16186C>T, 16189T>C, 16294C>T, 16519T>C, 73A>G, 152T>C, 195T>C	+4216NlaIII; +13704BstOI, +7025Alul
I763	JT	T2	16126T>C, 16294C>T, 16296C>T, 16304T>C, 16519T>C, 73A>G, 195T>C	+4216NlaIII; +13704BstOI, +7025Alul

Table 1: RFLP and Control-Region mtDNA Haplotypes from the French OPA1 mutation carrier (Continued)

1348	U	U3	16343A>G, 16390G>A, 16519T>C, 73A>G, 150C>T	-12308Hinfl,+7025Alul
1106	U	U4	16111C>T, 16140T>C, 16356T>C, 16362T>C, 16519T>C, 73A>G, 146T>C, 152T>C, 195T>C	-12308Hinfl,+7025Alul
1291	U	U5	16192C>T, 16256C>T, 16270C>T, 16291C>T, 16399A>G, 16519T>C, 73A>G	-12308Hinfl,+7025Alul
1742	U	U5 or U4	16270C>T, 16356T>C, 16519T>C, 73A>G, 152T>C, 195T>C	-12308Hinfl,+7025Alul
1363	U	U6	16172T>C, 16219A>G, 16261C>T, 16311T>C, 16361G>A, 73A>G	-12308Hinfl,+7025Alul
1278	U	K	16093T>C, 16224T>C, 16311T>C, 16519T>C, 73A>G	-12308Hinfl,+7025Alul
1455	U	K	16093T>C, 16224T>C, 16311T>C, 16519T>C, 73A>G	-12308Hinfl,+7025Alul
1587	U	K	16064T>K, 16129G>A, 16224T>C, 16311T>C, 16519T>C, 73A>G, 180T>Y	-12308Hinfl,+7025Alul
1799	U	K	16224T>C, 16311T>C, 16519T>C, 73A>G, 146T>C, 152T>C	-12308Hinfl,+7025Alul
1887	U	K	16213G>A, 16224T>C, 16311T>C, 16519T>C, 73A>G, 146T>C	-12308Hinfl,+7025Alul
1426	other	D (M)	16260C>T, 16261C>T, 16266C>T, 16301C>T, 16311T>C, 16319G>A, 16362T>C, 16527C>T, 64C>T, 73A>G, 195T>C	+7025Alul
1358	other	W	16083C>T, 16223C>T, 16292C>T, 16519T>C, 73A>G, 106G>A, 189A>G, 195T>C, 204T>C	+7025Alul
1588	other	W	16192C>T, 16223C>T, 16292C>T, 16325T>C, 16519T>C, 73A>G, 189A>G, 194C>T, 195T>C, 204T>C	+7025Alul
1759	other	X	16189T>C, 16194delA, 16223C>T, 16278C>T, 16519T>C, 73A>G, 153A>G, 195T>C	+7025Alul
1512	other	X	16189T>C, 16192C>T, 16223C>T, 16278C>T, 16519T>C, 73A>G, 153A>G, 195T>C	+7025Alul

41 carriers of a mutation on OPA1 (table 1). The range of haplogroups was extremely wide and included the typically European haplogroups H, preV, V, J, T1, T2, K, U2, U5, U4, X, and W, plus two representatives each of African haplogroup U6, and Asian haplogroup D (figure 1). The presence of the two last haplogroups is expected, since African mtDNAs are common in France and members of Asian super-haplogroup M are sporadically identified in Europe.

In light of these results (table 1), we investigated whether there was a difference in distribution among haplogroups of the patient population carrying mutation on OPA1 compared to the reference population. We therefore compared the distribution of 41 mutation carriers among European haplogroups with the frequencies obtained for 1385 unrelated French people [8]. The patients were distributed on a phylogenetic tree of the European popula-

tion based on recent total sequencing studies of mitochondrial DNA [10-13]. This tree shows the main haplogroups found in the French population. We used the new haplogroup nomenclature recently proposed [17] to place our patients on this tree, indicating the number of individuals per haplogroup (figure 1).

For the purposes of statistical comparison, we collected individuals in phylogenetic clusters centered on super-haplogroup R. This point was chosen as it corresponded to the node where the R line, which represents over 90% of the French population, divides off from the rest of the human phylogeny. We then analyzed whether each cluster constituted a risk factor compared to the rest of the population, using the odds ratio method, with Khi2 to test significance (table 2). We tested the 3 largest clusters representative of a phylogenetic reality, so that we could use Khi2 without correction. The statistic analyses of the larg-

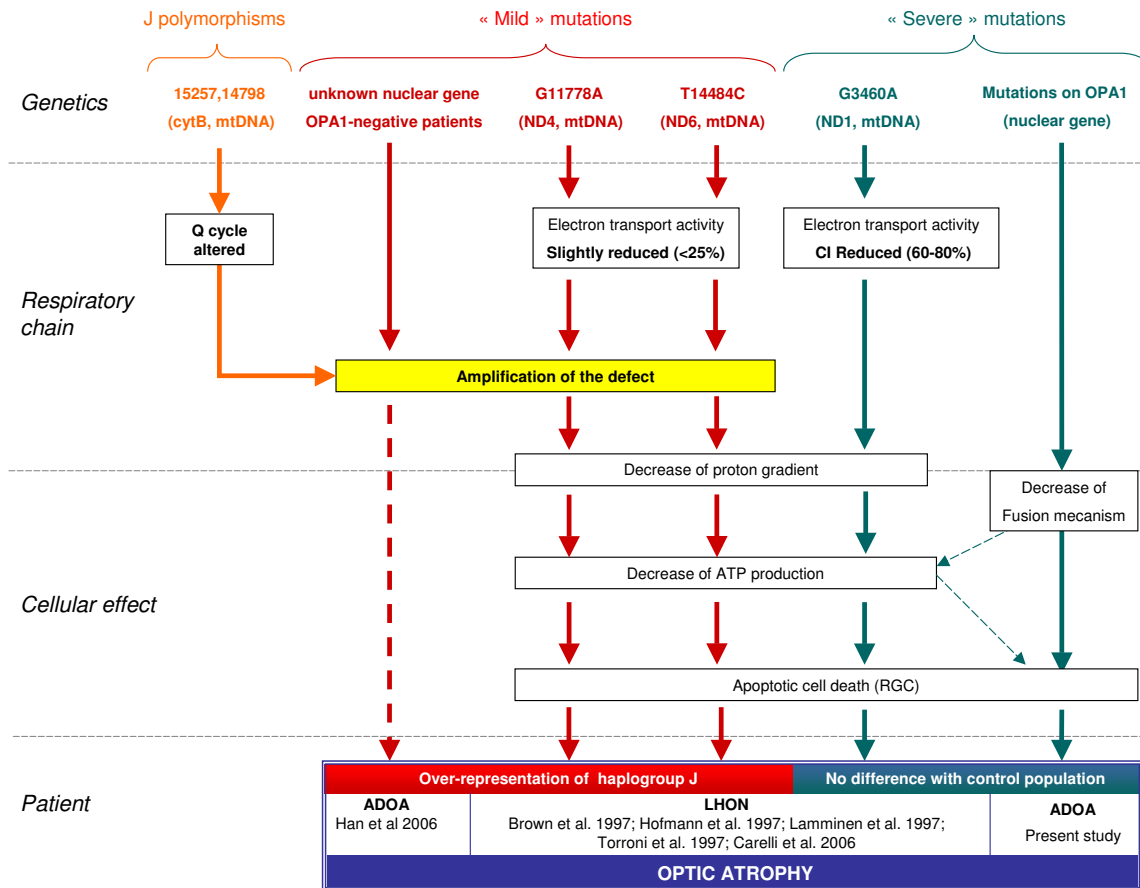


Figure 1
Phylogenical repartition of the 41 OPA1 mutation carriers. The phylogenetic tree was based on recent total sequencing studies of mitochondrial DNA [11,12]; the new haplogroup nomenclature is used [17]. The different polymorphisms used for haplogroup determination in this study are indicated in blue. The Red Cross indicate the number of individuals per haplogroup.

Table 2: Relative risk estimated of each haplogroup to develop the OPA1 pathology against the rest of the population

	OPA1 carriers	French Population	p-value 1 Fisher	O.R	p-value 2 Khi2	lower 95% CI	Upper 95% CI
R0	20	726	0,75	0,86	0,65	0,46	1,61
>H	17	660	0.52				
>HV0	3	66	0.44				
U	10	303	0.70	1,15	0,70	0,56	2,38
>U5	2	115	0.57				
>K	5	107	0.24				
JT	6	224	0.89	0,89	0,79	0,37	2,14
>J	3	106	0.95				
>T	3	118	0.86				
N	5	132	0,58				

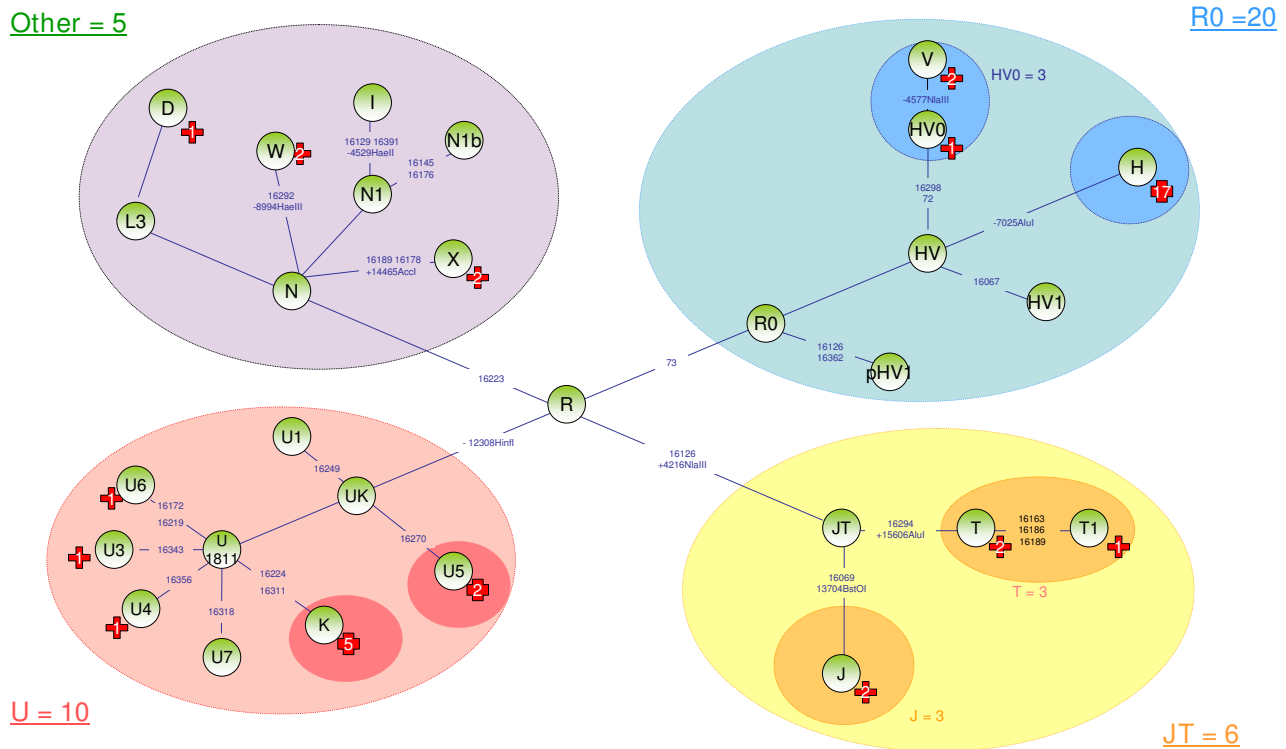


Figure 2
Influence of mtDNA background on the different optic atrophy forms. The influence of the mtDNA background (i.e., Haplogroup J) may vary according to deleterious effect of a mutation on mitochondrial oxidative phosphorylation. For example, "mild" mutations (i.e., T14484C and G11778A) could have to be amplified by the additional negative effect of mtDNA background (i.e., G15257A, T14798C) to influence the disease expression; where as "severe mutation" (i.e., G3460A, OPA1) does not necessarily need. The dashed spears describe unclear or unproved relations.

est clusters fail to show statistical differences ($p > 0.58$). To obtain more detailed results, we applied the fisher's exact test to the largest sub-cluster of R0 i.e. haplogroups H and haplogroup HV0. The frequency of haplogroup H and HV0 did not seem significantly different from that observed in the control population ($p \geq 0.44$). In view of the lack of data on the proportions of H sub-haplogroups in the French population, as well as the considerable variation in the frequency of sub-haplogroups in Western Europe [11], we have decided not to subdivide this haplogroup. As in the case of haplogroup R0, we tested for significant differences in the sub-cluster of UK and JT haplogroups, however, there was no significant difference for the sub-cluster ($p \geq 0.24$).

Discussion

The mitochondrial genetic background is known to influence the expression of the optic atrophy: i.e the LHON mitochondrial primary mutations G11778A and T14484C showed significant clustering on Caucasian mtDNA haplogroup J. It has been suggested that this clus-

tering result from an accumulation of non-synonymous J polymorphisms on the cytochrome b gene G15257A, T14798C [18]. Indeed these polymorphisms could affect the complex III efficiency and more specifically the Coenzyme Q binding sites, thus upsetting the proton pump Q cycle and, finally, oxidative phosphorylation (OXPHOS) coupling [19,20]. Consequently J polymorphisms could amplify the effect of the G11778A and T14484C mutations on the biochemical and phenotypical levels (figure 2). A recent study, based on a large cohort of relative patients confirm that the risk of visual failure is greater when the mutations are present in specific subgroups of haplogroup J [21]. However, we could not exclude that the association of the T14484C mutation with haplogroup J might be the result of an elevated specific mutation rate as proposed by [22].

In order to test the influence of the mitochondrial genetic background on the expression of the mutation of the OPA1 gene, we compared the distribution of 41 French patients among European haplogroups with a French

population sample constituted of 1385 individuals and described by a recent study [8]. The distribution of the patient population revealed no significant difference compared to the French population.

Interestingly, the clinical expression of the G3460A mtDNA mutation which is also responsible for LHON [23] is not influenced by mtDNA haplogroup J [24-27]. This mutations has a stronger deleterious effect on mitochondrial oxidative phosphorylation (OXPHOS) with a decrease of 60–80% electron transport activity [28]. In contrast, the LHON primary mutations influenced by haplogroup J (T14484C and G11778A) only display a decrease of 0–25% electron transport activity[28]. It is tempting to speculate that the deleterious effect of these later "mild" mutations could have to be amplified by the additional negative effect of mtDNA background to influence the disease expression.

The absence of strong influence of mtDNA background on OPA1-related ADOA expression suggests that, similarly to G3460A mtDNA mutation, the deleterious effect of OPA1 mutations could be responsible itself of the pathology and do not need additional mitochondrial factor. Interestingly, although the main function of OPA1 is devoted to mitochondrial inner membrane remodeling, recent data suggest that mutations of OPA1 could have strong deleterious effect on oxidative phosphorylation efficiency [29,30]. It is tempting to speculate that this energetic defect could be sufficient to induce clinical expression independently to the mtDNA background. Interestingly it has been proposed by Han *et al.* that the expression of OPA1-negative ADOA could remain dependant of additional negative effect of mtDNA background [7]. However, his study was based on an insufficient number of patients and the patients may have a heterogeneous genetic basis, given that multiple loci associated with DOA still await for other genes to be discovered.

Conclusion

These data allow to conclude that OPA1 should be considered as a "severe mutation", directly responsible of the optic atrophy, whereas OPA1-negative ADOA mutations need an external parameter to express the pathology (i.e. synergistic interaction with mitochondrial background). However, if the influence of mitochondrial background should be excluded for OPA1 mutation, other external parameters as well as environmental or nuclear factors could be implied in the phenotypic expression of the pathology. Although a large study remains to be performed to investigate the possible influences of mtDNA haplogroups on clinical phenotypes of OPA1-related ADOA (ie age of onset, progressivity of the disease, additional neurological symptoms such as neurosensorial deafness), our result indicates the absence of major influ-

ence of mtDNA haplogroups in the basic penetrance of the disease. In an attempt to try to better explain the interactions between mitochondrial energetic defect and hereditary optic neuropathies, we propose a model, presented in figure 2, that summarize the influence of mtDNA in optic atrophies.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DP carried out the molecular genetic studies, statistical analysis and drafted the manuscript. CR, MF, AC participated in the molecular genetic studies. PB, PR participated in the design of the study and drafted the manuscript performed the statistical analysis. PM and DT participated in the design of the study and helped to draft the manuscript. TL conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript

Acknowledgements

The authors specially thank the Association Française contre les Myopathies, INSERM, Université Victor Segalen Bordeaux-2, Aquitaine for financial support, and Françoise Tissot for her personal support. D. Pierron was supported by a grant from the Association contre les Maladies Mitochondriales and a grant from the CNRS. The Pôle Génotypage – Séquençage was constituted thanks to grants from the Conseil Régional d'Aquitaine (n°20030304002FA and n°20040305003FA) and from the FEDER (n°2003227). The authors gratefully thank Ivan Chang for the English corrections.

P. Reynier and P. Amati-Bonneau thank also the association "Retina France" et "Ouvrir les Yeux

References

1. Kjer B, Eiberg H, Kjer P, Rosenberg T: **Dominant optic atrophy mapped to chromosome 3q region. II. Clinical and epidemiological aspects.** *Acta Ophthalmol Scand* 1996, **74**(1):3-7.
2. Olichon A, Guillou E, Delettre C, Landes T, Arnaune-Pelloquin L, Emorine LJ, Mills V, Daloyau M, Hamel C, Amati-Bonneau P, *et al.*: **Mitochondrial dynamics and disease, OPA1.** *Biochim Biophys Acta* 2006, **1763**(5-6):500-509.
3. Alexander C, Votruba M, Pesch UE, Thiselton DL, Mayer S, Moore A, Rodriguez M, Kellner U, Leo-Kottler B, Auburger G, *et al.*: **OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28.** *Nat Genet* 2000, **26**(2):211-215.
4. Delettre C, Lenaers G, Griffoin JM, Gigarel N, Lorenzo C, Belenguer P, Pelloquin L, Grosgeorge J, Turc-Carel C, Perret E, *et al.*: **Nuclear gene OPA1, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy.** *Nat Genet* 2000, **26**(2):207-210.
5. Ferre M, Amati-Bonneau P, Tourmen Y, Malthiery Y, Reynier P: **eOPA1: an online database for OPA1 mutations.** *Hum Mutat* 2005, **25**(5):423-428.
6. Carelli V, La Morgia C, Iommarini L, Carroccia R, Mattiazzi M, Sangiorgi S, Farne S, Maresca A, Foscarini B, Lanzi L, *et al.*: **Mitochondrial optic neuropathies: how two genomes may kill the same cell type?** *Biosci Rep* 2007, **27**(1-3):173-184.
7. Han J, Thompson-Lowrey AJ, Reiss A, Mayorov V, Jia H, Biousse V, Newman NJ, Brown MD: **OPA1 mutations and mitochondrial DNA haplotypes in autosomal dominant optic atrophy.** *Genet Med* 2006, **8**(4):217-225.

8. Richard C, Pennarun E, Kivisild T, Tambets K, Tolk HV, Metspalu E, Reidla M, Chevalier S, Giraudeau S, Lauc LB, et al.: **An mtDNA perspective of French genetic variation.** *Ann Hum Biol* 2007, **34(1)**:68-79.
9. Pierron D, Rocher C, Amati-Bonneau P, Reynier P, Martin-Negrier ML, Allouche S, Batandier C, de Camaret BM, Godinot C, Rotig A, et al.: **New evidence of a mitochondrial genetic background paradox: impact of the J haplogroup on the A3243G mutation.** *BMC Med Genet* 2008, **9**:41.
10. Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti AS, et al.: **Saami and Berbers – an unexpected mitochondrial DNA link.** *Am J Hum Genet* 2005, **76(5)**:883-886.
11. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, et al.: **The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool.** *Am J Hum Genet* 2004, **75(5)**:910-918.
12. Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, et al.: **The role of selection in the evolution of human mitochondrial genomes.** *Genetics* 2006, **172(1)**:373-387.
13. Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP: **Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia.** *Am J Hum Genet* 2004, **75(6)**:966-978.
14. Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A: **The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs.** *Am J Hum Genet* 1999, **64(1)**:232-249.
15. Mogentale-Profizi N, Chollet L, Stevanovitch A, Dubut V, Poggi C, Pradie MP, Spadoni JL, Gilles A, Beraud-Colomb E: **Mitochondrial DNA sequence diversity in two groups of Italian Veneto speakers from Veneto.** *Ann Hum Genet* 2001, **65(Pt 2)**:153-166.
16. Miettinen O, Nurminen M: **Comparative analysis of two rates.** *Stat Med* 1985, **4(2)**:213-226.
17. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ: **Harvesting the fruit of the human mtDNA tree.** *Trends Genet* 2006, **22(6)**:339-345.
18. Carelli V, Achilli A, Valentino ML, Rengo C, Semino O, Pala M, Olivieri A, Mattiazzi M, Pallotti F, Carrara F, et al.: **Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber hereditary optic neuropathy pedigrees.** *Am J Hum Genet* 2006, **78(4)**:564-574.
19. Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC: **Effects of purifying and adaptive selection on regional variation in human mtDNA.** *Science* 2004, **303(5655)**:223-226.
20. Wallace DC: **The mitochondrial genome in human adaptive radiation and disease: on the road to therapeutics and performance enhancement.** *Gene* 2005, **354**:169-180.
21. Hudson G, Carelli V, Spruijt L, Gerards M, Mowbray C, Achilli A, Pyle A, Elson J, Howell N, La Morgia C, et al.: **Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background.** *Am J Hum Genet* 2007, **81(2)**:228-233.
22. Elliott HR, Samuels DC, Eden JA, Relton CL, Chinnery PF: **Pathogenic mitochondrial DNA mutations are common in the general population.** *Am J Hum Genet* 2008, **83(2)**:254-260.
23. Huoponen K, Vilkkij J, Aula P, Nikoskelainen EK, Savontaus ML: **A new mtDNA mutation associated with Leber hereditary optic neuropathy.** *Am J Hum Genet* 1991, **48(6)**:1147-1153.
24. Torroni A, Petrozzi M, D'Urbano L, Sellitto D, Zeviani M, Carrara F, Carducci C, Leuzzi V, Carelli V, Barboni P, et al.: **Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations I1778 and I4484.** *Am J Hum Genet* 1997, **60(5)**:1107-1121.
25. Lamminen T, Huoponen K, Sistonen P, Juvonen V, Lahermo P, Aula P, Nikoskelainen E, Savontaus ML: **mtDNA haplotype analysis in Finnish families with leber hereditary optic neuropathy.** *Eur J Hum Genet* 1997, **5(5)**:271-279.
26. Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A, Gerbitz KD: **Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease.** *Hum Mol Genet* 1997, **6(11)**:1835-1846.
27. Brown MD, Sun F, Wallace DC: **Clustering of Caucasian Leber hereditary optic neuropathy patients containing the I1778 or I4484 mutations on an mtDNA lineage.** *Am J Hum Genet* 1997, **60(2)**:381-387.
28. Yen MY, Wang AG, Wei YH: **Leber's hereditary optic neuropathy: a multifactorial disease.** *Prog Retin Eye Res* 2006, **25(4)**:381-396.
29. Lodi R, Tonon C, Valentino ML, Iotti S, Clementi V, Malucelli E, Barboni P, Longanesi L, Schimpf S, Wissinger B, et al.: **Deficit of in vivo mitochondrial ATP production in OPAI-related dominant optic atrophy.** *Ann Neurol* 2004, **56(5)**:719-723.
30. Amati-Bonneau P, Guichet A, Olichon A, Chevrollier A, Viala F, Miot S, Ayuso C, Odent S, Arrouet C, Verny C, et al.: **OPAI R445H mutation in optic atrophy associated with sensorineural deafness.** *Ann Neurol* 2005, **58(6)**:958-963.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2350/10/70/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



3.2.4 Synthèse

Les études de patients porteurs de mutations pathogènes d'*OPA1* ont révélé la grande diversité des présentations cliniques de l'AOAD, des déficits énergétiques associés à la maladie et une nouvelle relation établie entre le dysfonctionnement d'Opa1 et l'instabilité de l'ADNmt. Il reste encore beaucoup à faire pour élucider les liens complexes qui existent entre le métabolisme énergétique, la production de ROS, le métabolisme de l'ADNmt, la structure et la plasticité mitochondriale, la mort cellulaire, le transport axonal et la neurodégénérescence. Afin de contribuer à la compréhension du rôle majeur que semble jouer Opa1 dans l'ensemble de ces fonctions cellulaires et métaboliques, nous avons développé une base de données spécifique, nommée *eOPA1* (<http://lbbma.univ-angers.fr/eOPA1/>), dans le but de collecter et organiser les variations de séquence du gène *OPA1*, publiées ou non. Cette LSDB fournit un répertoire public des mutations pathogènes et des variants non pathogènes de ce gène. *eOPA1* est un outil qui s'est révélé utile pour le diagnostic moléculaire, les études statistiques et la détermination de corrélations génotype-phénotype dans les AOAD.

Notamment, nous avons analysé les résultats du dépistage moléculaire de 980 patients nous ayant été adressés pour suspicion de neuropathie optique héréditaire. Des causes moléculaires ont été identifiées chez 440 patients (45 % des patients examinés). Parmi eux, 295 patients (67 %) étaient porteurs d'une mutation d'*OPA1*, 131 patients (30 %) d'une mutation de l'ADNmt, et 14 patients (3 %), appartenant à trois familles non liées, avaient une mutation d'*OPA3*. Fait intéressant, une mutation d'*OPA1* a été trouvée dans 157 (40 %) des 392 cas apparemment sporadiques d'atrophie optique. Désormais, la LSDB *eOPA1* contient un total de 204 mutations, en incluant les 77 nouvelles mutations d'*OPA1* identifiées suite à cette étude. L'analyse statistique de ce vaste ensemble de mutations nous a conduits à proposer une stratégie diagnostique pour optimiser l'étude moléculaire des neuropathies optiques.

Parallèlement, l'influence de l'haplogroupe mitochondrial J sur l'expression de la NOHL a été démontrée et elle commençait à être suggérée sur l'AOAD. Dans ce contexte, nous avons étudié l'influence du patrimoine génétique mitochondrial sur les mutations d'*OPA1* en déterminant les haplogroupes de 41 patients français atteints d'AOAD liée à *OPA1* par séquençage de la région contrôle de leur ADNmt et technique RFLP. La comparaison entre les patients et les populations de référence n'a pas révélé de différence significative, invalidant l'hypothèse d'une forte influence du patrimoine génétique mitochondrial sur l'expression de l'AOAD. Ces données permettent de conclure que les mutations d'*OPA1* pourraient être considérées comme « sévères », directement responsables de l'atrophie optique, alors que les mutations impliquées dans la NOHL ou l'AOAD non liée à *OPA1*

auraient besoin de facteurs externes pour exprimer la pathologie, comme la synergie avec le patrimoine génétique mitochondrial.

3.3 AUTRES PUBLICATIONS

Au cours de ce travail de thèse, j'ai également participé à différentes études dont les résultats ont été publiés dans plusieurs articles, listés par ordre chronologique avec leur résumé original :

1. ***Sporadic optic atrophy due to synonymous codon change altering mRNA splicing of OPA1.***

Amati-Bonneau P., Pasquier L., Lainey E., Ferre M., Odent S., Malthiery Y., Bonneau D., and Reynier P., Clin Genet, 2005. **67**(1): p. 102-3.

TO THE EDITOR: Autosomal dominant optic atrophy (ADOA, OMIM 165500) is the most frequent form of hereditary optic atrophy first described by Kjer (1). This disease is, generally, characterized by the progressive decrease of visual acuity first appearing in childhood, loss of sensitivity in the central visual field (central, paracentral or coecocentral scotomas), optic nerve pallor and dyschromatopsia often predominating in the blue-yellow hues (tritanopia). Mutations in the OPA1 gene, located on chromosome 3q28–q29 and which codes for a mitochondrial dynamin-related GTPase, are found in about 60 % of patients affected with ADOA (2–4). We report, in this study, a sporadic case of progressive optic atrophy caused by a de novo heterozygous c.1770G > C variant in exon 18 of the OPA1 gene. This mutation was absent in 400 chromosome controls, but it did not modify the corresponding amino acid (R590R) of the protein. The pathogenicity of this new mutation was demonstrated at the mRNA level. Cloning and sequencing of the transcriptional product showed a frameshift caused by the entire exon 18 skipping.

PMID: 15617556

2. ***Mitochondrial coupling defect in Charcot-Marie-Tooth type 2A disease.***

Loiseau D., Chevrollier A., Verny C., Guillet V., Gueguen N., Pou de Crescenzo M.A., Ferre M., Malinge M.C., Guichet A., Nicolas G., Amati-Bonneau P., Malthiery Y., Bonneau D., and Reynier P., Ann Neurol, 2007. **61**(4): p. 315-23.

OBJECTIVE: Mutations of the mitofusin 2 gene (MFN2) may account for at least a third of the cases of Charcot-Marie-Tooth disease type 2 (CMT2). This study investigates mitochondrial cellular bioenergetics in MFN2-related CMT2A. METHODS: Mitochondrial network morphology and metabolism were studied in cultures of skin fibroblasts obtained from four CMT2A patients harboring novel missense mutations of the MFN2 gene. RESULTS: Although the mitochondrial network appeared morphologically unaltered, there was a significant defect of

mitochondrial coupling associated with a reduction of the mitochondrial membrane potential. INTERPRETATION: Our results suggest that the sharply reduced efficacy of oxidative phosphorylation in MFN2-related CMT2A may contribute to the pathophysiology of the axonal neuropathy.

PMID: 17444508

3. **Mitochondrial complex I deficiency in GDAP1-related autosomal dominant Charcot-Marie-Tooth disease (CMT2K).**

Cassereau J., Chevrollier A., Gueguen N., Malinge M.C., Letournel F., Nicolas G., Richard L., Ferre M., Verny C., Dubas F., Procaccio V., Amati-Bonneau P., Bonneau D., and Reynier P., Neurogenetics, 2008.

Mutations in GDAP1, an outer mitochondrial membrane protein responsible for recessive Charcot-Marie-Tooth disease (CMT4A), have also been associated with CMT2K, a dominant form of the disease. The three CMT2K patients we studied carried a novel dominant GDAP1 mutation, C240Y (c.719G > A). Mitochondrial respiratory chain complex I activity in fibroblasts from CMT2K patients was 40 % lower than in controls, whereas the tubular mitochondria were 33 % larger in diameter and the mitochondrial mass was 20 % greater. Thus, besides the regulatory role GDAP1 plays in mitochondrial network dynamics, it may also be involved in energy production and in the control of mitochondrial volume.

PMID: 19089472

4. **Hereditary optic neuropathies share a common mitochondrial coupling defect.**

Chevrollier A., Guillet V., Loiseau D., Gueguen N., de Crescenzo M.A., Verny C., Ferre M., Dollfus H., Odent S., Milea D., Goizet C., Amati-Bonneau P., Procaccio V., Bonneau D., and Reynier P., Ann Neurol, 2008. **63**(6): p. 794-8.

Hereditary optic neuropathies are heterogeneous diseases characterized by the degeneration of retinal ganglion cells leading to optic nerve atrophy and impairment of central vision. We found a common coupling defect of oxidative phosphorylation in fibroblasts of patients affected by autosomal dominant optic atrophy (mutations of OPA1), autosomal dominant optic atrophy associated with cataract (mutations of OPA3), and Leber's hereditary optic neuropathy, a disorder associated with point mutations of mitochondrial DNA complex I genes. Interestingly, the energetic defect was significantly more pronounced in Leber's hereditary optic neuropathy and autosomal dominant optic atrophy patients with a more complex phenotype, the so-called plus phenotype.

PMID: 18496845

5. **Reversible optic neuropathy with OPA1 exon 5b mutation.**

Cornille K., Milea D., Amati-Bonneau P., Procaccio V., Zazoun L., Guillet V., El Achouri G., Delettre C., Gueguen N., Loiseau D., Muller A., Ferre M., Chevrollier A., Wallace D.C., Bonneau D., Hamel C., Reynier P., and Lenaers G., *Ann Neurol*, 2008.

A new c.740G>A (R247H) mutation in OPA1 alternate spliced exon 5b was found in a patient presenting with bilateral optic neuropathy followed by partial, spontaneous visual recovery. R247H fibroblasts from the patient and his unaffected father presented unusual highly tubular mitochondrial network, significant increased susceptibility to apoptosis, oxidative phosphorylation uncoupling, and altered OPA1 protein profile, supporting the pathogenicity of this mutation. These results suggest that the clinical spectrum of the OPA1-associated optic neuropathies may be larger than previously described, and that spontaneous recovery may occur in cases harboring an exon 5b mutation. Ann Neurol 2008.

PMID: 18360822

6. **Acute and late-onset optic atrophy due to a novel OPA1 mutation leading to a mitochondrial coupling defect.**

Nochez Y., Arsene S., Gueguen N., Chevrollier A., Ferre M., Guillet V., Desquiret V., Toutain A., Bonneau D., Procaccio V., Amati-Bonneau P., Pisella P.J., and Reynier P., *Mol Vis*, 2009. **15**: p. 598-608.

PURPOSE: Autosomal dominant optic atrophy (ADOA, OMIM 165500), an inherited optic neuropathy that leads to retinal ganglion cell degeneration and reduced visual acuity during the early decades of life, is mainly associated with mutations in the OPA1 gene. Here we report a novel ADOA phenotype associated with a new pathogenic OPA1 gene mutation. METHODS: The patient, a 62-year-old woman, was referred for acute, painless, and severe visual loss in her right eye. Acute visual loss in her left eye occurred a year after initial presentation. MRI confirmed the diagnosis of isolated atrophic bilateral optic neuropathy. We performed DNA sequencing of the entire coding sequence and the exon/intron junctions of the OPA1 gene, and we searched for the mitochondrial DNA mutations responsible for Leber hereditary optic atrophy by sequencing entirely mitochondrial DNA. Mitochondrial respiratory chain complex activity and mitochondrial morphology were investigated in skin fibroblasts from the patient and controls. RESULTS: We identified a novel heterozygous missense mutation (c.2794C>T) in exon 27 of the OPA1 gene, resulting in an amino acid change (p.R932C) in the protein. This mutation, which affects a highly conserved amino acids, has not been previously reported, and was

absent in 400 control chromosomes. Mitochondrial DNA sequence analysis did not reveal any mutation associated with Leber hereditary optic neuropathy or any pathogenic mutations. The investigation of skin fibroblasts from the patient revealed a coupling defect of oxidative phosphorylation and a larger proportion of short mitochondria than in controls. CONCLUSIONS: The presence of an OPA1 mutation indicates that this sporadic, late-onset acute case of optic neuropathy is related to ADOA and to a mitochondrial energetic defect. This suggests that the mutational screening of the OPA1 gene would be justified in atypical cases of optic nerve atrophy with no evident cause.

PMID: 19325939

7. OPA1-associated disorders: Phenotypes and pathophysiology.

Amati-Bonneau P., Milea D., Bonneau D., Chevrollier A., Ferre M., Guillet V., Gueguen N., Loiseau D., Crescenzo M.A., Verny C., Procaccio V., Lenaers G., and Reynier P., *Int J Biochem Cell Biol*, 2009.

The OPA1 gene, encoding a dynamin-like mitochondrial GTPase, is involved in autosomal dominant optic atrophy (ADOA, OMIM #165500). ADOA, also known as Kjer's optic atrophy, affects retinal ganglion cells and the axons forming the optic nerve, leading to progressive visual loss. OPA1 gene sequencing in patients with hereditary optic neuropathies indicates that the clinical spectrum of ADOA is larger than previously thought. Specific OPA1 mutations are responsible for several distinct clinical presentations, such as ADOA with deafness (ADOAD), and severe multi-systemic syndromes, the so-called "ADOA plus" disorders, which involve neurological and neuromuscular symptoms similar to those due to mitochondrial oxidative phosphorylation defects or mitochondrial DNA instability. The study of the various clinical presentations of ADOA in conjunction with the investigation of OPA1 mutations in fibroblasts from patients with optic atrophy provides new insights into the pathophysiological mechanisms of the disease while underscoring the multiple physiological roles played by OPA1 in energetic metabolism, mitochondrial structure and maintenance, and cell death. Finally, OPA1 represents an important new paradigm for emerging neurodegenerative diseases affecting mitochondrial structure, plasticity and functions.

PMID: 19389487

8. Adenine nucleotide translocase is involved in a mitochondrial coupling defect in MFN2-related Charcot-Marie-Tooth type 2A disease.

Guillet V., Gueguen N., Verny C., Ferre M., Homedan C., Loiseau D., Procaccio V.,

Amati-Bonneau P., Bonneau D., Reynier P., and Chevrollier A., *Neurogenetics*, 2009.

Charcot-Marie-Tooth type 2A disease (CMT2A), a dominantly inherited peripheral neuropathy, is caused by mutations in MFN2, a mitochondrial fusion protein. Having previously demonstrated a mitochondrial coupling defect in CMT2A patients' fibroblasts, we here investigate mitochondrial oxygen consumption and the expression of adenine nucleotide translocase (ANT) and uncoupling proteins from eight other patients with the disease. The mitochondrial uncoupling was associated with a higher respiratory rate, essentially involving complex II proteins. Furthermore, a twofold increase in the expression of ANT led to the reduced efficiency of oxidative phosphorylation in CMT2A cells, suggesting that MFN2 plays a role in controlling ATP/ADP exchanges.

PMID: 19618221

CHAPITRE 4 : DISCUSSION GÉNÉRALE, CONCLUSION ET PERSPECTIVES

L'objectif de la détermination de l'ensemble des protéines mitochondriales se rapproche. Sur la base de l'étude la plus exhaustive à ce jour du protéome mitochondrial de *Saccharomyces cerevisiae*, environ 700 protéines sont identifiées comme étant localisées dans la mitochondrie [67]. Il s'agit d'une estimation prudente qui a produit un nombre relativement faible de protéines mitochondriales par rapport aux prévisions antérieures de 800 [81], et par rapport aux estimations correspondantes pour *Arabidopsis thaliana* (2 800) ou pour l'homme (1 500–4 200 [234, 235]). Les mitochondries des mammifères sont impliquées dans plusieurs processus qui ne se produisent pas dans les mitochondries de levure (par exemple la thermogenèse, la biosynthèse des stéroïdes et l'apoptose). Les protéines tissu-spécifiques et celles du développement, ainsi que leurs isoformes, sont susceptibles d'accroître encore le nombre de protéines mitochondriales dans les organismes supérieurs. Par conséquent, bien que cela ne soit pas formellement prouvé, il semble probable que le nombre des protéines présentes dans les mitochondries de mammifères et de plantes soit bien plus important que celui des levures.

Toutefois, de nombreuses protéines faiblement abondantes ou non extractibles, en particulier dans les cellules de mammifères, pourraient avoir résisté à l'identification à ce jour. Par conséquent, à l'heure actuelle, une estimation fiable du nombre total des protéines mitochondriales dans un organisme n'est pas possible. D'autres améliorations technologiques, *in vivo*, *in vitro* ou *in silico*, sont nécessaires qui pourront contribuer aux analyses intégratives. Des outils doivent être développés pour permettre l'identification de ces protéines « difficiles à obtenir ». Dans ce contexte, notre étude apporte une approche nouvelle pour l'identification de protéines mitochondriales dérivant de l'endosymbionte protomitochondrial procaryote. Elle mérite d'être approfondie, car nous avons également montré que ces protéines représentent des candidats préférentiels impliqués dans les mitochondriopathies. La découverte de nouvelles protéines mitochondriales par criblage bio-informatique à haut débit pourrait ainsi contribuer à l'identification de nouvelles clés dans la compréhension des métabolismes mitochondriaux et fournir des cibles thérapeutiques potentielles.

Un défi futur sera certainement d'assigner toutes les protéines à leur sous-localisation mitochondriale telle que la membrane externe, la membrane interne, la matrice et l'espace intermembranaire. Enfin, à ce jour, l'origine génétique de nombreuses pathologies mitochondriales est inconnue et de nombreux gènes restent à analyser. Ainsi, l'analyse détaillée de chaque protéine est indispensable pour permettre de mieux comprendre la physiopathologie de ces affections. La création de systèmes de référence pour décrire les informations phénotypiques, permettant de fédérer les résultats des travaux de toutes les équipes expertes d'un domaine spécifique, aura un rôle crucial. Puis, en facilitant la communication entre les LSDB, les registres de recueil des données cliniques et les autres bases de données, de puissantes requêtes pourront être effectuées pour résoudre des problèmes d'un haut niveau de complexité.

En nous focalisant sur notre domaine d'expertise, quelques années après l'implication du gène *OPA1* dans les AOAD, nous avons développé eOPA1 qui est, à ce jour, l'unique LSDB dédiée à ce gène. Après cinq années d'existence, ce service n'a pas connu d'interruption et le contenu de la base de données est régulièrement mis à jour, en intégrant principalement les nouvelles variations publiées, pathologiques ou non, mais également des soumissions directes de la part d'autres laboratoires. Bien que quelques améliorations aient été apportées au site Web pour l'adapter aux nouvelles menaces de sécurité apparues avec l'évolution d'Internet, l'aspect et le mode de consultation de eOPA1 ont été conservés afin de limiter autant que possible le temps passé à sa consultation. Cependant, cette uniformité nécessaire pour optimiser la navigation n'est valable que pour les utilisateurs de celle-ci exclusivement, ce qui ne correspond évidemment pas à la pratique courante des généticiens ou des biologistes moléculaires. Aujourd'hui, des centaines de LSDB sont librement accessibles sur Internet (<http://www.hgvs.org/dblist/glsdb.html>) et une étude effectuée en 2002 sur les LSDB a révélé une grande hétérogénéité dans leur contenu, dans leur fréquence de mise à jour et dans l'outil informatique utilisé pour leur développement [236], ce qui est certainement le facteur le plus limitant pour des études de grande ampleur.

Notamment, le *National Center for Biotechnology Information* (NCBI) prévoit d'annoter les séquences génomiques de l'homme avec toutes les variations et mutations identifiées, ce qui ne semble pas facilement envisageable dans le contexte actuel. Tout d'abord, seuls des experts peuvent réaliser la collecte et l'annotation des mutations. Aujourd'hui, beaucoup de mutations publiées sont mal nommées au niveau moléculaire, même si une nomenclature internationale existe. En outre, la collecte de ces informations demande généralement un temps très conséquent. Le succès de ce projet de grande envergure passera donc par des formats d'échange standards avec les LSDB. Alors, plutôt que faire évoluer l'architecture Web de notre base de données pour supporter les nouveaux gènes étudiés dans notre laboratoire comme cela avait été proposé dans notre article de 2005 [179], notre base de

données devrait évoluer vers un système plus efficace, capable d'échanger ses données avec les centres d'information en biologie moléculaire annotant la séquence du génome humain. Pour ce faire, elle devra subir une refonte de son architecture Web et faire évoluer son contenu.

Bien que les recommandations sur le contenu des LSDB [237, 238] aient été suivies pour la réalisation d'eOPA1, il serait dorénavant nécessaire de franchir une étape supplémentaire vers la standardisation, par exemple en utilisant les outils génériques *Universal Mutation Database* (UMD) [239, 240] ou *Leiden Open Variation Database* (LOVD) [241], à la fois pour faciliter l'utilisation et l'interopérabilité de la base de données, mais également pour accélérer sa mise en place en ne devant pas concevoir une nouvelle architecture logicielle dédiée. En outre, parce que les mutations d'*OPA1* étaient principalement famille-spécifiques, rendant difficile l'établissement de corrélations génotype-phénotype, nous avons pris le parti lors de la création de notre LSDB de ne répertorier les mutations qu'une seule fois, en faisant référence à leur première publication. Il faudrait désormais collecter de manière exhaustive les données moléculaires, biologiques et cliniques de chaque mutation. Ce dernier point semble en bonne voie, car les organisations nationales favorisent la création de registres de patients.

Le logiciel générique UMD [239, 240], déjà utilisé pour le développement des LSDB de 20 gènes impliqués dans le cancer ou les maladies génétiques (en septembre 2009, <http://www.umd.be>), semble être le plus adapté pour faire évoluer eOPA1 afin de répondre aux nouveaux besoins. En effet, cet outil qui est accessible gratuitement permet la création de LSDB pour pratiquement n'importe quel gène et comprend un vaste ensemble d'outils d'analyse. Entre autres, il permet d'intégrer des séquences non codantes, des données cliniques, des images, des anticorps monoclonaux et des variants non pathogènes. Ce logiciel fournit des outils spécifiquement conçus pour analyser les mutations à l'échelle moléculaire, ainsi qu'un ensemble de routines pour rechercher des corrélations génotype-phénotype. Il propose également des outils spécifiques pour les mutations rares tels que les grandes délétions, les duplications et les mutations introniques. En outre, UMD dispose de routines spécifiques pour aider les chercheurs à concevoir de nouvelles stratégies thérapeutiques, telles que le saut d'exon, la lecture ininterrompue de codons-stops par les aminoglycosides, ou la sélection d'anticorps monoclonaux et le balayage d'épitopes pour la thérapie génique.

Néanmoins, malgré des évolutions techniques nécessaires pour répondre aux nouvelles problématiques de recherche, la LSDB eOPA1 est devenue un outil utile pour les laboratoires réalisant le diagnostic moléculaire du gène *OPA1*, auxquels elle était initialement dédiée ; ceci se manifeste par les citations dans les publications scientifiques et par les échanges qui ont lieu entre l'administrateur et les utilisateurs. Le regroupement des

données sur la variabilité d'*OPA1* a également contribué à la réalisation d'une grande étude dont un des résultats les plus intéressants est la mise en évidence d'une mutation remarquable, significativement plus fréquente que les deux cents autres. De même, certains exons ont été observés plus fréquemment mutés, permettant ainsi d'optimiser le diagnostic moléculaire d'*OPA1*. Ces deux exemples appuient encore l'intérêt de faire évoluer e*OPA1* afin d'intégrer toutes les informations disponibles. Aussi, les variants non pathogènes d'*OPA1*, dont la publication dans la littérature peut sembler présenter peu d'intérêt, sont d'une importance pratique considérable dans le diagnostic moléculaire. La frontière entre mutation pathogène et variation non pathogène est souvent délicate et le statut d'une variation de séquence est amené à évoluer. Toute variation de séquence devrait alors être qualifiée par un indice reflétant la probabilité que celle-ci soit pathogène plutôt que par une catégorie (« mutation pathogène » vs « variant non pathogène »). Ce « score de pathogénicité » serait calculé en intégrant les méthodes déjà utilisées pour affecter le statut pathogène à une mutation telles que, entre autres, la conservation interespèces, la prévalence et la conséquence sur la structure 3D.

Enfin, l'influence du patrimoine génétique mitochondrial a été établie, y compris dans certaines neuropathies optiques héréditaires [156, 233], et ce type d'information, s'il est disponible, mériterait d'être intégré dans les LSDB. Dans le cas des AOAD liées à *OPA1*, le patrimoine génétique mitochondrial ne semble pas influencer l'expression de l'AOAD car Opa1 serait une protéine directement liée à la pathologie, touchant le mécanisme fondamental de la fusion mitochondriale ; aucune contribution supplémentaire ne serait nécessaire pour entraîner le déséquilibre à l'origine du dysfonctionnement observé. Il s'agit cependant d'une étude préliminaire qui devrait être confirmée sur une population de patients plus importante grâce à l'évolution des techniques d'haplogroupage. Ainsi, si l'influence du patrimoine mitochondrial sur les mutations d'*OPA1* est exclue, d'autres paramètres extérieurs pourraient être mis en évidence, comme des facteurs environnementaux ou des liens avec d'autres gènes nucléaires. Des facteurs liés à l'ADNmt, autres que les variations caractérisant les haplogroupes mitochondriaux, pourraient également être en cause, car des mutations sévères d'*OPA1* responsables de présentations cliniques syndromiques d'AOAD seraient associées à la présence de délétions dans le génome mitochondrial [167, 168]. Ainsi, la compréhension de l'expression phénotypique des pathologies sera certainement liée à celle de la synergie entre les génomes nucléaires et mitochondriaux.

Parallèlement à la disponibilité de LSDB et de registre de patients pour de nombreuses maladies génétiques humaines, on observe l'apparition de nouvelles approches thérapeutiques. Les thérapies géniques les plus récentes et les plus prometteuses seraient basées directement sur la lésion moléculaire en essayant de restaurer une protéine fonctionnelle en agissant sur un allèle spécifique. Une de ces stratégies consiste à

développer de petites molécules qui interagissent avec les protéines riches en sérine et arginine (*serine- and arginine-rich*, SR), jouant un rôle clé dans la régulation de la sélection des sites d'épissage et donc dans le contrôle des processus d'épissage. Les mutations perturbant ou créant des signaux d'épissage auxiliaires et conduisant à des événements d'épissage aberrants à l'origine de pathologies humaines seraient donc des cibles de cette approche thérapeutique [242, 243]. Un autre type d'approche, basée sur le saut d'exon, est utilisée pour tenter de restaurer une protéine fonctionnelle, en supprimant un ou plusieurs exons, afin de rétablir le cadre de lecture chez des patients porteurs d'une mutation non-sens ou d'une insertion ou délétion hors cadre de lecture. Le succès de ces types de stratégies fait souvent suite au regroupement des données moléculaires et cliniques au sein de LSDB [242]. Ces exemples soulignent encore l'importance de la caractérisation des mutations responsables de maladies humaines et de leur collecte pour les stratégies thérapeutiques.

Si beaucoup reste encore à découvrir sur la fonction d'Opa1 et la physiopathologie des AOAD, des essais thérapeutiques peuvent dès à présent être envisagés à partir des différents modèles de fibroblastes et de souris maintenant disponibles. En outre, la thérapie génique semble être une approche intéressante puisque l'œil s'y prêterait particulièrement bien [244] et parce que l'haplo-insuffisance serait le principal mécanisme impliqué, au moins dans les AOAD isolées. La situation de l'AOAD, à l'intersection de l'ophtalmologie, de la neurologie et des pathologies mitochondriales, permet d'envisager des stratégies de traitement fondées soit sur la thérapie génique, actuellement testée chez l'homme pour l'amaurose congénitale de Leber (MIM #204100) [245], soit sur des thérapies pharmacologiques, qui sont à ce jour les seuls moyens existant pour traiter les mitochondriopathies non rétiniennes et les maladies neurologiques.

Finalement, le séquençage du génome humain est presque complet et son annotation bien avancée. La pleine exploitation des nouvelles données, notamment pour la compréhension de l'étiologie des pathologies héréditaires et la prédisposition aux maladies, est toutefois susceptible d'être entravée par l'ignorance des processus fondamentaux de la diversité génétique interindividuelle, interpopulation et interespèce. Au niveau de la population, une telle compréhension est considérée comme essentielle pour toute interprétation de la prévalence ou de l'incidence des maladies avec une base génétique. Dans les familles, il s'agit d'une condition préalable pour être en mesure d'expliquer les variations interindividuelles et comment une même mutation d'un gène peut être associée à une expression phénotypique variable. Ainsi, la métaanalyse des mutations germinales pathologiques dans les gènes humains devrait permettre l'optimisation des stratégies de dépistage des mutations et faciliter l'identification et la compréhension des maladies à

pénétrance incomplète, en les rapprochant de mutations associées à des phénotypes cliniques établis.

RÉFÉRENCES

1. Margulis L., *Origin of eukaryotic cells; evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth*. 1970, New Haven: Yale University Press.
2. Andersson S.G., Zomorodipour A., Andersson J.O., Sicheritz-Ponten T., Alsmark U.C., Podowski R.M., Naslund A.K., Eriksson A.S., Winkler H.H., and Kurland C.G., *The genome sequence of Rickettsia prowazekii and the origin of mitochondria*. Nature, 1998. **396**(6707): p. 133-40.
3. Gray M.W., *The endosymbiont hypothesis revisited*. Int Rev Cytol, 1992. **141**: p. 233-357.
4. Gray M.W., Burger G., and Lang B.F., *Mitochondrial evolution*. Science, 1999. **283**(5407): p. 1476-81.
5. Gray M.W., Burger G., and Lang B.F., *The origin and early evolution of mitochondria*. Genome Biol, 2001. **2**(6): p. REVIEWS1018.
6. Martin W. and Muller M., *The hydrogen hypothesis for the first eukaryote*. Nature, 1998. **392**(6671): p. 37-41.
7. Sicheritz-Ponten T., Kurland C.G., and Andersson S.G., *A phylogenetic analysis of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae*. Biochim Biophys Acta, 1998. **1365**(3): p. 545-51.
8. Henze K. and Martin W., *How do mitochondrial genes get into the nucleus?* Trends Genet, 2001. **17**(7): p. 383-7.
9. Kurland C.G. and Andersson S.G., *Origin and evolution of the mitochondrial proteome*. Microbiol Mol Biol Rev, 2000. **64**(4): p. 786-820.
10. Tourmen Y., Baris O., Dessen P., Jacques C., Malthiery Y., and Reynier P., *Structure and chromosomal distribution of human mitochondrial pseudogenes*. Genomics, 2002. **80**(1): p. 71-7.
11. Woischnik M. and Moraes C.T., *Pattern of organization of human mitochondrial pseudogenes in the nuclear genome*. Genome Res, 2002. **12**(6): p. 885-93.
12. Wallace D.C., *Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine*. Annu Rev Biochem, 2007. **76**: p. 781-821.
13. 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., and Young I.G., *Sequence and organization of the human mitochondrial genome*. Nature, 1981. **290**(5806): p. 457-65.
14. Andrews R.M., Kubacka I., Chinnery P.F., Lightowlers R.N., Turnbull D.M., and Howell N., *Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA*. Nat Genet, 1999. **23**(2): p. 147.
15. Wolstenholme D.R., *Animal mitochondrial DNA: structure and evolution*. Int Rev Cytol, 1992. **141**: p. 173-216.

16. Boore J.L., *Animal mitochondrial genomes*. Nucleic Acids Res, 1999. **27**(8): p. 1767-80.
17. Gyllensten U., Wharton D., and Wilson A.C., *Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice*. J Hered, 1985. **76**(5): p. 321-4.
18. Awadalla P., Eyre-Walker A., and Smith J.M., *Linkage disequilibrium and recombination in hominid mitochondrial DNA*. Science, 1999. **286**(5449): p. 2524-5.
19. Kumar S., Hedrick P., Dowling T., and Stoneking M., *Questioning evidence for recombination in human mitochondrial DNA*. Science, 2000. **288**(5473): p. 1931.
20. Muller H.J., *The Relation of Recombination to Mutational Advance*. Mutat Res, 1964. **106**: p. 2-9.
21. Felsenstein J., *The evolutionary advantage of recombination*. Genetics, 1974. **78**(2): p. 737-56.
22. Tully L.A., Parsons T.J., Steighner R.J., Holland M.M., Marino M.A., and Prenger V.L., *A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region*. Am J Hum Genet, 2000. **67**(2): p. 432-43.
23. Barrell B.G., Bankier A.T., and Drouin J., *A different genetic code in human mitochondria*. Nature, 1979. **282**(5735): p. 189-94.
24. Mourier T., Hansen A.J., Willerslev E., and Arctander P., *The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus*. Mol Biol Evol, 2001. **18**(9): p. 1833-7.
25. Leister D., *Origin, evolution and genetic effects of nuclear insertions of organelle DNA*. Trends Genet, 2005. **21**(12): p. 655-63.
26. Brown W.M., George M., Jr., and Wilson A.C., *Rapid evolution of animal mitochondrial DNA*. Proc Natl Acad Sci U S A, 1979. **76**(4): p. 1967-71.
27. Nachman M.W. and Crowell S.L., *Estimate of the mutation rate per nucleotide in humans*. Genetics, 2000. **156**(1): p. 297-304.
28. Hasegawa M., Di Rienzo A., Kocher T.D., and Wilson A.C., *Toward a more accurate time scale for the human mitochondrial DNA tree*. J Mol Evol, 1993. **37**(4): p. 347-54.
29. Cann R.L., Stoneking M., and Wilson A.C., *Mitochondrial DNA and human evolution*. Nature, 1987. **325**(6099): p. 31-6.
30. Vigilant L., Stoneking M., Harpending H., Hawkes K., and Wilson A.C., *African populations and the evolution of human mitochondrial DNA*. Science, 1991. **253**(5027): p. 1503-7.
31. Stringer C.B. and Andrews P., *Genetic and fossil evidence for the origin of modern humans*. Science, 1988. **239**(4845): p. 1263-8.
32. Lahr M.M. and Foley R.A., *Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution*. Am J Phys Anthropol, 1998. **Suppl 27**: p. 137-76.
33. Wolpoff M.H., *Theories of modern human origins*, in *Continuity or Replacement: Controversies in Homo sapiens Evolution*, Bräuer G. and Smith F.H., Editors. 1992, Balkema: Rotterdam. p. 25-63.

34. Johnson M.J., Wallace D.C., Ferris S.D., Rattazzi M.C., and Cavalli-Sforza L.L., *Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns*. J Mol Evol, 1983. **19**(3-4): p. 255-71.
35. Seielstad M.T., Minch E., and Cavalli-Sforza L.L., *Genetic evidence for a higher female migration rate in humans*. Nat Genet, 1998. **20**(3): p. 278-80.
36. Wallace D.C., Brown M.D., and Lott M.T., *Mitochondrial DNA variation in human evolution and disease*. Gene, 1999. **238**(1): p. 211-30.
37. Richards M. and Macaulay V., *The mitochondrial gene tree comes of age*. Am J Hum Genet, 2001. **68**(6): p. 1315-20.
38. Torroni A., Bandelt H.J., Macaulay V., Richards M., Cruciani F., Rengo C., Martinez-Cabrera V., Villems R., Kivisild T., Metspalu E., Parik J., Tolk H.V., Tambets K., Forster P., Karger B., Francalacci P., Rudan P., Janicijevic B., Rickards O., Savontaus M.L., Huoponen K., Laitinen V., Koivumaki S., Sykes B., Hickey E., Novelletto A., Moral P., Sellitto D., Coppa A., Al-Zaheri N., Santachiara-Benerecetti A.S., Semino O., and Scozzari R., *A signal, from human mtDNA, of postglacial recolonization in Europe*. Am J Hum Genet, 2001. **69**(4): p. 844-52.
39. Underhill P.A., Passarino G., Lin A.A., Shen P., Mirazon Lahr M., Foley R.A., Oefner P.J., and Cavalli-Sforza L.L., *The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations*. Ann Hum Genet, 2001. **65**(Pt 1): p. 43-62.
40. Wallace D.C., *1994 William Allan Award Address. Mitochondrial DNA variation in human evolution, degenerative disease, and aging*. Am J Hum Genet, 1995. **57**(2): p. 201-23.
41. Wallace D.C. and Torroni A., *American Indian prehistory as written in the mitochondrial DNA: a review*. Hum Biol, 1992. **64**(3): p. 403-16.
42. Ruiz-Pesini E., Lott M.T., Procaccio V., Poole J.C., Brandon M.C., Mishmar D., Yi C., Kreuziger J., Baldi P., and Wallace D.C., *An enhanced MITOMAP with a global mtDNA mutational phylogeny*. Nucleic Acids Res, 2007. **35**(Database issue): p. D823-8.
43. Pereira L., Freitas F., Fernandes V., Pereira J.B., Costa M.D., Costa S., Maximo V., Macaulay V., Rocha R., and Samuels D.C., *The diversity present in 5140 human mitochondrial genomes*. Am J Hum Genet, 2009. **84**(5): p. 628-40.
44. Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., and Sayers E.W., *GenBank*. Nucleic Acids Res, 2009. **37**(Database issue): p. D26-31.
45. Wallace D.C., Singh G., Lott M.T., Hodge J.A., Schurr T.G., Lezza A.M., Elsas L.J., 2nd, and Nikoskelainen E.K., *Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy*. Science, 1988. **242**(4884): p. 1427-30.
46. Lestienne P. and Ponsot G., *Kearns-Sayre syndrome with muscle mitochondrial DNA deletion*. Lancet, 1988. **1**(8590): p. 885.
47. Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G., Hosseini S., Brandon M., Easley K., Chen E., Brown M.D., Sukernik R.I., Olckers A., and Wallace D.C., *Natural selection shaped regional mtDNA variation in humans*. Proc Natl Acad Sci U S A, 2003. **100**(1): p. 171-6.

48. Ruiz-Pesini E., Mishmar D., Brandon M., Procaccio V., and Wallace D.C., *Effects of purifying and adaptive selection on regional variation in human mtDNA*. Science, 2004. **303**(5655): p. 223-6.
49. Ruiz-Pesini E., Lapena A.C., Diez-Sanchez C., Perez-Martos A., Montoya J., Alvarez E., Diaz M., Urries A., Montoro L., Lopez-Perez M.J., and Enriquez J.A., *Human mtDNA haplogroups associated with high or reduced spermatozoa motility*. Am J Hum Genet, 2000. **67**(3): p. 682-96.
50. Tanaka M., Gong J.S., Zhang J., Yoneda M., and Yagi K., *Mitochondrial genotype associated with longevity*. Lancet, 1998. **351**(9097): p. 185-6.
51. De Benedictis G., Rose G., Carrieri G., De Luca M., Falcone E., Passarino G., Bonafe M., Monti D., Baggio G., Bertolini S., Mari D., Mattace R., and Franceschi C., *Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans*. FASEB J, 1999. **13**(12): p. 1532-6.
52. Ivanova R., Lepage V., Charron D., and Schachter F., *Mitochondrial genotype associated with French Caucasian centenarians*. Gerontology, 1998. **44**(6): p. 349.
53. Niemi A.K., Hervonen A., Hurme M., Karhunen P.J., Jylha M., and Majamaa K., *Mitochondrial DNA polymorphisms associated with longevity in a Finnish population*. Hum Genet, 2003. **112**(1): p. 29-33.
54. Ross O.A., McCormack R., Curran M.D., Duguid R.A., Barnett Y.A., Rea I.M., and Middleton D., *Mitochondrial DNA polymorphism: its role in longevity of the Irish population*. Exp Gerontol, 2001. **36**(7): p. 1161-78.
55. Chagnon P., Gee M., Filion M., Robitaille Y., Belouchi M., and Gauvreau D., *Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population*. Am J Med Genet, 1999. **85**(1): p. 20-30.
56. van der Walt J.M., Nicodemus K.K., Martin E.R., Scott W.K., Nance M.A., Watts R.L., Hubble J.P., Haines J.L., Koller W.C., Lyons K., Pahwa R., Stern M.B., Colcher A., Hiner B.C., Jankovic J., Ondo W.G., Allen F.H., Jr., Goetz C.G., Small G.W., Mastaglia F., Stajich J.M., McLaurin A.C., Middleton L.T., Scott B.L., Schmechel D.E., Pericak-Vance M.A., and Vance J.M., *Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease*. Am J Hum Genet, 2003. **72**(4): p. 804-11.
57. Chinnery P.F., Taylor G.A., Howell N., Andrews R.M., Morris C.M., Taylor R.W., McKeith I.G., Perry R.H., Edwardson J.A., and Turnbull D.M., *Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies*. Neurology, 2000. **55**(2): p. 302-4.
58. Moreno-Loshuertos R., Acin-Perez R., Fernandez-Silva P., Movilla N., Perez-Martos A., Rodriguez de Cordoba S., Gallardo M.E., and Enriquez J.A., *Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants*. Nat Genet, 2006. **38**(11): p. 1261-8.
59. Guan M.X., Enriquez J.A., Fischel-Ghodsian N., Puranam R.S., Lin C.P., Maw M.A., and Attardi G., *The deafness-associated mitochondrial DNA mutation at position 7445, which*

- affects tRNA^{Ser}(UCN) precursor processing, has long-range effects on NADH dehydrogenase subunit ND6 gene expression.* Mol Cell Biol, 1998. **18**(10): p. 5868-79.
60. Reid F.M., Vernham G.A., and Jacobs H.T., *A novel mitochondrial point mutation in a maternal pedigree with sensorineural deafness.* Hum Mutat, 1994. **3**(3): p. 243-7.
61. Shaw J.M. and Nunnari J., *Mitochondrial dynamics and division in budding yeast.* Trends Cell Biol, 2002. **12**(4): p. 178-84.
62. Jiang X. and Wang X., *Cytochrome C-mediated apoptosis.* Annu Rev Biochem, 2004. **73**: p. 87-106.
63. Trifunovic A., Wredenberg A., Falkenberg M., Spelbrink J.N., Rovio A.T., Bruder C.E., Bohlooly Y.M., Gidlof S., Oldfors A., Wibom R., Tornell J., Jacobs H.T., and Larsson N.G., *Premature ageing in mice expressing defective mitochondrial DNA polymerase.* Nature, 2004. **429**(6990): p. 417-23.
64. Mootha V.K., Bunkenborg J., Olsen J.V., Hjerrild M., Wisniewski J.R., Stahl E., Bolouri M.S., Ray H.N., Sihag S., Kamal M., Patterson N., Lander E.S., and Mann M., *Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria.* Cell, 2003. **115**(5): p. 629-40.
65. Ohlmeier S., Kastaniotis A.J., Hiltunen J.K., and Bergmann U., *The yeast mitochondrial proteome, a study of fermentative and respiratory growth.* J Biol Chem, 2004. **279**(6): p. 3956-79.
66. Pflieger D., Le Caer J.P., Lemaire C., Bernard B.A., Dujardin G., and Rossier J., *Systematic identification of mitochondrial proteins by LC-MS/MS.* Anal Chem, 2002. **74**(10): p. 2400-6.
67. Prokisch H., Scharfe C., Camp D.G., 2nd, Xiao W., David L., Andreoli C., Monroe M.E., Moore R.J., Gritsenko M.A., Kozany C., Hixson K.K., Mottaz H.M., Zischka H., Ueffing M., Herman Z.S., Davis R.W., Meitinger T., Oefner P.J., Smith R.D., and Steinmetz L.M., *Integrative analysis of the mitochondrial proteome in yeast.* PLoS Biol, 2004. **2**(6): p. e160.
68. Sickmann A., Reinders J., Wagner Y., Joppich C., Zahedi R., Meyer H.E., Schonfisch B., Perschil I., Chacinska A., Guiard B., Rehling P., Pfanner N., and Meisinger C., *The proteome of Saccharomyces cerevisiae mitochondria.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13207-12.
69. Taylor S.W., Fahy E., Zhang B., Glenn G.M., Warnock D.E., Wiley S., Murphy A.N., Gaucher S.P., Capaldi R.A., Gibson B.W., and Ghosh S.S., *Characterization of the human heart mitochondrial proteome.* Nat Biotechnol, 2003. **21**(3): p. 281-6.
70. Millar A.H. and Heazlewood J.L., *Genomic and proteomic analysis of mitochondrial carrier proteins in Arabidopsis.* Plant Physiol, 2003. **131**(2): p. 443-53.
71. Heazlewood J.L., Tonti-Filippini J.S., Gout A.M., Day D.A., Whelan J., and Millar A.H., *Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins.* Plant Cell, 2004. **16**(1): p. 241-56.
72. Heazlewood J.L., Howell K.A., Whelan J., and Millar A.H., *Towards an analysis of the rice mitochondrial proteome.* Plant Physiol, 2003. **132**(1): p. 230-42.

73. Bardel J., Louwagie M., Jaquinod M., Jourdain A., Luche S., Rabilloud T., Macherel D., Garin J., and Bourguignon J., *A survey of the plant mitochondrial proteome in relation to development*. Proteomics, 2002. **2**(7): p. 880-98.
74. Krufft V., Eubel H., Jansch L., Werhahn W., and Braun H.P., *Proteomic approach to identify novel mitochondrial proteins in Arabidopsis*. Plant Physiol, 2001. **127**(4): p. 1694-710.
75. Millar A.H., Sweetlove L.J., Giege P., and Leaver C.J., *Analysis of the Arabidopsis mitochondrial proteome*. Plant Physiol, 2001. **127**(4): p. 1711-27.
76. Taylor S.W., Warnock D.E., Glenn G.M., Zhang B., Fahy E., Gaucher S.P., Capaldi R.A., Gibson B.W., and Ghosh S.S., *An alternative strategy to determine the mitochondrial proteome using sucrose gradient fractionation and 1D PAGE on highly purified human heart mitochondria*. J Proteome Res, 2002. **1**(5): p. 451-8.
77. Scharfe C., Zaccaria P., Hoernagel K., Jaksch M., Klopstock T., Dembowski M., Lill R., Prokisch H., Gerbitz K.D., Neupert W., Mewes H.W., and Meitinger T., *MITOP, the mitochondrial proteome database: 2000 update*. Nucleic Acids Res, 2000. **28**(1): p. 155-8.
78. Andreoli C., Prokisch H., Hortnagel K., Mueller J.C., Munsterkotter M., Scharfe C., and Meitinger T., *MitoP2, an integrated database on mitochondrial proteins in yeast and man*. Nucleic Acids Res, 2004. **32**(Database issue): p. D459-62.
79. Dimmer K.S., Fritz S., Fuchs F., Messerschmitt M., Weinbach N., Neupert W., and Westermann B., *Genetic basis of mitochondrial function and morphology in Saccharomyces cerevisiae*. Mol Biol Cell, 2002. **13**(3): p. 847-53.
80. Steinmetz L.M., Scharfe C., Deutschbauer A.M., Mokranjac D., Herman Z.S., Jones T., Chu A.M., Giaever G., Prokisch H., Oefner P.J., and Davis R.W., *Systematic screen for human disease genes in yeast*. Nat Genet, 2002. **31**(4): p. 400-4.
81. Kumar A., Agarwal S., Heyman J.A., Matson S., Heidtman M., Piccirillo S., Umansky L., Drawid A., Jansen R., Liu Y., Cheung K.H., Miller P., Gerstein M., Roeder G.S., and Snyder M., *Subcellular localization of the yeast proteome*. Genes Dev, 2002. **16**(6): p. 707-19.
82. Huh W.K., Falvo J.V., Gerke L.C., Carroll A.S., Howson R.W., Weissman J.S., and O'Shea E.K., *Global analysis of protein localization in budding yeast*. Nature, 2003. **425**(6959): p. 686-91.
83. Claros M.G. and Vincens P., *Computational method to predict mitochondrially imported proteins and their targeting sequences*. Eur J Biochem, 1996. **241**(3): p. 779-86.
84. Nakai K. and Horton P., *PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization*. Trends Biochem Sci, 1999. **24**(1): p. 34-6.
85. Small I., Peeters N., Legeai F., and Lurin C., *Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences*. Proteomics, 2004. **4**(6): p. 1581-90.
86. Drawid A. and Gerstein M., *A Bayesian system integrating expression data with sequence patterns for localizing proteins: comprehensive application to the yeast genome*. J Mol Biol, 2000. **301**(4): p. 1059-75.

87. Lascaris R., Bussemaker H.J., Boorsma A., Piper M., van der Spek H., Grivell L., and Blom J., *Hap4p overexpression in glucose-grown Saccharomyces cerevisiae induces cells to enter a novel metabolic state*. Genome Biol, 2003. **4**(1): p. R3.
88. DeRisi J.L., Iyer V.R., and Brown P.O., *Exploring the metabolic and genetic control of gene expression on a genomic scale*. Science, 1997. **278**(5338): p. 680-6.
89. von Mering C., Krause R., Snel B., Cornell M., Oliver S.G., Fields S., and Bork P., *Comparative assessment of large-scale data sets of protein-protein interactions*. Nature, 2002. **417**(6887): p. 399-403.
90. Zischka H., Weber G., Weber P.J., Posch A., Braun R.J., Buhringer D., Schneider U., Nissum M., Meitinger T., Ueffing M., and Eckerskorn C., *Improved proteome analysis of Saccharomyces cerevisiae mitochondria by free-flow electrophoresis*. Proteomics, 2003. **3**(6): p. 906-16.
91. Ozawa T., Sako Y., Sato M., Kitamura T., and Umezawa Y., *A genetic approach to identifying mitochondrial proteins*. Nat Biotechnol, 2003. **21**(3): p. 287-93.
92. Tong A.H., Lesage G., Bader G.D., Ding H., Xu H., Xin X., Young J., Berriz G.F., Brost R.L., Chang M., Chen Y., Cheng X., Chua G., Friesen H., Goldberg D.S., Haynes J., Humphries C., He G., Hussein S., Ke L., Krogan N., Li Z., Levinson J.N., Lu H., Menard P., Munyana C., Parsons A.B., Ryan O., Tonikian R., Roberts T., Sdicu A.M., Shapiro J., Sheikh B., Suter B., Wong S.L., Zhang L.V., Zhu H., Burd C.G., Munro S., Sander C., Rine J., Greenblatt J., Peter M., Bretscher A., Bell G., Roth F.P., Brown G.W., Andrews B., Bussey H., and Boone C., *Global mapping of the yeast genetic interaction network*. Science, 2004. **303**(5659): p. 808-13.
93. Causton H.C., Ren B., Koh S.S., Harbison C.T., Kanin E., Jennings E.G., Lee T.I., True H.L., Lander E.S., and Young R.A., *Remodeling of yeast genome expression in response to environmental changes*. Mol Biol Cell, 2001. **12**(2): p. 323-37.
94. Hughes T.R., Marton M.J., Jones A.R., Roberts C.J., Stoughton R., Armour C.D., Bennett H.A., Coffey E., Dai H., He Y.D., Kidd M.J., King A.M., Meyer M.R., Slade D., Lum P.Y., Stepaniants S.B., Shoemaker D.D., Gachotte D., Chakraborty K., Simon J., Bard M., and Friend S.H., *Functional discovery via a compendium of expression profiles*. Cell, 2000. **102**(1): p. 109-26.
95. O'Brien E.A., Badidi E., Barbasiewicz A., deSousa C., Lang B.F., and Burger G., *GOBASE--a database of mitochondrial and chloroplast information*. Nucleic Acids Res, 2003. **31**(1): p. 176-8.
96. Kogelnik A.M., Lott M.T., Brown M.D., Navathe S.B., and Wallace D.C., *MITOMAP: a human mitochondrial genome database--1998 update*. Nucleic Acids Res, 1998. **26**(1): p. 112-5.
97. Guda C., Guda P., Fahy E., and Subramaniam S., *MITOPRED: a web server for the prediction of mitochondrial proteins*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W372-4.
98. Andersson S.G., Karlberg O., Canback B., and Kurland C.G., *On the origin of mitochondria: a genomics perspective*. Philos Trans R Soc Lond B Biol Sci, 2003. **358**(1429): p. 165-77; discussion 177-9.

99. Burger G. and Lang B.F., *Parallels in genome evolution in mitochondria and bacterial symbionts*. IUBMB Life, 2003. **55**(4-5): p. 205-12.
100. DiMauro S., *Mitochondrial medicine*. Biochim Biophys Acta, 2004. **1659**(2-3): p. 107-14.
101. DiMauro S., *Mitochondrial diseases*. Biochim Biophys Acta, 2004. **1658**(1-2): p. 80-8.
102. Wallace D.C., *Mitochondrial diseases in man and mouse*. Science, 1999. **283**(5407): p. 1482-8.
103. Zeviani M. and Carelli V., *Mitochondrial disorders*. Curr Opin Neurol, 2007. **20**(5): p. 564-71.
104. Bourgeron T., Rustin P., Chretien D., Birch-Machin M., Bourgeois M., Viegas-Pequignot E., Munnich A., and Rotig A., *Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency*. Nat Genet, 1995. **11**(2): p. 144-9.
105. Schaefer A.M., Taylor R.W., Turnbull D.M., and Chinnery P.F., *The epidemiology of mitochondrial disorders--past, present and future*. Biochim Biophys Acta, 2004. **1659**(2-3): p. 115-20.
106. Verma M., Kagan J., Sidransky D., and Srivastava S., *Proteomic analysis of cancer-cell mitochondria*. Nat Rev Cancer, 2003. **3**(10): p. 789-95.
107. Koutnikova H., Campuzano V., Foury F., Dolle P., Cazzalini O., and Koenig M., *Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin*. Nat Genet, 1997. **16**(4): p. 345-51.
108. Wilson R.B. and Roof D.M., *Respiratory deficiency due to loss of mitochondrial DNA in yeast lacking the frataxin homologue*. Nat Genet, 1997. **16**(4): p. 352-7.
109. Jin H., May M., Tranebjaerg L., Kendall E., Fontan G., Jackson J., Subramony S.H., Arena F., Lubs H., Smith S., Stevenson R., Schwartz C., and Vetrie D., *A novel X-linked gene, DDP, shows mutations in families with deafness (DFN-1), dystonia, mental deficiency and blindness*. Nat Genet, 1996. **14**(2): p. 177-80.
110. Roesch K., Curran S.P., Tranebjaerg L., and Koehler C.M., *Human deafness dystonia syndrome is caused by a defect in assembly of the DDP1/TIMM8a-TIMM13 complex*. Hum Mol Genet, 2002. **11**(5): p. 477-86.
111. Rothbauer U., Hofmann S., Muhlenbein N., Paschen S.A., Gerbitz K.D., Neupert W., Brunner M., and Bauer M.F., *Role of the deafness dystonia peptide 1 (DDP1) in import of human Tim23 into the inner membrane of mitochondria*. J Biol Chem, 2001. **276**(40): p. 37327-34.
112. Bauer M.F., Rothbauer U., Muhlenbein N., Smith R.J., Gerbitz K., Neupert W., Brunner M., and Hofmann S., *The mitochondrial TIM22 preprotein translocase is highly conserved throughout the eukaryotic kingdom*. FEBS Lett, 1999. **464**(1-2): p. 41-7.
113. Alexander C., Votruba M., Pesch U.E., Thiselton D.L., Mayer S., Moore A., Rodriguez M., Kellner U., Leo-Kottler B., Auburger G., Bhattacharya S.S., and Wissinger B., *OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28*. Nat Genet, 2000. **26**(2): p. 211-5.
114. Delettre C., Lenaers G., Griffoin J.M., Gigarel N., Lorenzo C., Belenguer P., Pelloquin L., Grosgeorge J., Turc-Carel C., Perret E., Astarie-Dequeker C., Lasquelléc L., Arnaud B., Ducommun B., Kaplan J., and Hamel C.P., *Nuclear gene OPA1, encoding a mitochondrial*

- dynamin-related protein, is mutated in dominant optic atrophy*. Nat Genet, 2000. **26**(2): p. 207-10.
115. Wong E.D., Wagner J.A., Scott S.V., Okreglak V., Holewinske T.J., Cassidy-Stone A., and Nunnari J., *The intramitochondrial dynamin-related GTPase, Mgm1p, is a component of a protein complex that mediates mitochondrial fusion*. J Cell Biol, 2003. **160**(3): p. 303-11.
116. Sesaki H., Southard S.M., Yaffe M.P., and Jensen R.E., *Mgm1p, a dynamin-related GTPase, is essential for fusion of the mitochondrial outer membrane*. Mol Biol Cell, 2003. **14**(6): p. 2342-56.
117. Zuchner S., Mersiyanova I.V., Muglia M., Bissar-Tadmouri N., Rochelle J., Dadali E.L., Zappia M., Nelis E., Patitucci A., Senderek J., Parman Y., Evgrafov O., Jonghe P.D., Takahashi Y., Tsuji S., Pericak-Vance M.A., Quattrone A., Battaloglu E., Polyakov A.V., Timmerman V., Schroder J.M., and Vance J.M., *Mutations in the mitochondrial GTPase mitofusin 2 cause Charcot-Marie-Tooth neuropathy type 2A*. Nat Genet, 2004. **36**(5): p. 449-51.
118. Gerbitz K.D., Gempel K., and Brdiczka D., *Mitochondria and diabetes. Genetic, biochemical, and clinical implications of the cellular energy circuit*. Diabetes, 1996. **45**(2): p. 113-26.
119. Orth M. and Schapira A.H., *Mitochondrial involvement in Parkinson's disease*. Neurochem Int, 2002. **40**(6): p. 533-41.
120. Hashimoto M., Rockenstein E., Crews L., and Masliah E., *Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases*. Neuromolecular Med, 2003. **4**(1-2): p. 21-36.
121. Batten B., *A family suffering from hereditary optic atrophy*. Trans Ophthalmol Soc UK, 1896. **16**: p. 125.
122. Snell S., *Disease of the optic nerve: I. Hereditary or congenital optic atrophy and allied cases*. Trans Ophthalmol Soc UK, 1897. **17**: p. 66-81.
123. Leber T., *Ueber hereditäre und congenital-angelegte Sehnervenleiden*. Graefes Arch Ophthalmol, 1871. **17**(2): p. 249-291.
124. Kjer P., *Infantile optic atrophy with dominant mode of inheritance: a clinical and genetic study of 19 Danish families*. Acta Ophthalmol Suppl, 1959. **164**(Suppl 54): p. 1-147.
125. Eiberg H., Kjer B., Kjer P., and Rosenberg T., *Dominant optic atrophy (OPA1) mapped to chromosome 3q region. I. Linkage analysis*. Hum Mol Genet, 1994. **3**(6): p. 977-80.
126. Lunkes A., Hartung U., Magarino C., Rodriguez M., Palmero A., Rodriguez L., Heredero L., Weissenbach J., Weber J., and Auburger G., *Refinement of the OPA1 gene locus on chromosome 3q28-q29 to a region of 2-8 cM, in one Cuban pedigree with autosomal dominant optic atrophy type Kjer*. Am J Hum Genet, 1995. **57**(4): p. 968-70.
127. Kerrison J.B., Koenekoop R.K., Arnould V.J., Zee D., and Maumenee I.H., *Clinical features of autosomal dominant congenital nystagmus linked to chromosome 6p12*. Am J Ophthalmol, 1998. **125**(1): p. 64-70.
128. Barbet F., Hakiki S., Orssaud C., Gerber S., Perrault I., Hanein S., Ducroq D., Dufier J.L., Munnich A., Kaplan J., and Rozet J.M., *A third locus for dominant optic atrophy on chromosome 22q*. J Med Genet, 2005. **42**(1): p. e1.

129. Assink J.J., Tijmes N.T., ten Brink J.B., Oostra R.J., Riemsdag F.C., de Jong P.T., and Bergen A.A., *A gene for X-linked optic atrophy is closely linked to the Xp11.4-Xp11.2 region of the X chromosome*. Am J Hum Genet, 1997. **61**(4): p. 934-9.
130. Anikster Y., Kleta R., Shaag A., Gahl W.A., and Elpeleg O., *Type III 3-methylglutaconic aciduria (optic atrophy plus syndrome, or Costeff optic atrophy syndrome): identification of the OPA3 gene and its founder mutation in Iraqi Jews*. Am J Hum Genet, 2001. **69**(6): p. 1218-24.
131. Barbet F., Gerber S., Hakiki S., Perrault I., Hanein S., Ducroq D., Tanguy G., Dufier J.L., Munnich A., Rozet J.M., and Kaplan J., *A first locus for isolated autosomal recessive optic atrophy (ROA1) maps to chromosome 8q*. Eur J Hum Genet, 2003. **11**(12): p. 966-71.
132. Hanein S., Perrault I., Roche O., Gerber S., Khadom N., Rio M., Boddaert N., Jean-Pierre M., Brahimi N., Serre V., Chretien D., Delphin N., Fares-Taie L., Lachheb S., Rotig A., Meire F., Munnich A., Dufier J.L., Kaplan J., and Rozet J.M., *TMEM126A, encoding a mitochondrial protein, is mutated in autosomal-recessive nonsyndromic optic atrophy*. Am J Hum Genet, 2009. **84**(4): p. 493-8.
133. Man P.Y., Griffiths P.G., Brown D.T., Howell N., Turnbull D.M., and Chinnery P.F., *The epidemiology of Leber hereditary optic neuropathy in the North East of England*. Am J Hum Genet, 2003. **72**(2): p. 333-9.
134. Puomila A., Hamalainen P., Kivioja S., Savontaus M.L., Koivumaki S., Huoponen K., and Nikoskelainen E., *Epidemiology and penetrance of Leber hereditary optic neuropathy in Finland*. Eur J Hum Genet, 2007. **15**(10): p. 1079-89.
135. Lyle W.M., *Genetic risks : a reference for eye care practitioners*. 1990, Waterloo, Ont., Canada: University of Waterloo Press. 1 v. (various pagings).
136. Kjer B., Eiberg H., Kjer P., and Rosenberg T., *Dominant optic atrophy mapped to chromosome 3q region. II. Clinical and epidemiological aspects*. Acta Ophthalmol Scand, 1996. **74**(1): p. 3-7.
137. Thiselton D.L., Alexander C., Morris A., Brooks S., Rosenberg T., Eiberg H., Kjer B., Kjer P., Bhattacharya S.S., and Votruba M., *A frameshift mutation in exon 28 of the OPA1 gene explains the high prevalence of dominant optic atrophy in the Danish population: evidence for a founder effect*. Hum Genet, 2001. **109**(5): p. 498-502.
138. Reynier P., Amati-Bonneau P., Verny C., Olichon A., Simard G., Guichet A., Bonnemains C., Malecaze F., Malinge M.C., Pelletier J.B., Calvas P., Dollfus H., Belenguer P., Malthiery Y., Lenaers G., and Bonneau D., *OPA3 gene mutations responsible for autosomal dominant optic atrophy and cataract*. J Med Genet, 2004. **41**(9): p. e110.
139. Cornille K., Milea D., Amati-Bonneau P., Procaccio V., Zazoun L., Guillet V., El Achouri G., Delettre C., Gueguen N., Loiseau D., Muller A., Ferre M., Chevrollier A., Wallace D.C., Bonneau D., Hamel C., Reynier P., and Lenaers G., *Reversible optic neuropathy with OPA1 exon 5b mutation*. Ann Neurol, 2008. **63**(5): p. 667-71.
140. Hoyt C.S., *Autosomal dominant optic atrophy. A spectrum of disability*. Ophthalmology, 1980. **87**(3): p. 245-51.

141. Votruba M., Moore A.T., and Bhattacharya S.S., *Clinical features, molecular genetics, and pathophysiology of dominant optic atrophy*. J Med Genet, 1998. **35**(10): p. 793-800.
142. Johnston P.B., Gaster R.N., Smith V.C., and Tripathi R.C., *A clinicopathologic study of autosomal dominant optic atrophy*. Am J Ophthalmol, 1979. **88**(5): p. 868-75.
143. Kjer P., Jensen O.A., and Klinken L., *Histopathology of eye, optic nerve and brain in a case of dominant optic atrophy*. Acta Ophthalmol (Copenh), 1983. **61**(2): p. 300-12.
144. Ito Y., Nakamura M., Yamakoshi T., Lin J., Yatsuya H., and Terasaki H., *Reduction of inner retinal thickness in patients with autosomal dominant optic atrophy associated with OPA1 mutations*. Invest Ophthalmol Vis Sci, 2007. **48**(9): p. 4079-86.
145. Milea D., Sander B., Wegener M., Jensen H., Kjer B., Jorgensen T.M., Lund-Andersen H., and Larsen M., *Axonal loss occurs early in dominant optic atrophy*. Acta Ophthalmol, 2009.
146. Carelli V., Ross-Cisneros F.N., and Sadun A.A., *Mitochondrial dysfunction as a cause of optic neuropathies*. Prog Retin Eye Res, 2004. **23**(1): p. 53-89.
147. Baris O., Delettre C., Amati-Bonneau P., Surget M.O., Charlin J.F., Catier A., Derieux L., Guyomard J.L., Dollfus H., Jonveaux P., Ayuso C., Maumenee I., Lorenz B., Mohammed S., Tourmen Y., Bonneau D., Malthiery Y., Hamel C., and Reynier P., *Fourteen novel OPA1 mutations in autosomal dominant optic atrophy including two de novo mutations in sporadic optic atrophy*. Hum Mutat, 2003. **21**(6): p. 656.
148. Cohn A.C., Toomes C., Potter C., Towns K.V., Hewitt A.W., Inglehearn C.F., Craig J.E., and Mackey D.A., *Autosomal dominant optic atrophy: penetrance and expressivity in patients with OPA1 mutations*. Am J Ophthalmol, 2007. **143**(4): p. 656-62.
149. Delettre C., Griffoin J.M., Kaplan J., Dollfus H., Lorenz B., Faivre L., Lenaers G., Belenguer P., and Hamel C.P., *Mutation spectrum and splicing variants in the OPA1 gene*. Hum Genet, 2001. **109**(6): p. 584-91.
150. Marchbank N.J., Craig J.E., Leek J.P., Toohey M., Churchill A.J., Markham A.F., Mackey D.A., Toomes C., and Inglehearn C.F., *Deletion of the OPA1 gene in a dominant optic atrophy family: evidence that haploinsufficiency is the cause of disease*. J Med Genet, 2002. **39**(8): p. e47.
151. Pesch U.E., Leo-Kottler B., Mayer S., Jurklics B., Kellner U., Apfelstedt-Sylla E., Zrenner E., Alexander C., and Wissinger B., *OPA1 mutations in patients with autosomal dominant optic atrophy and evidence for semi-dominant inheritance*. Hum Mol Genet, 2001. **10**(13): p. 1359-68.
152. Thiselton D.L., Alexander C., Taanman J.W., Brooks S., Rosenberg T., Eiberg H., Andreasson S., Van Regemorter N., Munier F.L., Moore A.T., Bhattacharya S.S., and Votruba M., *A comprehensive survey of mutations in the OPA1 gene in patients with autosomal dominant optic atrophy*. Invest Ophthalmol Vis Sci, 2002. **43**(6): p. 1715-24.
153. Toomes C., Marchbank N.J., Mackey D.A., Craig J.E., Newbury-Ecob R.A., Bennett C.P., Vize C.J., Desai S.P., Black G.C., Patel N., Teimory M., Markham A.F., Inglehearn C.F., and Churchill A.J., *Spectrum, frequency and penetrance of OPA1 mutations in dominant optic atrophy*. Hum Mol Genet, 2001. **10**(13): p. 1369-78.

154. Kivlin J.D., Lovrien E.W., Bishop D.T., and Maumenee I.H., *Linkage analysis in dominant optic atrophy*. Am J Hum Genet, 1983. **35**(6): p. 1190-5.
155. Puomila A., Huoponen K., Mantyjarvi M., Hamalainen P., Paananen R., Sankila E.M., Savontaus M.L., Somer M., and Nikoskelainen E., *Dominant optic atrophy: correlation between clinical and molecular genetic studies*. Acta Ophthalmol Scand, 2005. **83**(3): p. 337-46.
156. Han J., Thompson-Lowrey A.J., Reiss A., Mayorov V., Jia H., Biousse V., Newman N.J., and Brown M.D., *OPA1 mutations and mitochondrial DNA haplotypes in autosomal dominant optic atrophy*. Genet Med, 2006. **8**(4): p. 217-25.
157. Ferre M., Bonneau D., Milea D., Chevrollier A., Verny C., Dollfus H., Ayuso C., Defoort S., Vignal C., Zanlonghi X., Charlin J.F., Kaplan J., Odent S., Hamel C.P., Procaccio V., Reynier P., and Amati-Bonneau P., *Molecular screening of 980 cases of suspected hereditary optic neuropathy with a report on 77 Novel OPA1 mutations*. Hum Mutat, 2009.
158. Lodi R., Tonon C., Valentino M.L., Iotti S., Clementi V., Malucelli E., Barboni P., Longanesi L., Schimpf S., Wissinger B., Baruzzi A., Barbiroli B., and Carelli V., *Deficit of in vivo mitochondrial ATP production in OPA1-related dominant optic atrophy*. Ann Neurol, 2004. **56**(5): p. 719-23.
159. Chevrollier A., Guillet V., Loiseau D., Gueguen N., de Crescenzo M.A., Verny C., Ferre M., Dollfus H., Odent S., Milea D., Goizet C., Amati-Bonneau P., Procaccio V., Bonneau D., and Reynier P., *Hereditary optic neuropathies share a common mitochondrial coupling defect*. Ann Neurol, 2008. **63**(6): p. 794-8.
160. Nochez Y., Arsene S., Gueguen N., Chevrollier A., Ferre M., Guillet V., Desquirit V., Toutain A., Bonneau D., Procaccio V., Amati-Bonneau P., Pisella P.J., and Reynier P., *Acute and late-onset optic atrophy due to a novel OPA1 mutation leading to a mitochondrial coupling defect*. Mol Vis, 2009. **15**: p. 598-608.
161. Shimizu S., Mori N., Kishi M., Sugata H., Tsuda A., and Kubota N., *A novel mutation in the OPA1 gene in a Japanese patient with optic atrophy*. Am J Ophthalmol, 2003. **135**(2): p. 256-7.
162. Amati-Bonneau P., Odent S., Derrien C., Pasquier L., Malthiery Y., Reynier P., and Bonneau D., *The association of autosomal dominant optic atrophy and moderate deafness may be due to the R445H mutation in the OPA1 gene*. Am J Ophthalmol, 2003. **136**(6): p. 1170-1.
163. Payne M., Yang Z., Katz B.J., Warner J.E., Weight C.J., Zhao Y., Pearson E.D., Treft R.L., Hillman T., Kennedy R.J., Meire F.M., and Zhang K., *Dominant optic atrophy, sensorineural hearing loss, ptosis, and ophthalmoplegia: a syndrome caused by a missense mutation in OPA1*. Am J Ophthalmol, 2004. **138**(5): p. 749-55.
164. Li C., Kosmorsky G., Zhang K., Katz B.J., Ge J., and Traboulsi E.I., *Optic atrophy and sensorineural hearing loss in a family caused by an R445H OPA1 mutation*. Am J Med Genet A, 2005. **138A**(3): p. 208-11.
165. Amati-Bonneau P., Guichet A., Olichon A., Chevrollier A., Viala F., Miot S., Ayuso C., Odent S., Arrouet C., Verny C., Calmels M.N., Simard G., Belenguer P., Wang J., Puel J.L., Hamel

- C., Malthiery Y., Bonneau D., Lenaers G., and Reynier P., *OPA1 R445H mutation in optic atrophy associated with sensorineural deafness*. *Ann Neurol*, 2005. **58**(6): p. 958-63.
166. Ke T., Nie S.W., Yang Q.B., Liu J.P., Zhou L.N., Ren X., Liu J.Y., Wang Q., and Liu M.G., *The G401D mutation of OPA1 causes autosomal dominant optic atrophy and hearing loss in a Chinese family*. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*, 2006. **23**(5): p. 481-5.
167. Amati-Bonneau P., Valentino M.L., Reynier P., Gallardo M.E., Bornstein B., Boissiere A., Campos Y., Rivera H., de la Aleja J.G., Carroccia R., Iommarini L., Labauge P., Figarella-Branger D., Marcocelles P., Furby A., Beauvais K., Letournel F., Liguori R., La Morgia C., Montagna P., Liguori M., Zanna C., Rugolo M., Cossarizza A., Wissinger B., Verny C., Schwarzenbacher R., Martin M.A., Arenas J., Ayuso C., Garesse R., Lenaers G., Bonneau D., and Carelli V., *OPA1 mutations induce mitochondrial DNA instability and optic atrophy 'plus' phenotypes*. *Brain*, 2008. **131**(Pt 2): p. 338-51.
168. Hudson G., Amati-Bonneau P., Blakely E.L., Stewart J.D., He L., Schaefer A.M., Griffiths P.G., Ahlqvist K., Suomalainen A., Reynier P., McFarland R., Turnbull D.M., Chinnery P.F., and Taylor R.W., *Mutation of OPA1 causes dominant optic atrophy with external ophthalmoplegia, ataxia, deafness and multiple mitochondrial DNA deletions: a novel disorder of mtDNA maintenance*. *Brain*, 2008. **131**(Pt 2): p. 329-37.
169. Liguori M., La Russa A., Manna I., Andreoli V., Caracciolo M., Spadafora P., Cittadella R., and Quattrone A., *A phenotypic variation of dominant optic atrophy and deafness (ADOAD) due to a novel OPA1 mutation*. *J Neurol*, 2008. **255**(1): p. 127-9.
170. Spinazzi M., Cazzola S., Bortolozzi M., Baracca A., Loro E., Casarin A., Solaini G., Sgarbi G., Casalena G., Cenacchi G., Malena A., Frezza C., Carrara F., Angelini C., Scorrano L., Salviati L., and Vergani L., *A novel deletion in the GTPase domain of OPA1 causes defects in mitochondrial morphology and distribution, but not in function*. *Hum Mol Genet*, 2008. **17**(21): p. 3291-302.
171. Verny C., Loiseau D., Scherer C., Lejeune P., Chevrollier A., Gueguen N., Guillet V., Dubas F., Reynier P., Amati-Bonneau P., and Bonneau D., *Multiple sclerosis-like disorder in OPA1-related autosomal dominant optic atrophy*. *Neurology*, 2008. **70**(13 Pt 2): p. 1152-3.
172. Harding A.E., Sweeney M.G., Miller D.H., Mumford C.J., Kellar-Wood H., Menard D., McDonald W.I., and Compston D.A., *Occurrence of a multiple sclerosis-like illness in women who have a Leber's hereditary optic neuropathy mitochondrial DNA mutation*. *Brain*, 1992. **115** (Pt 4): p. 979-89.
173. Konigsmark B.W., Knox D.L., Hussels I.E., and Moses H., *Dominant congenital deafness and progressive optic nerve atrophy. Occurrence in four generations of a family*. *Arch Ophthalmol*, 1974. **91**(2): p. 99-103.
174. Meire F., De Laey J.J., de Bie S., van Staey M., and Matton M.T., *Dominant optic nerve atrophy with progressive hearing loss and chronic progressive external ophthalmoplegia (CPEO)*. *Ophthalmic Paediatr Genet*, 1985. **5**(1-2): p. 91-7.
175. Mets M.B. and Mhoon E., *Probable autosomal dominant optic atrophy with hearing loss*. *Ophthalmic Paediatr Genet*, 1985. **5**(1-2): p. 85-9.

176. Zeviani M., *OPA1 mutations and mitochondrial DNA damage: keeping the magic circle in shape*. Brain, 2008. **131**(Pt 2): p. 314-7.
177. Stewart J.D., Hudson G., Yu-Wai-Man P., Blakeley E.L., He L., Horvath R., Maddison P., Wright A., Griffiths P.G., Turnbull D.M., Taylor R.W., and Chinnery P.F., *OPA1 in multiple mitochondrial DNA deletion disorders*. Neurology, 2008. **71**(22): p. 1829-31.
178. Milone M., Younge B.R., Wang J., Zhang S., and Wong L.J., *Mitochondrial disorder with OPA1 mutation lacking optic atrophy*. Mitochondrion, 2009. **9**(4): p. 279-81.
179. Ferre M., Amati-Bonneau P., Tourmen Y., Malthiery Y., and Reynier P., *eOPA1: an online database for OPA1 mutations*. Hum Mutat, 2005. **25**(5): p. 423-8.
180. Fuhrmann N., Alavi M.V., Bitoun P., Woernle S., Auburger G., Leo-Kottler B., Yu-Wai-Man P., Chinnery P., and Wissinger B., *Genomic rearrangements in OPA1 are frequent in patients with autosomal dominant optic atrophy*. J Med Genet, 2009. **46**(2): p. 136-44.
181. Misaka T., Miyashita T., and Kubo Y., *Primary structure of a dynamin-related mouse mitochondrial GTPase and its distribution in brain, subcellular localization, and effect on mitochondrial morphology*. J Biol Chem, 2002. **277**(18): p. 15834-42.
182. Olichon A., Emorine L.J., Descoins E., Pelloquin L., Bricchese L., Gas N., Guillou E., Delettre C., Valette A., Hamel C.P., Ducommun B., Lenaers G., and Belenguer P., *The human dynamin-related protein OPA1 is anchored to the mitochondrial inner membrane facing the inter-membrane space*. FEBS Lett, 2002. **523**(1-3): p. 171-6.
183. Satoh M., Hamamoto T., Seo N., Kagawa Y., and Endo H., *Differential sublocalization of the dynamin-related protein OPA1 isoforms in mitochondria*. Biochem Biophys Res Commun, 2003. **300**(2): p. 482-93.
184. Ishihara N., Fujita Y., Oka T., and Mihara K., *Regulation of mitochondrial morphology through proteolytic cleavage of OPA1*. EMBO J, 2006. **25**(13): p. 2966-77.
185. Lenaers G., Reynier P., Elachouri G., Soukkaieh C., Olichon A., Belenguer P., Baricault L., Ducommun B., Hamel C., and Delettre C., *OPA1 functions in mitochondria and dysfunctions in optic nerve*. Int J Biochem Cell Biol, 2009. **41**(10): p. 1866-74.
186. Olichon A., Baricault L., Gas N., Guillou E., Valette A., Belenguer P., and Lenaers G., *Loss of OPA1 perturbs the mitochondrial inner membrane structure and integrity, leading to cytochrome c release and apoptosis*. J Biol Chem, 2003. **278**(10): p. 7743-6.
187. Olichon A., Landes T., Arnaune-Pelloquin L., Emorine L.J., Mills V., Guichet A., Delettre C., Hamel C., Amati-Bonneau P., Bonneau D., Reynier P., Lenaers G., and Belenguer P., *Effects of OPA1 mutations on mitochondrial morphology and apoptosis: relevance to ADOA pathogenesis*. J Cell Physiol, 2007. **211**(2): p. 423-30.
188. Zanna C., Ghelli A., Porcelli A.M., Karbowski M., Youle R.J., Schimpf S., Wissinger B., Pinti M., Cossarizza A., Vidoni S., Valentino M.L., Rugolo M., and Carelli V., *OPA1 mutations associated with dominant optic atrophy impair oxidative phosphorylation and mitochondrial fusion*. Brain, 2008. **131**(Pt 2): p. 352-67.

189. Legros F., Lombes A., Frachon P., and Rojo M., *Mitochondrial fusion in human cells is efficient, requires the inner membrane potential, and is mediated by mitofusins*. Mol Biol Cell, 2002. **13**(12): p. 4343-54.
190. Guillery O., Malka F., Landes T., Guillou E., Blackstone C., Lombes A., Belenguer P., Arnoult D., and Rojo M., *Metalloprotease-mediated OPA1 processing is modulated by the mitochondrial membrane potential*. Biol Cell, 2008. **100**(5): p. 315-25.
191. Cipolat S., Martins de Brito O., Dal Zilio B., and Scorrano L., *OPA1 requires mitofusin 1 to promote mitochondrial fusion*. Proc Natl Acad Sci U S A, 2004. **101**(45): p. 15927-32.
192. Olichon A., Elachouri G., Baricault L., Delettre C., Belenguer P., and Lenaers G., *OPA1 alternate splicing uncouples an evolutionary conserved function in mitochondrial fusion from a vertebrate restricted function in apoptosis*. Cell Death Differ, 2007. **14**(4): p. 682-92.
193. Scorrano L., Ashiya M., Buttle K., Weiler S., Oakes S.A., Mannella C.A., and Korsmeyer S.J., *A distinct pathway remodels mitochondrial cristae and mobilizes cytochrome c during apoptosis*. Dev Cell, 2002. **2**(1): p. 55-67.
194. Frezza C., Cipolat S., Martins de Brito O., Micaroni M., Beznoussenko G.V., Rudka T., Bartoli D., Polishuck R.S., Danial N.N., De Strooper B., and Scorrano L., *OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion*. Cell, 2006. **126**(1): p. 177-89.
195. Yamaguchi R., Lartigue L., Perkins G., Scott R.T., Dixit A., Kushnareva Y., Kuwana T., Ellisman M.H., and Newmeyer D.D., *Opa1-mediated cristae opening is Bax/Bak and BH3 dependent, required for apoptosis, and independent of Bak oligomerization*. Mol Cell, 2008. **31**(4): p. 557-69.
196. Schimpf S., Fuhrmann N., Schaich S., and Wissinger B., *Comprehensive cDNA study and quantitative transcript analysis of mutant OPA1 transcripts containing premature termination codons*. Hum Mutat, 2008. **29**(1): p. 106-12.
197. Aijaz S., Erskine L., Jeffery G., Bhattacharya S.S., and Votruba M., *Developmental expression profile of the optic atrophy gene product: OPA1 is not localized exclusively in the mammalian retinal ganglion cell layer*. Invest Ophthalmol Vis Sci, 2004. **45**(6): p. 1667-73.
198. Bette S., Schlaszus H., Wissinger B., Meyermann R., and Mittelbronn M., *OPA1, associated with autosomal dominant optic atrophy, is widely expressed in the human brain*. Acta Neuropathol, 2005. **109**(4): p. 393-9.
199. Kamei S., Chen-Kuo-Chang M., Cazevieille C., Lenaers G., Olichon A., Belenguer P., Roussignol G., Renard N., Eybalin M., Michelin A., Delettre C., Brabet P., and Hamel C.P., *Expression of the Opa1 mitochondrial protein in retinal ganglion cells: its downregulation causes aggregation of the mitochondrial network*. Invest Ophthalmol Vis Sci, 2005. **46**(11): p. 4288-94.
200. Pesch U.E., Fries J.E., Bette S., Kalbacher H., Wissinger B., Alexander C., and Kohler K., *OPA1, the disease gene for autosomal dominant optic atrophy, is specifically expressed in ganglion cells and intrinsic neurons of the retina*. Invest Ophthalmol Vis Sci, 2004. **45**(11): p. 4217-25.

201. Huizing M., Brooks B.P., and Anikster Y., *Optic atrophies in metabolic disorders*. Mol Genet Metab, 2005. **86**(1-2): p. 51-60.
202. Kim J.Y., Hwang J.M., Ko H.S., Seong M.W., Park B.J., and Park S.S., *Mitochondrial DNA content is decreased in autosomal dominant optic atrophy*. Neurology, 2005. **64**(6): p. 966-72.
203. Mayorov V.I., Lowrey A.J., Biousse V., Newman N.J., Cline S.D., and Brown M.D., *Mitochondrial oxidative phosphorylation in autosomal dominant optic atrophy*. BMC Biochem, 2008. **9**: p. 22.
204. Alavi M.V., Bette S., Schimpf S., Schuettauf F., Schraermeyer U., Wehrl H.F., Ruttiger L., Beck S.C., Tonagel F., Pichler B.J., Knipper M., Peters T., Laufs J., and Wissinger B., *A splice site mutation in the murine Opa1 gene features pathology of autosomal dominant optic atrophy*. Brain, 2007. **130**(Pt 4): p. 1029-42.
205. Davies V.J., Hollins A.J., Piechota M.J., Yip W., Davies J.R., White K.E., Nicols P.P., Boulton M.E., and Votruba M., *Opa1 deficiency in a mouse model of autosomal dominant optic atrophy impairs mitochondrial morphology, optic nerve structure and visual function*. Hum Mol Genet, 2007. **16**(11): p. 1307-18.
206. Cipolat S., Rudka T., Hartmann D., Costa V., Serneels L., Craessaerts K., Metzger K., Frezza C., Annaert W., D'Adamio L., Derks C., Dejaegere T., Pellegrini L., D'Hooge R., Scorrano L., and De Strooper B., *Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling*. Cell, 2006. **126**(1): p. 163-75.
207. Yarosh W., Monserrate J., Tong J.J., Tse S., Le P.K., Nguyen K., Brachmann C.B., Wallace D.C., and Huang T., *The molecular mechanisms of OPA1-mediated optic atrophy in Drosophila model and prospects for antioxidant treatment*. PLoS Genet, 2008. **4**(1): p. e6.
208. Waterham H.R., Koster J., van Roermund C.W., Mooyer P.A., Wanders R.J., and Leonard J.V., *A lethal defect of mitochondrial and peroxisomal fission*. N Engl J Med, 2007. **356**(17): p. 1736-41.
209. Rusch S.L. and Kendall D.A., *Protein transport via amino-terminal targeting sequences: common themes in diverse systems*. Mol Membr Biol, 1995. **12**(4): p. 295-307.
210. Schatz G. and Dobberstein B., *Common principles of protein translocation across membranes*. Science, 1996. **271**(5255): p. 1519-26.
211. Tourmen Y., Ferre M., Malthiery Y., Dessen P., and Reynier P., *Mitochondrial diseases preferentially involve proteins with prokaryote homologues*. C R Biol, 2004. **327**(12): p. 1095-101.
212. Lavenier D., Georges G., and Liu X., *A Reconfigurable Index FLASH Memory tailored to Seed-Based Genomic Sequence Comparison Algorithms*. The Journal of VLSI Signal Processing, 2007. **48**(3): p. 255-269.
213. Schneider M., Lane L., Boutet E., Lieberherr D., Tognolli M., Bougueleret L., and Bairoch A., *The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program*. J Proteomics, 2009. **72**(3): p. 567-73.
214. Kuhn R.M., Karolchik D., Zweig A.S., Wang T., Smith K.E., Rosenbloom K.R., Rhead B., Raney B.J., Pohl A., Pheasant M., Meyer L., Hsu F., Hinrichs A.S., Harte R.A., Giardine B.,

- Fujita P., Diekhans M., Dreszer T., Clawson H., Barber G.P., Haussler D., and Kent W.J., *The UCSC Genome Browser Database: update 2009*. Nucleic Acids Res, 2009. **37**(Database issue): p. D755-61.
215. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., and Lipman D.J., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
216. Giraud M., Kucherov G., Lavenier D., Noe L., and Peterlongo P. *Utilization of Subset Seeds on a Reconfigurable Architecture*. in *LAW 2007, London Algorithm Workshop*. 2007. King's College London, UK.
217. Peterlongo P., Noe L., Lavenier D., Georges G., Jacques J., Kucherov G., and Giraud M. *Protein similarity search with subset seeds on a dedicated reconfigurable hardware*. in *PBC 2007, Workshop on Parallel Computational Biology*. 2007. Gdansk, Poland.
218. Finocchiaro G., Colombo I., Garavaglia B., Gellera C., Valdameri G., Garbuglio N., and Didonato S., *cDNA cloning and mitochondrial import of the beta-subunit of the human electron-transfer flavoprotein*. Eur J Biochem, 1993. **213**(3): p. 1003-8.
219. Pushpa-Rekha T.R., Burdsall A.L., Oleksa L.M., Chisolm G.M., and Driscoll D.M., *Rat phospholipid-hydroperoxide glutathione peroxidase. cDNA cloning and identification of multiple transcription and translation start sites*. J Biol Chem, 1995. **270**(45): p. 26993-9.
220. Emanuelsson O., Nielsen H., Brunak S., and von Heijne G., *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence*. J Mol Biol, 2000. **300**(4): p. 1005-16.
221. Hua S. and Sun Z., *Support vector machine approach for protein subcellular localization prediction*. Bioinformatics, 2001. **17**(8): p. 721-8.
222. Xie D., Li A., Wang M., Fan Z., and Feng H., *LOCSVMPSI: a web server for subcellular localization of eukaryotic proteins using SVM and profile of PSI-BLAST*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W105-10.
223. Pruitt K.D., Tatusova T., and Maglott D.R., *NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins*. Nucleic Acids Res, 2007. **35**(Database issue): p. D61-5.
224. Burge C. and Karlin S., *Prediction of complete gene structures in human genomic DNA*. J Mol Biol, 1997. **268**(1): p. 78-94.
225. Blanco E., Parra G., and Guigo R., *Using geneid to identify genes*. Curr Protoc Bioinformatics, 2007. **Chapter 4**: p. Unit 4 3.
226. Gross S.S. and Brent M.R., *Using multiple alignments to improve gene prediction*. J Comput Biol, 2006. **13**(2): p. 379-93.
227. Ingram V.M., *Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin*. Nature, 1957. **180**(4581): p. 326-8.
228. Pauling L., Itano H.A., and et al., *Sickle cell anemia a molecular disease*. Science, 1949. **110**(2865): p. 543-8.

229. Sanger F., Nicklen S., and Coulson A.R., *DNA sequencing with chain-terminating inhibitors*. Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
230. McKusick V.A., *Mendelian Inheritance in Man and its online version, OMIM*. Am J Hum Genet, 2007. **80**(4): p. 588-604.
231. Antonarakis S.E., *Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group*. Hum Mutat, 1998. **11**(1): p. 1-3.
232. den Dunnen J.T. and Antonarakis S.E., *Nomenclature for the description of human sequence variations*. Hum Genet, 2001. **109**(1): p. 121-4.
233. Carelli V., La Morgia C., Iommarini L., Carroccia R., Mattiazzi M., Sangiorgi S., Farne S., Maresca A., Foscari B., Lanzi L., Amadori M., Bellan M., and Valentino M.L., *Mitochondrial optic neuropathies: how two genomes may kill the same cell type?* Biosci Rep, 2007. **27**(1-3): p. 173-84.
234. Taylor S.W., Fahy E., and Ghosh S.S., *Global organellar proteomics*. Trends Biotechnol, 2003. **21**(2): p. 82-8.
235. Richly E., Chinnery P.F., and Leister D., *Evolutionary diversification of mitochondrial proteomes: implications for human disease*. Trends Genet, 2003. **19**(7): p. 356-62.
236. Claustres M., Horaitis O., Vanevski M., and Cotton R.G., *Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases*. Genome Res, 2002. **12**(5): p. 680-8.
237. Scriver C.R., Nowacki P.M., and Lehvaslaiho H., *Guidelines and recommendations for content, structure, and deployment of mutation databases: II. Journey in progress*. Hum Mutat, 2000. **15**(1): p. 13-5.
238. Scriver C.R., Nowacki P.M., and Lehvaslaiho H., *Guidelines and recommendations for content, structure, and deployment of mutation databases*. Hum Mutat, 1999. **13**(5): p. 344-50.
239. Beroud C., Hamroun D., Collod-Beroud G., Boileau C., Soussi T., and Claustres M., *UMD (Universal Mutation Database): 2005 update*. Hum Mutat, 2005. **26**(3): p. 184-91.
240. Beroud C., Collod-Beroud G., Boileau C., Soussi T., and Junien C., *UMD (Universal mutation database): a generic software to build and analyze locus-specific databases*. Hum Mutat, 2000. **15**(1): p. 86-94.
241. Fokkema I.F., den Dunnen J.T., and Taschner P.E., *LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach*. Hum Mutat, 2005. **26**(2): p. 63-8.
242. Beroud C., Tuffery-Giraud S., Matsuo M., Hamroun D., Humbertclaude V., Monnier N., Moizard M.P., Voelckel M.A., Calemard L.M., Boisseau P., Blayau M., Philippe C., Cossee M., Pages M., Rivier F., Danos O., Garcia L., and Claustres M., *Multiexon skipping leading to an artificial DMD protein lacking amino acids from exons 45 through 55 could rescue up to 63% of patients with Duchenne muscular dystrophy*. Hum Mutat, 2007. **28**(2): p. 196-202.
243. Soret J., Gabut M., and Tazi J., *SR proteins as potential targets for therapy*. Prog Mol Subcell Biol, 2006. **44**: p. 65-87.

-
244. Acland G.M., Aguirre G.D., Ray J., Zhang Q., Aleman T.S., Cideciyan A.V., Pearce-Kelling S.E., Anand V., Zeng Y., Maguire A.M., Jacobson S.G., Hauswirth W.W., and Bennett J., *Gene therapy restores vision in a canine model of childhood blindness*. Nat Genet, 2001. **28**(1): p. 92-5.
245. Cideciyan A.V., Hauswirth W.W., Aleman T.S., Kaushal S., Schwartz S.B., Boye S.L., Windsor E.A., Conlon T.J., Sumaroka A., Pang J.J., Roman A.J., Byrne B.J., and Jacobson S.G., *Human RPE65 Gene Therapy for Leber Congenital Amaurosis: Persistence of Early Visual Improvements and Safety at 1 Year*. Hum Gene Ther, 2009.

ANNEXES

ANNEXE A : LE FILTRE *MITOPID1*

Le filtre *mitopid1* est écrit en langage Perl. Il analyse les fichiers résultants de l'alignement des protéines procaryotes contre le génome humain réalisé avec le programme *tblastn* (iBLAST) sur ReMIX. Ces fichiers iBLAST (ReMIX) sont identiques aux fichiers BLAST classiques. Ils contiennent des *hits* contenant eux-mêmes des HSP. Les HSP correspondent à des alignements locaux significatifs entre une séquence requête (*query*, les séquences des protéines procaryotes dans notre cas) et une séquence sujet (*subject*, les séquences nucléiques des chromosomes humains dans notre cas). Pour être significatifs, ces alignements doivent présenter une *E-value* ou un score supérieurs à un seuil qui est défini dans les paramètres du programme BLAST.

L'algorithme de *mitopid1* fusionne tous les HSP proches. Puis, il ne retient que les groupes fusionnés d'HSP qui ont une couverture suffisante et qui débutent et se terminent autour de positions fixées en parties C- et N-terminales. Le fichier de résultat comporte des lignes correspondant à des HSP fusionnés. Il pourrait s'agir d'exons contigus.

Tout d'abord, l'algorithme lit le fichier ReMIX ligne par ligne et ne traite que celles qui ne sont pas des lignes de commentaire (celles ne commençant pas par le caractère « % ») ou qui ne sont pas des lignes vides. Pour chacune de ces lignes, le nom de la requête (*query_name*) est récupéré. Si une ligne a le même nom que la précédente, alors elle est ajoutée à une liste temporaire en attente de traitement. Si une ligne n'a pas le même nom que la précédente, alors la liste en attente qui ne contient que des lignes de même nom (c'est-à-dire un *hit*) est traitée. En d'autres termes, l'algorithme de *mitopid1* traite les HSP du fichier ReMIX par groupes ayant la même protéine requête et la même séquence sujet (appelés *hits* dans un rapport BLAST).

Entrée/sortie

L'option « -i » en ligne de commande indique l'emplacement du fichier d'entrée de type ReMIX. L'en-tête du fichier ReMIX (précédé par le symbole « % ») contient les trois lignes suivantes :

1. le nom du programme lancé, le nom d'utilisateur de la personne qui a lancé l'analyse et la date de l'analyse ;
2. les paramètres passés au programme ;
3. la légende du tableau.

Le corps du fichier ReMIX, contient les résultats de l'alignement sous forme d'un tableau comportant des colonnes suivantes :

- *query_name* : le nom de la requête, c'est-à-dire ses numéros d'accèsion (dans différentes banques de données, par exemple, « gi|16127165|ref|NP_421729.1| ») ;
- *query_description* : la description de la requête (par exemple, «cytochrome c, membrane-bound...») ;
- *query_length* : la taille de la requête en acides aminés ;
- *score* : le score de l'alignement, dépendant de la similarité entre les séquences alignées ;
- *evaluate* : valeur exprimant la significativité de l'alignement (*E-value* ; plus celle-ci est faible, plus l'alignement est fiable ;
- *identity* : pourcentage d'identité entre les séquences alignées ;
- *strand* : la polarité de l'alignement qui peut être positive (alignement sur le brin sens) ou négative (alignement sur le brin antisens) ;
- *query_start* : la position, sur la requête, du début de l'alignement ;
- *query_end* : la position, sur la requête, de la fin de l'alignement ;
- *sjct_start* : la position, sur la séquence sujet, du début de l'alignement ;
- *sjct_end* : la position, sur la séquence sujet, de la fin de l'alignement.

Un extrait de ce type de fichier est présenté Tableau 14.

Tableau 14. Extrait d'un fichier de sortie de ReMIX.

% iBLAST started on ReMIX by georges on Sat Mar 3 00:43:06 CET 2007										
% Parameters: -i=/local/dup/db/gen2-db/hg18/chr19 -d=/local/dup/db/gen2-db/bact/allbact_split_19 -o=/local/dup/apps/gen2-app/RESULTS/R										
% query_name	query_descr	query_length	score	evalue	identity	strand	query_start	query_end	sbjct_start	sbjct_end
gj 15832538 ref N	cell division p	332	99	4.510e-02	31.2500	+	80	183	368715	369048
gj 15832538 ref N	cell division p	332	100	3.453e-02	29.7872	+	86	176	843837	844110
gj 15832538 ref N	cell division p	332	126	3.337e-05	32.7586	-	72	187	62375398	62375099
gj 15832538 ref N	cell division p	332	94	1.714e-01	30.8642	-	87	166	61742139	61741906
gj 15832538 ref N	cell division p	332	101	2.644e-02	29.4118	-	94	185	60580655	60580350
gj 15832538 ref N	cell division p	332	90	4.987e-01	31.4286	-	101	188	60285100	60284792
gj 15832538 ref N	cell division p	332	96	1.005e-01	30.3571	+	73	171	3888668	3888998
gj 15832538 ref N	cell division p	332	94	1.714e-01	31.4050	-	86	183	56884490	56884140
gj 15832538 ref N	cell division p	332	99	4.510e-02	31.1927	-	75	183	47566610	47566332
gj 15832538 ref N	cell division p	332	119	2.163e-04	29.2308	-	62	185	47120096	47119710
gj 15832538 ref N	cell division p	332	112	1.402e-03	33.6842	-	85	192	43918546	43918283
gj 15832538 ref N	cell division p	332	123	7.435e-05	33.0579	-	75	196	43918506	43918165
gj 15832538 ref N	cell division p	332	97	7.694e-02	34.6154	-	74	148	43918362	43918138
gj 15832538 ref N	cell division p	332	110	2.392e-03	38.4615	-	84	161	40482823	40482611
gj 15832538 ref N	cell division p	332	112	1.402e-03	34.6535	-	94	193	34297175	34296915
gj 15832538 ref N	cell division p	332	112	1.402e-03	33.6000	-	69	179	17548897	17548544
gj 15832538 ref N	cell division p	332	94	1.714e-01	36.3636	-	101	165	17442722	17442534
gj 15832538 ref N	cell division p	332	97	7.694e-02	30.1471	+	79	192	47134315	47134702
gj 15832538 ref N	cell division p	332	104	1.187e-02	30.5556	+	86	193	47171957	47172257
gj 15832538 ref N	cell division p	332	104	1.187e-02	37.6344	-	86	172	15237425	15237165
gj 15832538 ref N	cell division p	332	92	2.924e-01	32.2917	+	120	213	48697703	48697961
gj 15832538 ref N	cell division p	332	99	4.510e-02	28.9474	+	85	197	48703255	48703588
gj 15832538 ref N	cell division p	332	99	4.510e-02	41.6667	-	136	192	13965619	13965452
gj 15832538 ref N	cell division p	332	101	2.644e-02	32.7103	-	96	200	13965640	13965347

Le fichier de sortie est soit un fichier dont le nom a été donné en ligne de commande avec l'option « -o », soit un fichier dont le nom sera celui du fichier d'entrée avec l'extension « .mpid1 ». Ce fichier possède le même format que le fichier d'entrée avec certaines lignes fusionnées (voir fonction *merge* ci-après). Différentes options permettent de configurer les paramètres du filtrage. Nous verrons les traitements réalisés par le filtre *mitopid1* et les arguments nécessaires dans la suite de cette annexe.

Traitement des *hits*

Avant d'écrire le résultat du traitement d'un *hit* dans le fichier de sortie, il faut traiter tous les HSP qu'il contient. Pour cela, la fonction *filter* appelle successivement quatre traitements. Le premier traitement est toujours réalisé alors que les trois suivants dépendent des paramètres passés en ligne de commande. La sortie d'un traitement sert d'entrée au suivant. Ces quatre fonctions sont les suivantes :

- *merge* : fusionne les données de plusieurs HSP en une seule ligne du fichier de sortie si leurs positions sont proches ;
- *fCoverage* : ne retient que les HSP fusionnés (ou lignes du fichier de sortie) qui atteignent un pourcentage seuil fixé de couverture de l'alignement ;

- *fNterm* : ne retient que les HSP dont l'extrémité N-terminale de l'alignement se situe à une distance fixée de l'extrémité N-terminale de la protéine requête. Une certaine plage de tolérance est admise.
- *fCterm* : agit comme *fNterm* mais pour les extrémités C-terminales.

Nous allons voir en détail chacune de ces fonctions.

Fusion

La fonction *merge* rassemble sur une même ligne du fichier de sortie *mitopid1* des éléments qui se suivent dans le fichier ReMIX et qui pourraient être sur le génome les exons d'un même gène, entrecoupés par des introns. Également, de par son algorithme, BLAST peut créer des HSP chevauchants dans son fichier de sortie. Il convient donc de s'intéresser à la distance entre deux HSP trouvés, leur superposition éventuelle et leur polarité pour décider de la fusion des lignes. Pour chacun des éléments du tableau à traiter, il faut considérer la polarité, le début et la fin de la requête et du sujet. Chaque HSP est comparé au groupement en cours de fusion. Dans le fichier de sortie, le groupe d'HSP fusionné contient certaines colonnes avec une liste des valeurs des différents HSP non fusionnés, séparées par des virgules. Ainsi, les colonnes 3, 4, 5, 7, 8, 9, 10 qui correspondent respectivement au score, à la *E-value*, à l'identité, au début et la fin de la requête et du sujet contiennent des listes dans le cas de HSP fusionnés (Tableau 15).

Tableau 15. Exemple de fusion de deux HSP en sortie de la fonction *merge*.

% query_name	query_descri	query_len	score	evalue	identity	strand	query_start	query_end	sbjct_start	sbjct_end
gj 15832538 ref NF	cell division	332	99	4.510e-02	31.2500	+	80	183	368715	369048
gj 15832538 ref NF	cell division	332	100	3.453e-02	29.7872	+	86	176	843837	844110

Après fusion des deux HSP précédents par la fonction *merge* :

% query_name	query_descri	query_len	score	evalue	identity	strand	query_start	query_end	sbjct_start	sbjct_end
gj 15832538 ref NF	cell division	332	99,100	4.510e-02,3.	31.2500,29.7872	+	80,86	183,176	368715,843837	369048,844110

Trois tests sont réalisés dans la fonction *merge* pour la fusion des HSP, qui déterminent si un HSP est retenu dans le groupe en cours de fusion ou bien si un nouveau groupe est initié. Ceux-ci sont définis dans les trois sections suivantes.

Test sur la requête

Le test sur la requête est positif si le début de la séquence requête alignée d'un HSP est situé au niveau de la fin du précédent dans le groupement en cours de fusion (Figure 23). Une variable *Ta* (pour « tolérance en acides aminés ») sur la tolérance des positions est

fixée à 5 acides aminés. Au-delà de cette distance, deux HSP ne font pas partie du même groupement.

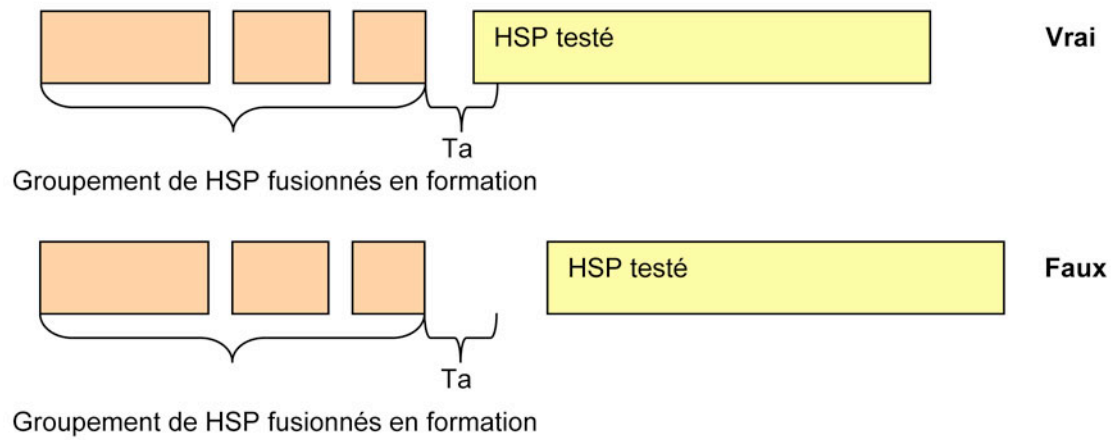


Figure 23. Test de fusion d'un HSP au niveau de sa séquence requête.

Le chevauchement des HSP est aussi testé, en utilisant également la variable *Ta* égale à 5 acides aminés (Figure 24).

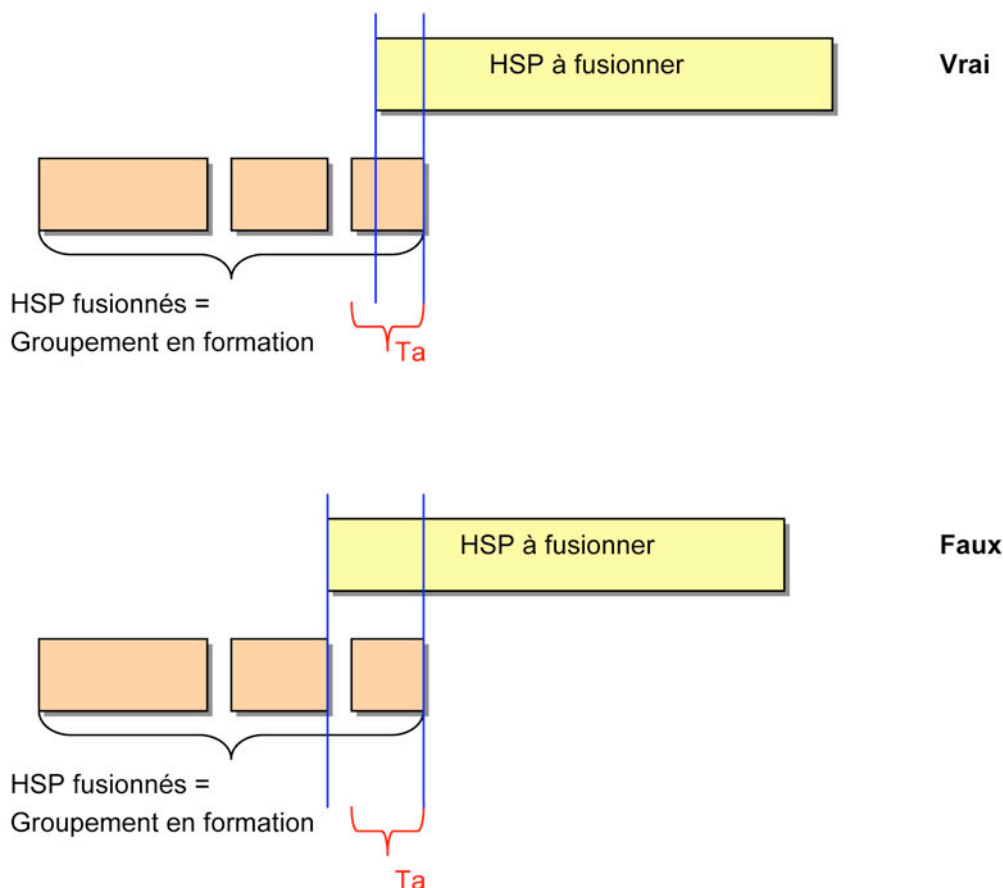


Figure 24. Test de chevauchement d'un HSP au niveau de sa séquence requête.

Test sur le sujet

Les mêmes tests sont effectués pour la séquence sujet. Il y a une seule différence par rapport au test précédent : la variable de tolérance est ajustable. C'est un paramètre passé en ligne de commande avec l'option « -g » (en bases) qui permet de prendre en compte le découpage d'un gène en introns. En moyenne, un intron possède une taille de 3 400 bases, un gène de 27 kb et une séquence codante (*coding sequence*, CDS) de 1 100 bases. De plus, il y a en moyenne 6 ± 3 introns/kb de CDS. Il paraît donc raisonnable de fixer cette limite à 10 000 bases, qui est la valeur par défaut si aucun argument n'est passé en ligne de commande. Lors de notre étude, ce paramètre a été conservé à sa valeur par défaut.

Pour le chevauchement, le fonctionnement est rigoureusement le même que pour la requête avec l'utilisation d'une valeur de tolérance de 5 bases.

Test sur la polarité

Seuls les HSP possédant la même polarité sont ajoutés à un groupement.

Couverture

Pour chaque groupement de HSP, la couverture résultante est calculée par la fonction *fCoverage*. Pour ce faire, la fonction calcule le rapport entre la taille du HSP fusionné (la distance entre la fin du dernier HSP moins le début du premier HSP) et la taille de la requête. La limite de couverture, exprimée en pour cent, est passée en argument (option « -p ») du programme de filtrage. Ainsi, tout HSP (fusionné ou non) est retenu par le filtre *mitopid1* si sa couverture est supérieure à cette limite fixée en ligne de commande. Sinon, il est supprimé. Ce test n'est pas réalisé si aucun paramètre de couverture n'est spécifié.

Extrémité N-terminale

La taille de la zone qui n'est pas similaire (c'est-à-dire non alignée) en début de protéine requête (extrémité N-terminale) est calculée par la fonction *fNterm*. Cette fonction vérifie si l'alignement débute dans une certaine région dont la position dépend de la taille totale de la protéine requête. Cette région doit se situer à une position limite *nterm_limit* plus ou moins une variance *nterm_var* par rapport au début de la séquence protéique requête (Figure 25). Ces limites sont fixées par deux arguments passés en ligne de commande avec respectivement les options « -n » et « -v » en acides aminés. Le test n'est pas réalisé si ces deux paramètres ne sont pas spécifiés.

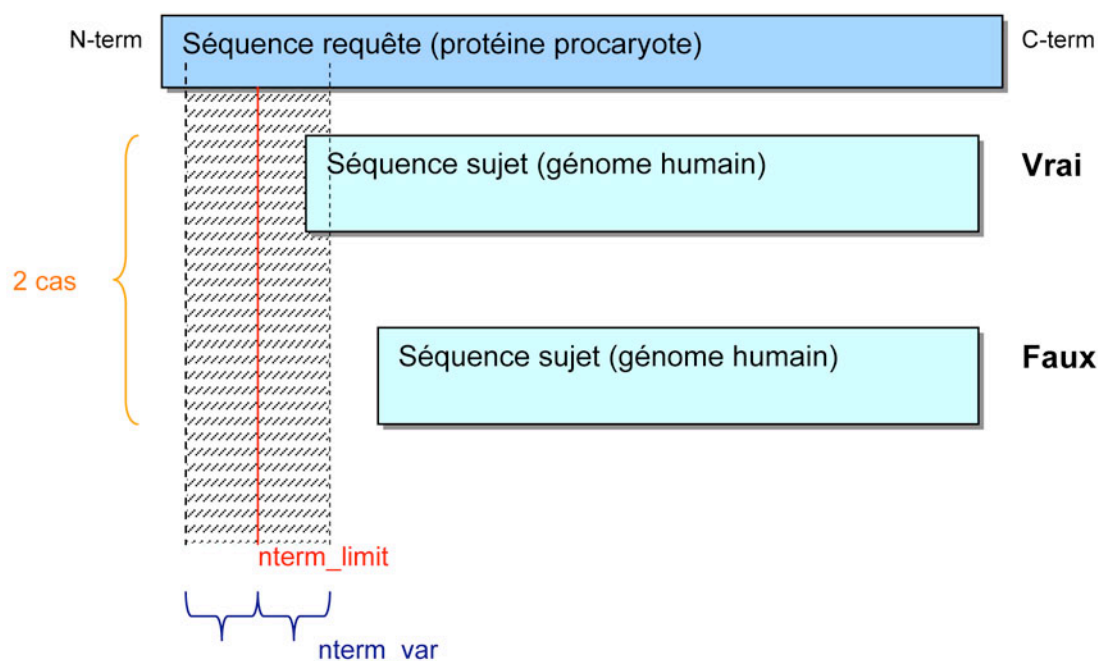


Figure 25. Test sur la partie N-terminale de la requête.

Extrémité C-terminale

La fonction *fCterm* réalise le même test que la fonction *fNterm* appliqué à l'extrémité C-terminale de la séquence protéique requête. Les arguments passés en ligne de commande *cterm_var* par rapport cette position. Les options en ligne de commande sont respectivement « -c » et « -w » (en acides aminés) pour la position C-terminale limite et la variance. Le test n'est pas réalisé si ces deux paramètres ne sont pas fixés.

Code source

```
#!/usr/bin/perl
# -----
# FILE      : mitopid1.pl
#
# PURPOSE   : Identification of mitochondrial proteins (step 1)
#
# VERSION   : 0.5
#
# AUTHOR    : Marc FERRE   <MaFerre@chu-angers.fr>
#
# CREATED   : June, 2005
# REVISION  : September 16, 2005
#            October 13, 2006
#            November 13, 2006
#            December 20, 2006
#            February 15, 2007
#            March 2, 2007
#
# PROTOTYPE : ./mitopid1.pl -p 60 -g 10000 -i file.remix -o file.mpid1
#
# PRECONDITION : Require a tab-delimited file as input, with columns:
#                [0] query name           [3] score           [7] query start
#                [1] query description    [4] expected value  [8] query end
#                [2] query length         [5] identity        [9] subject start
#                [6] strand                [10] subject end
#
# POSTCONDITION : N/A
# -----
# Laboratoire de biochimie et biologie moleculaire, INSERM U694
# CHU d'Angers, 4 rue Larrey, F-49933 ANGERS CEDEX 9, FRANCE
# -----

use strict;
use Getopt::Std;

# >>>Command line options<<<
my %opts = ();
my (
    $remix_file, $mpid1_file, $nterm_lim, $cterm_lim,
    $nterm_var,  $cterm_var,  $cov_lim,   $gap_lim
);
getopts( 'i:o:n:c:v:w:p:g:h', \%opts );
unless ( defined $opts{'h'} ) {
    $remix_file = $opts{'i'} or $remix_file = $ARGV[0];
    $mpid1_file = $opts{'o'}
        or $mpid1_file =
        $remix_file =~ /\.remix$/ ? $` . '.mpid1' : $remix_file . '.mpid1';
    $nterm_lim = $opts{'n'};
    $cterm_lim = $opts{'c'};
    $nterm_var = $opts{'v'};
    $cterm_var = $opts{'w'};
    $cov_lim   = $opts{'p'};
}
```



```

    $gap_lim = $opts{'g'} or $gap_lim = 10000;
}
else {
    print 'Usage : '
        . 'mitopid1.pl'
        . ' [OPTION...]' . "\n\n"
        . 'Options :' . "\n"
        . ' -i Input filename (must be a ReMIX output file)' . "\n"
        . ' -o Output filename (tab delimited-file)' . "\n"
        . ' -n Limit on N-term position of alignment (in a.a.)' . "\n"
        . ' -c Limit on C-term position of alignment (in a.a.)' . "\n"
        . ' -v Variance on N-term limit position (in a.a.)' . "\n"
        . ' -w Variance on C-term limit position (in a.a.)' . "\n"
        . ' -p Limit on percentage of coverage (in percent)' . "\n"
        . ' -g Limit on gap between contiguous HSPs (in b.)' . "\n"
        . ' -h Show help message' . "\n";
    exit;
}

# >>>Initialise Files<<<
open( INPUT, "<$remix_file" ) or die "Can't read file '$remix_file'";
open( OUTPUT, ">$mpid1_file" ) or die "Can't write file '$mpid1_file'";

# Write te output header.
# The output contains general info and the tab-delimited fields:
# [0] query name [3] score [7] query start
# [1] query description [4] expected value [8] query end
# [2] query length [5] identity [9] subject start
# [6] strand [10] subject end
my @header = (
    'query_name', 'query_description',
    'query_length', 'score',
    'evalue', 'identity',
    'strand', 'query_start',
    'query_end', 'sbjct_start',
    'sbjct_end'
);

printf OUTPUT "%% mitopid1.pl started by %s on %s GMT\n", getlogin,
    scalar(gmtime);
print OUTPUT "%% Parameters: -i=$remix_file -o=$mpid1_file";
print OUTPUT " -n=$nterm_lim" if ($nterm_lim);
print OUTPUT " -c=$cterm_lim" if ($cterm_lim);
print OUTPUT " -v=$nterm_var" if ($nterm_var);
print OUTPUT " -w=$cterm_var" if ($cterm_var);
print OUTPUT " -p=$cov_lim" if ($cov_lim);
print OUTPUT " -g=$gap_lim" if ($gap_lim);
print OUTPUT "\n";
printf OUTPUT "%% " . join( "\t", @header ) . "\n";

# >>>Process Filtering<<<
my @grp = ();
my $id_grp = 0;
my $prev_name = '';
while ( my $line = <INPUT> ) {
    unless ( $line =~ m/^[%|\s]/ ) { # Ignore comments and empty lines

```

```

    chomp $line;
    my @col      = split "\t", $line;
    my $cur_name = $col[0];

    # Fix bug
    @col = fixBug(@col);

    # Process a group of lines with the same id
    if ( ( $cur_name ne $prev_name ) and ( $prev_name ne '' ) ) {
        fPrint(
            \@grp,      $nterm_lim, $cterm_lim,
$nterm_var,
                        $cterm_var, $cov_lim,  $gap_lim );
        @grp      = ();
        $id_grp = 0;
    }
    @{ $grp[ $id_grp++ ] } = @col;
    $prev_name = $cur_name;
}
}
fPrint( \@grp,      $nterm_lim, $cterm_lim, $nterm_var,
        $cterm_var, $cov_lim,  $gap_lim
); # Process the last group of lines with the same id

close OUTPUT;
exit;

# -----
# SUBROUTINES
# -----

# Fix bug on sbjct_start and sbjct_end strand
# Swap positions if minus strand
sub fixBug {
    my @col = @_;

    my ( $strand, $start, $end ) = ( $col[6], $col[9], $col[10] );
    ( $col[9], $col[10] ) = ( $end, $start ) if ( $strand eq '-' );

    return @col;
}

# Print a filtered group of line(s)
sub fPrint {
    my (
        $grp_ref,  $nterm_lim, $cterm_lim, $nterm_var,
        $cterm_var, $cov_lim,  $gap_lim
    )
    = @_;

    my @filtered_grp = filter( $grp_ref,  $nterm_lim, $cterm_lim,
$nterm_var,
                                $cterm_var, $cov_lim,  $gap_lim
);
    if (@filtered_grp) { # Print filtered result (may be a group) if it
exists

```

```

        for my $i ( 0 .. $#filtered_grp ) {
            print OUTPUT join( "\t", @{$ filtered_grp[$i] } ), "\n";
        }
    }
}

# Return 2dim-table of HSP(s)
# filtered according to N-term, C-term and gap size
sub filter {
    my (
        $grp_ref,    $nterm_lim, $cterm_lim, $nterm_var,
        $cterm_var, $cov_lim,   $gap_lim
    )
    = @_;

    ###
    # Merge
    my $merged_grp_ref = merge( $grp_ref, $gap_lim );
    ###

    ###
    # Filter
    my $filtered_grp_ref = $merged_grp_ref;
    if ( $cov_lim ) {
        $filtered_grp_ref = fCoverage( $filtered_grp_ref, $cov_lim );
    }
    if ( $nterm_lim and $nterm_var ) {
        $filtered_grp_ref = fNterm( $filtered_grp_ref, $nterm_lim,
$nterm_var );
    }
    if ( $cterm_lim and $cterm_var ) {
        $filtered_grp_ref = fCterm( $filtered_grp_ref, $cterm_lim,
$cterm_var );
    }
    ###

    return @$filtered_grp_ref;
}

# Return reference to 2dim-table of HSP(s)
# merged according to gap limit
sub merge {
    my ( $grp_ref, $gap_lim ) = @_;

    my $ta = 5;          # Tolerance (in a.a.)
    my @grp = @$grp_ref;

    # Merge according to gap limit
    my @merged_grp;
    @{$ merged_grp[0] } = @{$ grp[0] };
    for my $i ( 1 .. $#grp ) {
        my @mrg_query_ends = split( ',', $merged_grp[$#merged_grp][8] );
        my @mrg_subject_ends = split( ',', $merged_grp[$#merged_grp][10] );

        my ( $grp_strand, $grp_query_start, $grp_subject_start ) =

```

```

        ( $grp[$i][6], $grp[$i][7], $grp[$i][9] );
        my ( $mrg_strand, $mrg_query_end, $mrg_subject_end ) = (
$mmerged_grp[$#mmerged_grp][6],
$mrg_query_ends[$#mrg_query_ends],
$mrg_subject_ends[$#mrg_subject_ends]
        );

        my $query_gap_OK = ( $mrg_query_end + $ta ) >= $grp_query_start ? 1
: 0;
        my $query_overlap_OK =
        ( $mrg_query_end - $ta ) <= $grp_query_start ? 1 : 0;
        my $subject_gap_OK =
        ( $mrg_subject_end + $gap_lim ) >= $grp_subject_start ? 1 : 0;
        my $subject_overlap_OK =
        ( $mrg_subject_end - $ta ) <= $grp_subject_start ? 1 : 0;
        my $strand_OK = $mrg_strand eq $grp_strand ? 1 : 0;

        if (
            $query_gap_OK
            and $query_overlap_OK
            and $subject_gap_OK
            and $subject_overlap_OK
            and $strand_OK )
        {
            for my $j ( 3, 4, 5, 7, 8, 9, 10 )
            {
                # List [query|subject]_[start|end], score, evalue,
identity
                    $mmerged_grp[$#mmerged_grp][$j] =
                    $mmerged_grp[$#mmerged_grp][$j] . ',' . $grp[$i][$j];
            }
        }
        else {
            @{$mmerged_grp[ ++$#mmerged_grp ]} = @{$grp[$i]};
        }
    }

    return \@mmerged_grp;
}

# Return reference to 2dim-table of HSP(s)
# filtered according to N-term limit
sub fNterm {
    my ( $grp_ref, $nterm_lim, $nterm_var ) = @_;

    my @grp = @$grp_ref;
    my @filtered_grp;
    for my $i ( 0 .. $#grp ) {
        my ( $strand, $length ) = ( $grp[$i][6], $grp[$i][2] );
        my ( $p, $l, $v );

        if ( $strand eq '+' ) {
            my @query_starts = split( ',', $grp[$i][7] );

```

```

        ( $p, $l, $v ) = ( $query_starts[0], $nterm_lim, $nterm_var
);
    }
    elsif ( $strand eq '-' ) {
        my @query_ends = split( ',', $grp[$i][8] );
        ( $p, $l, $v ) =
            ( $query_ends[$#query_ends], $length - $nterm_lim,
$nterm_var );
    }

    if ( ( $p >= ( $l - $v ) ) and ( $p <= ( $l + $v ) ) ) {
        @{ $filtered_grp[ ++$#filtered_grp ] } = @{ $grp[$i] };
    }
}

return \@filtered_grp;
}

# Return reference to 2dim-table of HSP(s)
# filtered according to C-term limit
sub fCterm {
    my ( $grp_ref, $cterm_lim, $cterm_var ) = @_;

    my @grp = @$grp_ref;
    my @filtered_grp;
    for my $i ( 0 .. $#grp ) {
        my ( $strand, $length ) = ( $grp[$i][6], $grp[$i][2] );
        my ( $p, $l, $v );

        if ( $strand eq '+' ) {
            my @query_ends = split( ',', $grp[$i][8] );
            ( $p, $l, $v ) =
                ( $query_ends[$#query_ends], $length - $cterm_lim,
$cterm_var );
        }
        elsif ( $strand eq '-' ) {
            my @query_starts = split( ',', $grp[$i][7] );
            ( $p, $l, $v ) = ( $query_starts[0], $cterm_lim, $cterm_var
);
        }

        if ( ( $p >= ( $l - $v ) ) and ( $p <= ( $l + $v ) ) ) {
            @{ $filtered_grp[ ++$#filtered_grp ] } = @{ $grp[$i] };
        }
    }

    return \@filtered_grp;
}

# Return reference to 2dim-table of HSP(s)
# filtered according coverage limit
sub fCoverage {
    my ( $grp_ref, $cov_lim ) = @_;

    my @grp = @$grp_ref;

```

```
my @filtered_grp;
for my $i ( 0 .. $#grp ) {
    my $query_length = $grp[$i][2];
    my @query_starts = split( ',', $grp[$i][7] );
    my @query_ends   = split( ',', $grp[$i][8] );
    die "Query starts and ends count doesn't match: ("
        . $grp[$i][7] . ")!=("
        . $grp[$i][8] . ")"
        unless ( scalar(@query_starts) == scalar(@query_ends) );

    my ( $coverage, $limit ) = (
        (
            $query_ends[$#query_ends] - $query_starts[0] ) /
$query_length
        ) * 100,
        $cov_lim
    );
    if ( $coverage >= $limit ) {
        @{ $filtered_grp[ ++$#filtered_grp ] } = @{ $grp[$i] };
    }
}

return \@filtered_grp;
}
```


ANNEXE B : LE FILTRE *MITOPID2*

Le filtre *mitopid2* est écrit en langage Perl. Il est conçu pour traiter les données de sortie de *mitopid1*. Il recherche les informations contenues dans la base de données *UCSC Genome Browser Database* pour chaque zone chromosomique (groupe de HSP fusionnés) retenue par *mitopid1*. Plusieurs tables de la base de l'UCSC, réparties en trois catégories dans *mitopid2*, sont interrogées : la catégorie *known* pour la table *knownGene* ; *pred* pour les tables *geneid*, *genscan*, *mgcGenes*, *refGene*, *sgpGene* ; *mrna* pour les tables *chr[1–22, X, Y]_mrna*.

Entrée/sortie

L'option « -i » en ligne de commande indique l'emplacement du fichier d'entrée de type *mitopid1*. Le fichier de sortie est soit un fichier dont le nom a été donné en ligne de commande avec l'option « -o », soit un fichier dont le nom sera celui du fichier d'entrée avec l'extension « .mpid2 ». Ce fichier possède le même format que le fichier d'entrée avec l'ajout de colonnes pour renseigner les données de la base de l'UCSC si la requête a correspondu à une entrée dans celle-ci. Différentes options permettent de configurer les paramètres du filtrage.

Recherche dans la table des gènes connus

Il est possible de restreindre les requêtes de *mitopid2* à une seule catégorie de tables avec l'option « -c » en ligne de commande. Pour notre étude, nous faisons exclusivement des requêtes dans la table des gènes connus (*knownGene*) de la base de l'UCSC. Chaque requête effectuée par *mitopid2* dans la table se décompose en deux étapes. D'abord une requête récupérant toutes les entrées selon la position génomique du CDS, en prenant en compte une tolérance *Tb* (Figure 26). Ensuite, pour chaque résultat positif sur le CDS, le découpage des HSP fusionnés par *mitopid1* est comparé aux exons de l'entrée dans la base de l'UCSC, avec une tolérance *Tb* (Figure 27). Si un HSP fusionné correspond avec au moins un exon d'un gène de la base de l'UCSC, les informations suivantes concernant le gène sont ajoutées à celles de la ligne du fichier d'entrée (*mitopid1*) :

- *section* : la catégorie de la table (*known* dans notre cas) ;
- *table* : le nom de la table (*knownGene* dans notre cas) ;
- *name* : le nom du gène ;
- *proteinID* : l'identifiant Swiss-Prot de la protéine correspondante ;
- *alignID* : un identifiant unique pour chaque couple gène connu/position de l'alignement ;
- *strand* : la polarité (« + » ou « - ») ;

- *cdsStart* : le début de la région codante ;
- *cdsEnd* : la fin de la région codante ;
- *exonCount* : le nombre d'exons ;
- *exonStarts* : la liste des positions du début des exons ;
- *exonEnds* : la liste des positions de la fin des exons.

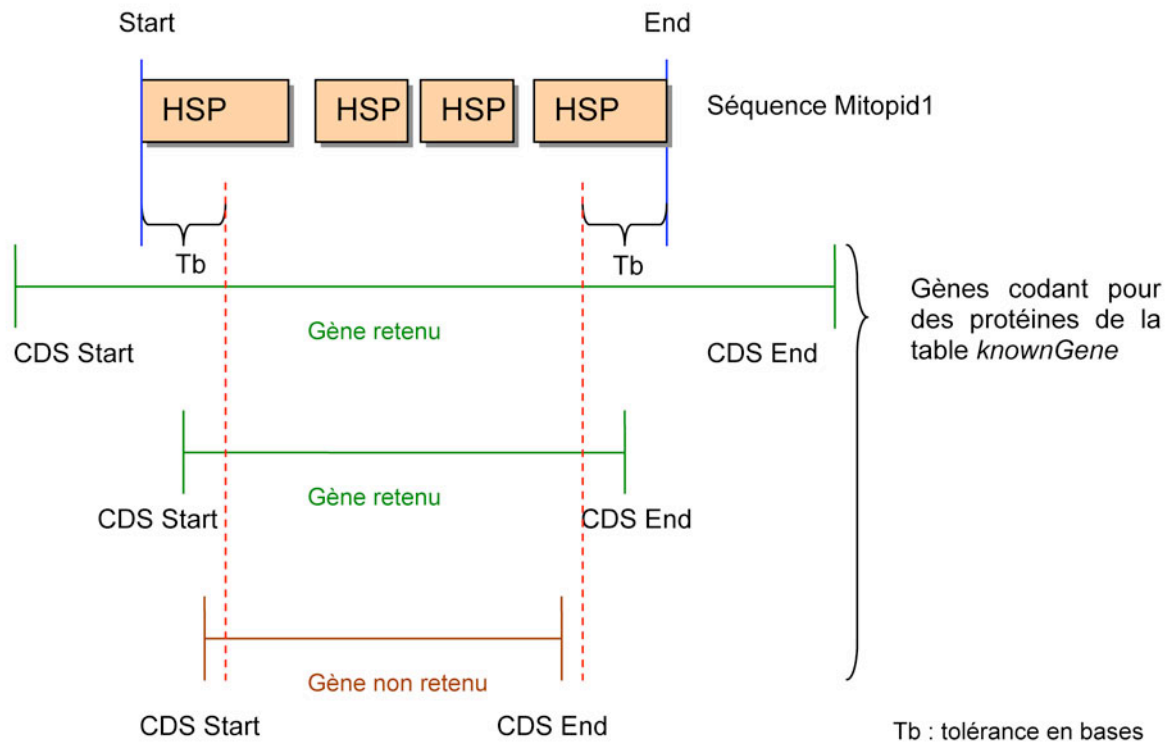


Figure 26. Recherche dans la table des gènes connus selon la position du CDS par *mitopid2*.

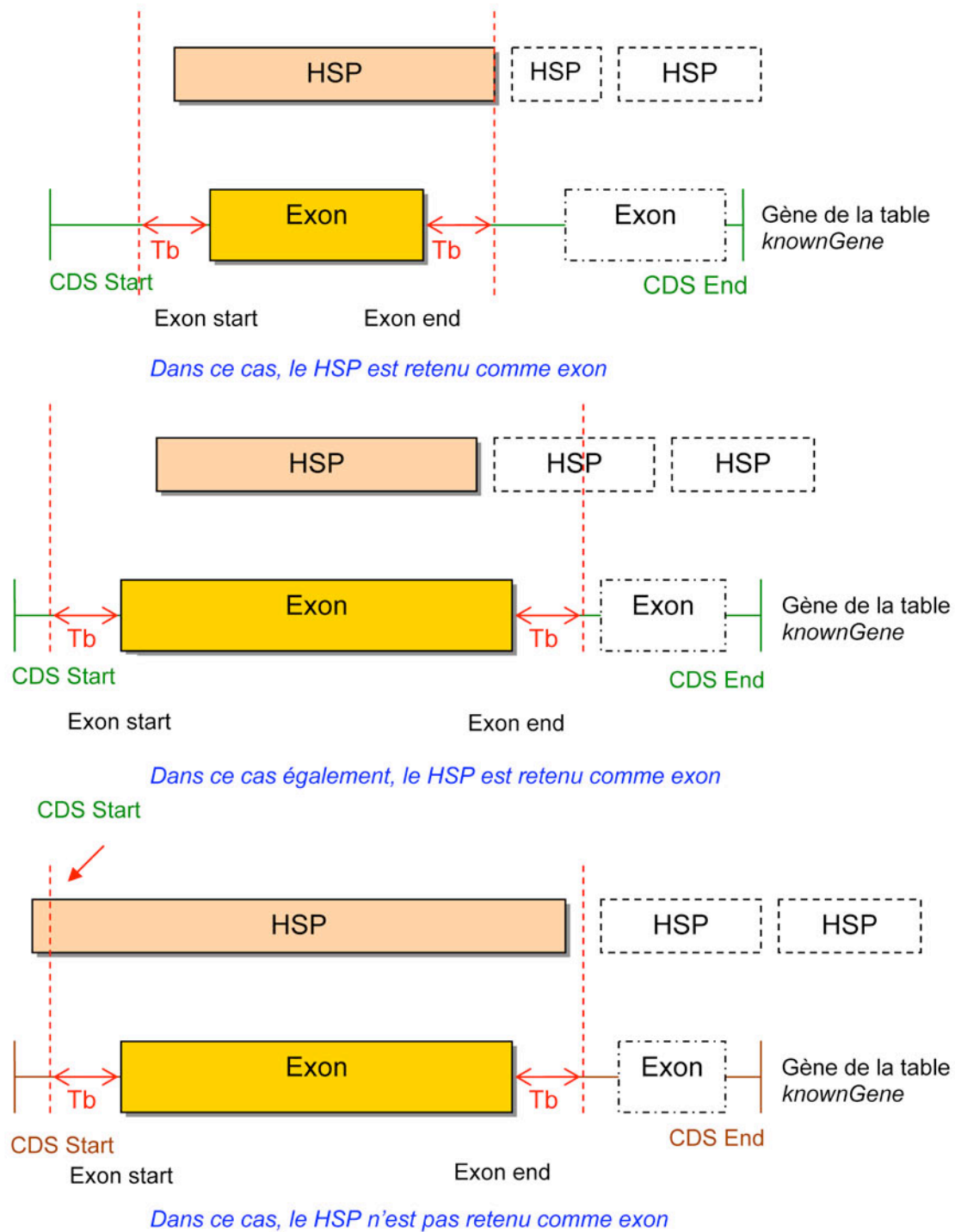


Figure 27. Analyse de la correspondance des HSP avec les exons des gènes de la base de l'UCSC.

Code source

```
#!/usr/bin/perl
# -----
# FILE      : mitopid2.pl
#
# PURPOSE   : Identification of mitochondrial proteins (step 2)
#
# VERSION   : 0.4
#
# AUTHOR    : Marc FERRE   <MaFerre@chu-angers.fr>
#
# CREATED   : June, 2005
# REVISION  : September 16, 2005
#            November 24, 2006
#            December 28, 2006
#            February 26, 2007
#
# PROTOTYPE : ./mitopid2.pl -i file.mpid1 -o file.mpid2
#
# PRECONDITION : Require a tab-delimited file as input, with columns:
#                [0] query name          [3] score          [7] query start
#                [1] query description  [4] expected value [8] query end
#                [2] query length       [5] identity       [9] subject start
#                [6] strand              [10] subject end
#
#                Input filename must be set as following:
#                <chromosome_name>-<free_text>.mpid1
#                UCSC genome annotation database hg18 must be installed
#                and mitopid user (password: mitopid) acces allowed
#
# POSTCONDITION : N/A
# -----
# Laboratoire de biochimie et biologie moleculaire, INSERM U694
# CHU d'Angers, 4 rue Larrey, F-49033 ANGERS CEDEX 9, FRANCE
# -----

use strict;
use Getopt::Std;
use DBI();

# >>>Command line options<<<
my %opts = ();
my ( $mpid1_file, $mpid2_file, $class, $db_name, $db_user, $db_pass, $db_host
);
getopts( 'i:o:c:d:s:u:p:h', \%opts );
unless ( defined $opts{'h'} ) {
    $mpid1_file = $opts{'i'} or $mpid1_file = $ARGV[0];
    $mpid2_file = $opts{'o'}
    or $mpid2_file =
        $mpid1_file =~ /remix\mpid1$/
        ? $` . 'mpid2'
        : $mpid1_file . '.mpid2';
    $class = $opts{'c'} or $class = 'all';
    $db_name = $opts{'d'} or $db_name = 'hg18';
    $db_host = $opts{'s'} or $db_host = 'localhost';
}
```

```

$db_user = $opts{'u'} or $db_user = 'mitopid';
$db_pass = $opts{'p'} or $db_pass = 'mitopid';
}
else {
    print 'Usage   : '
      . 'mitopid2.pl'
      . ' [OPTION...]' . "\n\n"
      . 'Options :' . "\n"
      . ' -i   Input filename (must be a ReMIX or MPID1 output file)' . "\n"
      . ' -o   Output filename (tab delimited-file)' . "\n"
      . ' -c   Class of results ([known|pred|mrnal|all])' . "\n"
      . ' -d   Human genome annotation database name (MySQL)' . "\n"
      . ' -s   Human genome annotation database host' . "\n"
      . ' -u   MySQL user name' . "\n"
      . ' -p   MySQL user password' . "\n"
      . ' -h   Show help message' . "\n";
    exit;
}

# >>>Initialise working set<<<
open( INPUT, "<$mpid1_file" ) or die "Can't read file '$mpid1_file'";
open( OUTPUT, ">$mpid2_file" ) or die "Can't write file '$mpid2_file'";

my $dbh = DBI->connect( "DBI:mysql:database=$db_name;host=$db_host",
                      $db_user, $db_pass, { 'RaiseError' => 1 }
);

$mpid1_file =~ /(chr\w+)/ # Get chromosome name
or die "Can't catch chromosome name from filename '$mpid1_file'";
my $chrom = $1;

# Write te output header
# The input contains general info and the tab-delimited fields:
# [0] query name      [8] query end      [15] gp alignID
# [1] query description [9] subject start [16] gp strand
# [2] query length    [10] subject end   [17] gp cdsStart
# [3] score           [11] gp section   [18] gp cdsEnd
# [4] expected value  [12] gp table     [19] gp exonCount
# [5] identity        [13] gp name      [20] gp exonStarts
# [6] strand          [14] gp proteinID [21] gp exonEnds
# [7] query start
my @header = (
    'query_name',    'query_description',
    'query_length', 'score',
    'evalue',        'identity',
    'strand',        'query_start',
    'query_end',     'sbjct_start',
    'sbjct_end',     'gp_section',
    'gp_table',      'gp_name',
    'gp_proteinID', 'gp_alignID',
    'gp_strand',     'gp_cdsStart',
    'gp_cdsEnd',     'gp_exonCount',
    'gp_exonStarts', 'gp_exonEnds'
);

```

```

printf OUTPUT "%% mitopid2.pl started by %s on %s GMT\n", getlogin,
    scalar(gmtime);
printf OUTPUT "%% Parameters: -i=%s -o=%s -c=%s -d=%s -s=%s -u=%s -p=%s\n",
    $mpid1_file, $mpid2_file, $class, $db_name, $db_host, $db_user, $db_pass;
print OUTPUT '% ', join( "\t", @header ), "\n";

# >>>Process search<<<
while ( my $line = <INPUT> ) {
    unless ( $line =~ m/^[%\s]/ ) { # Ignore comments and empty lines
        chomp $line;
        my @col = split( "\t", $line );

        my $strand = $col[6];
        my @starts = split( ',', $col[9] );
        my @ends = split( ',', $col[10] );
        die "Subject starts and ends count doesn't match: ("
            . $col[9] . ")!=("
            . $col[10] . ")"
            unless ( scalar(@starts) == scalar(@ends) );
        my $tb = 30; # Tolerance (in b.)

        my $found = 0;

        # -> Search known genes
        $found = 1
            if printSearch( \@col,
                searchKnown( $class, $strand, \@starts,
                    \@ends, $tb ) );

        # -> Search mRNA
        $found = 1
            if printSearch( \@col,
                searchMrna( $class, $strand, \@starts,
                    \@ends, $tb ) );

        # -> Search predicted genes
        $found = 1
            if printSearch( \@col,
                searchPred( $class, $strand, \@starts,
                    \@ends, $tb ) );

        # -> If no match found
        printEmpty( \@col ) unless ( $found );
    }
}
$dbh->disconnect();
close OUTPUT;
exit;

# -----
# SUBROUTINES
# -----

# Return reference to table of known genes (if applicable)
# *****

```

```

# Retrieve data from the table knownGene (GoldenPath)
# table knownGene
# "Protein coding genes based on proteins from SWISS-PROT, TrEMBL, and
# TrEMBL-NEW and their corresponding mRNAs from GenBank"
# (
#   string name;           "Name of gene"
#   string chrom;         "Chromosome name"
#   char[1] strand;       "+ or - for strand"
#   uint txStart;         "Transcription start position"
#   uint txEnd;           "Transcription end position"
#   uint cdsStart;        "Coding region start"
#   uint cdsEnd;          "Coding region end"
#   uint exonCount;       "Number of exons"
#   uint[exonCount] exonStarts; "Exon start positions"
#   uint[exonCount] exonEnds;  "Exon end positions"
#   string proteinID;     "SWISS-PROT ID"
#   string alignID;       "Unique identifier for each (known
#                           gene, alignment position) pair"
# )
# *****
sub searchKnown {
    my ( $class, $strand, $starts_ref, $ends_ref, $tb ) = @_;

    my $section = 'known';
    if ( ( $class eq $section ) or ( $class eq 'all' ) ) {
        my @result;
        my @starts = @$starts_ref;
        my @ends   = @$ends_ref;
        my $table  = 'knownGene';
        my $query  =
            "SELECT * FROM $table WHERE (chrom LIKE '$chrom')
            . " AND (strand LIKE '$strand')
            . " AND ( cdsStart <= "
            . ( $starts[0] + $tb ) . " )"
            . " AND ( cdsEnd >= "
            . ( $ends[$#ends] - $tb ) . " )";
        my $sth = $dbh->prepare($query);
        $sth->execute();

        while ( my $ref = $sth->fetchrow_hashref() ) {
            my @exonStarts = split( ',', $ref->{'exonStarts'} );
            my @exonEnds   = split( ',', $ref->{'exonEnds'} );
            my $found_all  = 1;

            for my $i ( 0 .. $#starts ) { # Browse HSP starts
                my $found_one = 0;
                for ( my $j = 0 ; $j < $ref->{'exonCount'} ; $j++ )
                { # Browse database exons
                    $found_one++
                    if ( ( $starts[$i] >= ( $exonStarts[$j] -
$tb ) )
                        and ( $ends[$i] <= ( $exonEnds[$j] +
$tb ) ) );
                }
                $found_all &&= $found_one;
            }
        }
    }
}

```

```

    }
    @{ $result[ ++$#result ] } = (
        $section,
    $table,
        $ref->{'proteinID'},
        $ref->{'alignID'},
        $ref->{'strand'},
        $ref->{'cdsStart'},
        $ref->{'cdsEnd'},
        $ref->{'exonCount'},
        $ref->{'exonStarts'},
        $ref->{'exonEnds'}
    )
    if ($found_all);
}
$sth->finish();

return \@result;
}
}

# Return reference to table of predicted genes (if applicable)
# *****
# Retrieve data from gene prediction tables (GoldenPath)
# "The following definition is used for gene prediction tables.
# In alternative-splicing situations, each transcript has a row
# in this table.
# (
#     string name;           "Name of gene"
#     string chrom;         "Chromosome name"
#     char[1] strand;       "+ or - for strand"
#     uint txStart;         "Transcription start position"
#     uint txEnd;           "Transcription end position"
#     uint cdsStart;        "Coding region start"
#     uint cdsEnd;          "Coding region end"
#     uint exonCount;       "Number of exons"
#     uint[exonCount] exonStarts; "Exon start positions"
#     uint[exonCount] exonEnds;  "Exon end positions"
# )
# *****
sub searchPred {
    my ( $class, $strand, $starts_ref, $ends_ref, $tb ) = @_;

    my $section = 'pred';
    if ( ( $class eq $section ) or ( $class eq 'all' ) ) {
        my @result;
        my @starts = @$starts_ref;
        my @ends = @$ends_ref;
        my @tables = ( 'geneid', 'genscan', 'mgcGenes', 'refGene',
'sgpGene' );

        foreach my $table (@tables) {
            my $query =
                "SELECT * FROM $table WHERE (chrom LIKE '$chrom')"
```

```

        . " AND (strand LIKE '$strand')"
        . " AND ( cdsStart <= "
        . ( $starts[0] + $tb ) . " )"
        . " AND ( cdsEnd >= "
        . ( $ends[$#ends] - $tb ) . " )";
my $sth = $dbh->prepare($query);
$sth->execute();

while ( my $ref = $sth->fetchrow_hashref() ) {
    my @exonStarts = split( ',', $ref->{'exonStarts'} );
    my @exonEnds   = split( ',', $ref->{'exonEnds'} );
    my $found_all  = 1;

    for my $i ( 0 .. $#starts ) {    # Browse HSP starts
        my $found_one = 0;
        for ( my $j = 0 ; $j < $ref->{'exonCount'} ;
$j++ )
            {
                # Browse database
                exons
                    $found_one++
                    if ( ( $starts[$i] >= (
$exonStarts[$j] - $tb ) )
                        and ( $ends[$i] <= (
$exonEnds[$j] + $tb ) ) );
            }
            $found_all &&= $found_one;
        }
        @{ $result[ ++$#result ] } = (
$section,
$ref->{'name'},
'///',
$ref->{'strand'},
$ref->{'cdsStart'},
$ref->{'cdsEnd'},
$ref->{'exonCount'},
$ref->{'exonStarts'},
$ref->{'exonEnds'}
        )
        if ($found_all);
    }
    $sth->finish();
}

return \@result;
}
}

# Return reference to table of mRNA (if applicable)
# *****
# Retrieve data from mRNA tables (GoldenPath)
# "Summary info about a patSpace alignment"
# (
#     ushort bin;           "A field to speed indexing"

```



```

#     uint     matches;           "Number of bases that match
#                                     that aren't repeats"
#     uint     misMatches;       "Number of bases that don't match"
#     uint     repMatches;       "Number of bases that match but are
#                                     part of repeats"
#     uint     nCount;           "Number of 'N' bases"
#     uint     qNumInsert;       "Number of inserts in query"
#     int      qBaseInsert;      "Number of bases inserted in query"
#     uint     tNumInsert;       "Number of inserts in target"
#     int      tBaseInsert;      "Number of bases inserted in target"
#     char[2]  strand;           "+ or - for query strand. For
#                                     translated alignments, second +/-
#                                     is for genomic strand"
#
#     string   qName;            "Query sequence name"
#     uint     qSize;            "Query sequence size"
#     uint     qStart;           "Alignment start position in query"
#     uint     qEnd;             "Alignment end position in query"
#     string   tName;            "Target sequence name"
#     uint     tSize;            "Target sequence size"
#     uint     tStart;           "Alignment start position in target"
#     uint     tEnd;             "Alignment end position in target"
#     uint     blockCount;       "Number of blocks in alignment
#                                     (a block contains no gaps)"
#     uint[blockCount] blockSizes; "Size of each block"
#     uint[blockCount] qStarts;   "Start of each block in query"
#     uint[blockCount] tStarts;   "Start of each block in target"
# )
# *****
sub searchMrna {
    my ( $class, $strand, $starts_ref, $ends_ref, $tb ) = @_;

    my $section = 'mrna';
    if ( ( $class eq $section ) or ( $class eq 'all' ) ) {
        my @result;
        my @starts = @$starts_ref;
        my @ends   = @$ends_ref;
        my $table  = $chrom . '_mrna';
        my $query  =
            "SELECT * FROM $table WHERE (strand LIKE '$strand')
            . " AND ( tStart <= "
            . ( $starts[0] + $tb ) . " )"
            . " AND ( tEnd >= "
            . ( $ends[$#ends] - $tb ) . " )";
        my $sth = $dbh->prepare($query);
        $sth->execute();

        while ( my $ref = $sth->fetchrow_hashref() ) {
            my @tStarts = split( ',', $ref->{'tStarts'} );
            my @blockSizes = split( ',', $ref->{'blockSizes'} );
            my @tEnds;
            my $found_all = 1;
            for ( my $i = 0 ; $i < $ref->{'blockCount'} ; $i++ ) {
                push( @tEnds, $tStarts[$i] + $blockSizes[$i] );
            }
            $ref->{'tEnds'} = join( ',', @tEnds );
        }
    }
}

```

```

        for my $i ( 0 .. $#starts ) {    # Browse HSP starts
            my $found_one = 0;
            for ( my $j = 0 ; $j < $ref->{'blockCount'} ; $j++ )
            {
                # Browse database exons
                $found_one++
                if (      ( $starts[$i] >= ( $tStarts[$j] - $tb
) )
                    and ( $ends[$i] <= ( $tEnds[$j] + $tb )
) );
            }
            $found_all &&= $found_one;
        }
        @{$result[ ++$#result ] } = (
            $table,
            $ref->{'qName'},
            '//',
            $ref->{'strand'},
            $ref->{'tEnd'},
            $ref->{'tStart'},
            $ref->{'blockCount'}, $ref->{'tStarts'},
            $ref->{'tEnds'}
        )
        if ($found_all);
    }
    $sth->finish();

    return \@result;
}

# Print a search result
sub printSearch {
    my ( $col_ref, $res_ref ) = @_;

    if ($res_ref) {
        my @res = @{$res_ref};
        if ( scalar(@res) ) {
            my @col = @{$col_ref};
            for my $i ( 0 .. $#res ) {
                print OUTPUT join( "\t", @col ), "\t",
                    join( "\t", @{$res[$i]} ), "\n";
            }
            return 1;
        }
    }
    else {
        return 0;
    }
}

# Print an empty result

```

```
sub printEmpty {
  my ($col_ref) = @_;

  my @result = @{$col_ref};
  push( @result,
    ( '///', '///', '///', '///', '///', '///', '///', '///', '///', '///',
    '///' )
  );

  print OUTPUT join( "\t", @result ), "\n";
}
```

ANNEXE C : LE FILTRE *MITOPID3*

Le filtre *mitopid3* est écrit en langage Perl. Il est conçu pour traiter les données de sortie de *mitopid2*. Son algorithme lit chaque ligne du fichier *mitopid2* et ne traite que les lignes qui contiennent un résultat trouvé dans les tables de la base de l'UCSC. Ce filtre va effectuer trois tests^a qui permettront de sélectionner les gènes candidats prédits par notre méthode : 1) sur la couverture de l'alignement ; 2) sur la région non alignée N-terminale ; et 3) sur la région non alignée C-terminale. Il s'agit du dernier filtre du flux de traitement des alignements ReMIX ; il retient les gènes mitochondriaux PH+ candidats selon l'hypothèse que nous avons formulée (voir Étude complémentaire non publiée : méthode de prédiction à haut débit de nouvelles protéines mitochondriales par génomique comparative, p. 83).

Entrée/sortie

L'option « -i » en ligne de commande indique l'emplacement du fichier d'entrée de type *mitopid2*. Le fichier de sortie est soit un fichier dont le nom a été donné en ligne de commande avec l'option « -o », soit un fichier dont le nom sera celui du fichier d'entrée avec l'extension « .mpid3 ».

Couverture

La fonction *fCoverage*^b calcule la couverture, en pour cent, de la protéine procaryote alignée par rapport aux exons correspondants dans la base de données de l'UCSC. L'option « -p » en ligne de commande définit la limite de couverture. Ici, seule la taille de la protéine alignée par rapport au CDS est testée, car la couverture des exons est vérifiée en amont par *mitopid2*. La Figure 28 résume les différents cas envisageables.

Extrémités N- et C-terminales

La fonction *fLim* est utilisée à la fois pour les extrémités N- et C-terminales. Elle vérifie la présence d'une zone non alignée pour chacune des extrémités de la protéine codée par le gène identifié sur le génome humain. Dans le cas de l'hypothèse que nous avons formulée, nous fixons une limite en partie N-terminale de la protéine, ce qui correspondrait à

^a **NB** : Les trois tests réalisés par *mitopid3* peuvent prêter à confusion avec ceux de *mitopid1*. Mais, il ne s'agit pas des mêmes tests : *mitopid1* effectue un filtrage sur les alignements « protéine procaryote vs génome humain » alors que *mitopid3* traite les alignements « protéine procaryote vs gène humain correspondant à l'alignement ». C'est donc à ce stade du flux de traitement que l'hypothèse d'identification des gènes mitochondriaux PH+ candidats est vérifiée.

^b La fonction *fCoverage* du filtre *mitopid3* est différente de celle de *mitopid1*. Il ne faut donc pas la confondre avec cette dernière dont elle est une adaptation.

l'extension, supposée d'origine eucaryote, présente dans la protéine codée par le génome nucléaire par rapport à sa protéine procaryote correspondante.

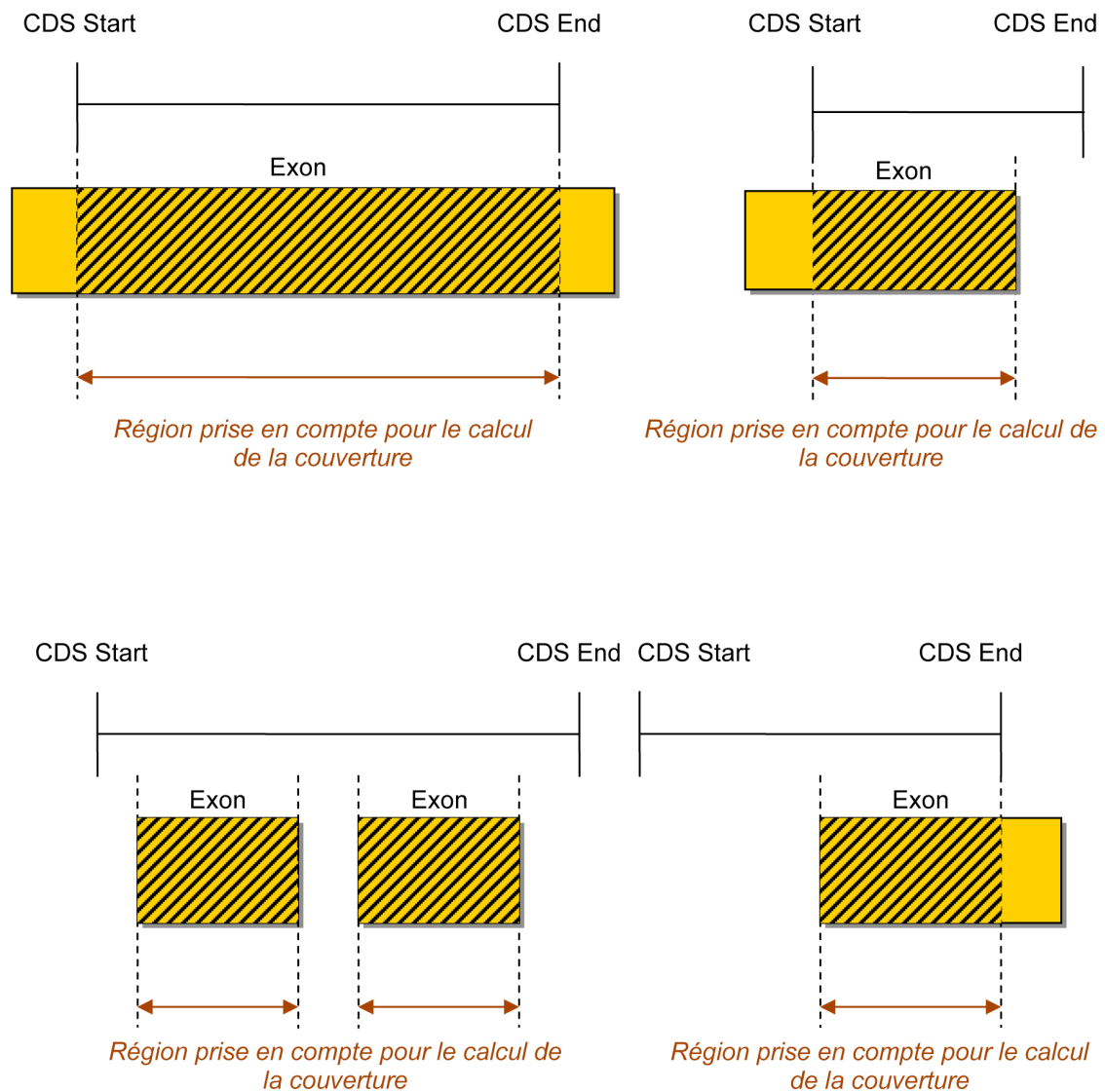


Figure 28. Test de la couverture des groupes de HSP fusionnés avec les exons des gènes correspondants.

Code source

```
#!/usr/bin/perl
# -----
# FILE      : mitopid3.pl
#
# PURPOSE   : Identification of mitochondrial proteins (step 3)
#
# VERSION   : 0.2
#
# AUTHOR    : Marc FERRE <MaFerre@chu-angers.fr>
#
# CREATED   : December, 2006
# REVISION  : March 5, 2007
#           : June 19, 2007
#
# PROTOTYPE : ./mitopid3.pl -p 60 -i file.mpid2 -o file.mpid3
#
# PRECONDITION : Require a tab-delimited file as input, with columns:
#               [0] query name      [8] query end      [15] gp alignID
#               [1] query description [9] subject start [16] gp strand
#               [2] query length     [10] subject end [17] gp cdsStart
#               [3] score            [11] gp section  [18] gp cdsEnd
#               [4] expected value   [12] gp table   [19] gp exonCount
#               [5] identity         [13] gp name    [20] gp exonStarts
#               [6] strand           [14] gp proteinID [21]gp exonEnds
#               [7] query start
#
# POSTCONDITION : N/A
# -----
# Laboratoire de biochimie et biologie moleculaire, INSERM U694
# CHU d'Angers, 4 rue Larrey, F-49033 ANGERS CEDEX 9, FRANCE
# -----

use strict;
use Getopt::Std;

# >>>Command line options<<<
my %opts = ();
my (
    $mpid2_file, $mpid3_file, $nterm_lim, $cterm_lim,
    $nterm_var,  $cterm_var,  $cov_lim
);
getopts( 'i:o:n:c:v:w:p:h', \%opts );
unless ( defined $opts{'h'} ) {
    $mpid2_file = $opts{'i'} or $mpid2_file = $ARGV[0];
    $mpid3_file = $opts{'o'}
        or $mpid3_file =
            $mpid2_file =~ /\.mpid2$/
            ? $` . '.mpid3'
            : $mpid2_file . '.mpid3';
    $nterm_lim = $opts{'n'};
    $cterm_lim = $opts{'c'};
    $nterm_var = $opts{'v'};
    $cterm_var = $opts{'w'};
}
```

```

    $cov_lim = $opts{'p'};
}
else {
    print 'Usage : '
        . 'mitopid3.pl'
        . ' [OPTION...]' . "\n\n"
        . 'Options :' . "\n"
        . ' -i Input filename (must be a MPID2 output file)' . "\n"
        . ' -o Output filename (tab delimited-file)' . "\n"
        . ' -n Limit on N-term position of alignment (in b.)' . "\n"
        . ' -c Limit on C-term position of alignment (in b.)' . "\n"
        . ' -v Variance on N-term limit position (in b.)' . "\n"
        . ' -w Variance on C-term limit position (in b.)' . "\n"
        . ' -p Limit on percentage of coverage (in percent)' . "\n"
        . ' -h Show help message' . "\n";
    exit;
}

# >>>Initialise files<<<
open( INPUT, "<$mpid2_file" ) or die "Can't read file '$mpid2_file'";
open( OUTPUT, ">$mpid3_file" ) or die "Can't write file '$mpid3_file'";

# Write te output header
# The input contains general info and the tab-delimited fields:
# [0] query name      [8] query end      [15] gp alignID
# [1] query description [9] subject start [16] gp strand
# [2] query length    [10] subject end   [17] gp cdsStart
# [3] score           [11] gp section   [18] gp cdsEnd
# [4] expected value  [12] gp table     [19] gp exonCount
# [5] identity        [13] gp name      [20] gp exonStarts
# [6] strand          [14] gp proteinID [21]gp exonEnds
# [7] query start
my @header = (
    'query_name',    'query_description',
    'query_length', 'score',
    'evalue',        'identity',
    'strand',        'query_start',
    'query_end',    'sbjct_start',
    'sbjct_end',    'gp_section',
    'gp_table',     'gp_name',
    'gp_proteinID', 'gp_alignID',
    'gp_strand',    'gp_cdsStart',
    'gp_cdsEnd',    'gp_exonCount',
    'gp_exonStarts', 'gp_exonEnds'
);

printf OUTPUT "%% mitopid3.pl started by %s on %s GMT\n", getlogin,
    scalar(gmtime);
print OUTPUT "%% Parameters: -i=$mpid2_file -o=$mpid3_file";
print OUTPUT " -n=$nterm_lim" if ($nterm_lim);
print OUTPUT " -c=$cterm_lim" if ($cterm_lim);
print OUTPUT " -v=$nterm_var" if ($nterm_var);
print OUTPUT " -w=$cterm_var" if ($cterm_var);
print OUTPUT " -p=$cov_lim" if ($cov_lim);
print OUTPUT "\n";
printf OUTPUT "%% " . join( "\t", @header ) . "\n";

```

```

# >>>Process Filtering<<<
while ( my $line = <INPUT> ) {
    unless ( $line =~ m/^[%|\s]/ ) { # Ignore comments and empty lines
        chomp $line;
        my @col      = split "\t", $line;
        my $gp_section = $col[11];

        unless ( $gp_section eq '/' ) {

            # Subject data
            my ( $sj_strand, $sj_starts, $sj_ends ) =
                ( $col[6], $col[9], $col[10] );
            my @sj_starts = split( ',', $sj_starts );
            my @sj_ends   = split( ',', $sj_ends );
            die "Subject starts and ends count doesn't match: ("
                . $sj_starts . ")!=("
                . $sj_ends . ")"
                unless ( scalar(@sj_starts) == scalar(@sj_ends) );

            # GoldenPath data
            my (
                $gp_strand,    $gp_cdsStart, $gp_cdsEnd,
                $gp_exonStarts, $gp_exonEnds
            )
            = ( $col[16], $col[17], $col[18], $col[20], $col[21] );
            my @gp_exonStarts = split( ',', $gp_exonStarts );
            my @gp_exonEnds   = split( ',', $gp_exonEnds );
            die "Subject starts and ends count doesn't match: ("
                . $gp_exonStarts . ")!=("
                . $gp_exonEnds . ")"
                unless ( scalar(@gp_exonStarts) == scalar(@gp_exonEnds) );

            # Filtering
            print( OUTPUT $line, "\n" )
                if filter(
                    $cov_lim,    $nterm_lim,    $nterm_var,
                    $cterm_lim, $cterm_var,    \@sj_starts,
                    \@sj_ends,  $gp_strand,
                    $gp_cdsStart,
                    $gp_cdsEnd, \@gp_exonStarts,
                    \@gp_exonEnds
                );
        }
    }
}

close OUTPUT;
exit;

# -----
# SUBROUTINES
# -----

# Return tables of CDS starts and ends

```



```

sub cds {
  my ( $cdsStart, $cdsEnd, $exonStarts_ref, $exonEnds_ref ) = @_;

  my @exonStarts = @$exonStarts_ref;
  my @exonEnds   = @$exonEnds_ref;
  my ( @cdsStarts, @cdsEnds );
  for my $i ( 0 .. $#exonStarts ) {
    if ( ( $cdsStart >= $exonStarts[$i] ) and ( $cdsEnd <=
$exonEnds[$i] ) )
    {
      $cdsStarts[0] = $cdsStart;
      $cdsEnds[0]   = $cdsEnd;
    }
    elsif ( ( $cdsStart >= $exonStarts[$i] )
and ( $cdsStart <= $exonEnds[$i] ) )
    {
      $cdsStarts[0] = $cdsStart;
      $cdsEnds[0]   = $exonEnds[$i];
    }
    elsif ( ( $cdsStart < $exonStarts[$i] )
and ( $cdsEnd > $exonEnds[$i] ) )
    {
      $cdsStarts[ ++$#cdsStarts ] = $exonStarts[$i];
      $cdsEnds[ ++$#cdsEnds ]     = $exonEnds[$i];
    }
    elsif ( ( $cdsEnd >= $exonStarts[$i] )
and ( $cdsEnd <= $exonEnds[$i] ) )
    {
      $cdsStarts[ ++$#cdsStarts ] = $exonStarts[$i];
      $cdsEnds[ ++$#cdsEnds ]     = $cdsEnd;
    }
  }
  return ( \@cdsStarts, \@cdsEnds );
}

# Return the length of the non-align starting/ending segments of the HSP
sub selength {
  my ( $sj_start, $sj_end, $gp_starts_ref, $gp_ends_ref ) = @_;

  my @gp_starts = @$gp_starts_ref;
  my @gp_ends   = @$gp_ends_ref;

  my ( $sLength, $eLength ) = ( 0, 0 );
  for my $i ( 0 .. $#gp_starts ) {

    # Start length
    if ( $sj_start > $gp_ends[$i] ) {
      $sLength += ( $gp_ends[$i] - $gp_starts[$i] );
    }
    elsif ( ( $sj_start >= $gp_starts[$i] )
and ( $sj_start <= $gp_ends[$i] ) )
    {
      $sLength += ( $sj_start - $gp_starts[$i] );
    }
  }
}

```

```

# End length
if ( ( $sj_end >= $gp_starts[$i] )
    and ( $sj_end <= $gp_ends[$i] ) )
{
    $eLength += ( $gp_ends[$i] - $sj_end );
}
elseif ( $sj_end < $gp_starts[$i] ) {
    $eLength += ( $gp_ends[$i] - $gp_starts[$i] );
}
}

return ( $sLength, $eLength );
}

# Return TRUE/FALSE according to coverage limit
sub fCoverage {
    my ( $limit, $sj_starts_ref, $sj_ends_ref, $gp_starts_ref, $gp_ends_ref )
    =
        @_;

    my @sj_starts = @$sj_starts_ref;
    my @sj_ends   = @$sj_ends_ref;
    my @gp_starts = @$gp_starts_ref;
    my @gp_ends   = @$gp_ends_ref;

    # Alignment lengths
    sub alnLength {
        my @starts = shift;
        my @ends   = shift;

        my $length = 0;
        for my $i ( 0 .. $#starts ) {
            $length += ( $ends[$i] - $starts[$i] );
        }
        return $length;
    }
    my ( $sj_length, $gp_length ) =
        ( alnLength( @sj_starts, @sj_ends ), alnLength( @gp_starts, @gp_ends )
    );

    # Coverage
    my $coverage = ( ( $sj_length / $gp_length ) * 100 );

    return ( $coverage >= $limit ) ? 1 : 0;
}

# Return TRUE/FALSE according to position limit
sub fLim {
    my ( $lim, $var, $strand, $sLength, $eLength ) = @_;

    my $length = ( $strand eq '+' ) ? $sLength : $eLength;

    return (
        ( $length >= ( $lim - $var ) )
        and ( $length <= ( $lim + $var ) ) );
}

```

```

# Return TRUE if all test are successfully passed
# else FALSE
sub filter {
  my (
    $cov_lim,      $nterm_lim,      $nterm_var,
    $cterm_lim,   $cterm_var,      $sj_starts_ref,
    $sj_ends_ref, $gp_strand,      $gp_cdsStart,
    $gp_cdsEnd,   $gp_exonStarts_ref, $gp_exonEnds_ref
  )
  = @_;

  my ( $gp_starts_ref, $gp_ends_ref ) =
    cds( $gp_cdsStart, $gp_cdsEnd, $gp_exonStarts_ref, $gp_exonEnds_ref );
  my ( $sj_start, $sj_end ) =
    ( ${ @{$sj_starts_ref} }[0], ${ @{$sj_ends_ref} }[ $#{$sj_ends_ref} ]
  );

  my ( $sLength, $eLength ) =
    selength( $sj_start, $sj_end, $gp_starts_ref, $gp_ends_ref );

  my ( $cov_ok, $nterm_ok, $cterm_ok );
  if ( $cov_lim ) {
    $cov_ok = fCoverage(      $cov_lim,      $sj_starts_ref,
$sj_ends_ref,
                                $gp_starts_ref, $gp_ends_ref );
  }
  else {
    $cov_ok = 1;
  }
  if ( $nterm_lim and $nterm_var ) {
    $nterm_ok =
      fLim( $cterm_lim, $cterm_var, $gp_strand, $sLength, $eLength );
  }
  else {
    $nterm_ok = 1;
  }
  if ( $cterm_lim and $cterm_var ) {
    $cterm_ok =
      fLim( $cterm_lim, $cterm_var, $gp_strand, $sLength, $eLength );
  }
  else {
    $cterm_ok = 1;
  }

  return ( $cov_ok and $nterm_ok and $cterm_ok ) ? 1 : 0;
}

```

ANNEXE D : LE FILTRE *MITOPID11*

Le flux de traitement des alignements ReMIX est composé de trois filtres *mitopid1-3*. Pour les besoins de notre étude visant à fixer certains paramètres de ces filtres, un quatrième filtre *mitopid11* a été développé. Il est écrit en langage Perl et contient le filtrage selon la couverture du *mitopid1*. Il reprend l'algorithme de ce dernier en l'adaptant pour son fichier d'entrée qui diffère légèrement.

Entrée/sortie

L'option « -i » en ligne de commande indique l'emplacement du fichier d'entrée de type *mitopid2*. Le fichier de sortie est soit un fichier dont le nom a été donné en ligne de commande avec l'option « -o », soit un fichier dont le nom sera celui du fichier d'entrée avec l'extension « .mpid11 ».

Code source

```
#!/usr/bin/perl
# -----
# FILE      : mitopid11.pl
#
# PURPOSE   : Identification of mitochondrial proteins
#             MPID1 like filter for post-MIPD2 filtering (bench only)
#
# VERSION   : 0.1
#
# AUTHOR    : Marc FERRE <MaFerre@chu-angers.fr>
#
# CREATED   : June, 2007
# REVISION  : NA
#
# PROTOTYPE : ./mitopid11.pl -p 15 -g 10000 -i file.mpid2 -o file.mpid11
#
# PRECONDITION : Bug on ReMIX output file must be fixed (swap of positions
#                                     if minus strand)
#
#             Require a tab-delimited file as input, with columns:
#             [0] query name      [8] query end      [15] gp alignID
#             [1] query description [9] subject start [16] gp strand
#             [2] query length     [10] subject end [17] gp cdsStart
#             [3] score            [11] gp section  [18] gp cdsEnd
#             [4] expected value   [12] gp table   [19] gp exonCount
#             [5] identity         [13] gp name    [20] gp exonStarts
#             [6] strand           [14] gp proteinID [21] gp exonEnds
#             [7] query start
#
# POSTCONDITION : N/A
# -----
# Laboratoire de biochimie et biologie moleculaire, INSERM U694
# CHU d'Angers, 4 rue Larrey, F-49933 ANGERS CEDEX 9, FRANCE
# -----

use strict;
use Getopt::Std;

# >>>Command line options<<<
my %opts = ();
my (
    $mpid2_file, $mpid11_file, $nterm_lim, $cterm_lim,
    $nterm_var,  $cterm_var,  $cov_lim,  $gap_lim
);
getopts( 'i:o:n:c:v:w:p:h', \%opts );
unless ( defined $opts{'h'} ) {
    $mpid2_file = $opts{'i'} or $mpid2_file = $ARGV[0];
    $mpid11_file = $opts{'o'}
    or $mpid11_file =
        $mpid2_file =~ /\.mpid2$/
        ? $` . '.mpid11'
        : $mpid2_file . '.mpid11';
    $nterm_lim = $opts{'n'};
}
```

```

    $cterm_lim = $opts{'c'};
    $nterm_var = $opts{'v'};
    $cterm_var = $opts{'w'};
    $cov_lim   = $opts{'p'};
}
else {
    print 'Usage   : '
        . 'mitopid1.pl'
        . ' [OPTION...]' . "\n\n"
        . 'Options :' . "\n"
        . ' -i   Input filename (must be a ReMIX output file)' . "\n"
        . ' -o   Output filename (tab delimited-file)' . "\n"
        . ' -n   Limit on N-term position of alignment (in a.a.)' . "\n"
        . ' -c   Limit on C-term position of alignment (in a.a.)' . "\n"
        . ' -v   Variance on N-term limit position      (in a.a.)' . "\n"
        . ' -w   Variance on C-term limit position      (in a.a.)' . "\n"
        . ' -p   Limit on percentage of coverage      (in percent)' . "\n"
        . ' -h   Show help message' . "\n";
    exit;
}

# >>>Initialise Files<<<
open( INPUT, "<$mpid2_file" ) or die "Can't read file '$mpid2_file'";
open( OUTPUT, ">$mpid11_file" ) or die "Can't write file '$mpid11_file'";

# Write te output header.
# The input contains general info and the tab-delimited fields:
# [0] query name      [8] query end      [15] gp alignID
# [1] query description [9] subject start [16] gp strand
# [2] query length    [10] subject end   [17] gp cdsStart
# [3] score           [11] gp section   [18] gp cdsEnd
# [4] expected value  [12] gp table     [19] gp exonCount
# [5] identity        [13] gp name      [20] gp exonStarts
# [6] strand          [14] gp proteinID [21]gp exonEnds
# [7] query start
my @header = (
    'query_name',    'query_description',
    'query_length', 'score',
    'evalue',        'identity',
    'strand',        'query_start',
    'query_end',    'sbjct_start',
    'sbjct_end',    'gp_section',
    'gp_table',     'gp_name',
    'gp_proteinID', 'gp_alignID',
    'gp_strand',    'gp_cdsStart',
    'gp_cdsEnd',    'gp_exonCount',
    'gp_exonStarts', 'gp_exonEnds'
);

);
printf OUTPUT "%% mitopid11.pl started by %s on %s GMT\n", getlogin,
    scalar(gmtime);
print OUTPUT "%% Parameters: -i=$mpid2_file -o=$mpid11_file";
print OUTPUT " -n=$nterm_lim" if ($nterm_lim);
print OUTPUT " -c=$cterm_lim" if ($cterm_lim);
print OUTPUT " -v=$nterm_var" if ($nterm_var);
print OUTPUT " -w=$cterm_var" if ($cterm_var);

```

```

print OUTPUT " -p=$cov_lim"  if ($cov_lim);
print OUTPUT "\n";
printf OUTPUT "%% " . join( "\t", @header ) . "\n";

# >>>Process Filtering<<<
while ( my $line = <INPUT> ) {
    unless ( $line =~ m/^[%\s]/ ) {    # Ignore comments and empty lines
        chomp $line;
        my @col = split "\t", $line;

        # Query data
        my ( $length, $strand, $query_starts, $query_ends ) =
            ( $col[2], $col[6], $col[7], $col[8] );
        my @query_starts = split( ',', $query_starts );
        my @query_ends   = split( ',', $query_ends );
        die "Subject starts and ends count doesn't match: ("
            . $query_starts . ")!=( "
            . $query_ends . ")"
            unless ( scalar(@query_starts) == scalar(@query_ends) );

        # Filtering
        print( OUTPUT $line, "\n" )
            if filter(
                $cov_lim,  $nterm_lim,
                $nterm_var, $cterm_lim,
                $cterm_var, $strand,
                $length,  $query_starts[0],
                $query_ends[$#query_ends]
            );
    }
}

close OUTPUT;
exit;

# -----
# SUBROUTINES
# -----

# Return TRUE/FALSE according to coverage limit
sub fCoverage {
    my ( $cov_lim, $length, $start, $end ) = @_;

    my $coverage = ( ( $end - $start ) / $length ) * 100;

    return ( $coverage >= $cov_lim ) ? 1 : 0;
}

# Return TRUE/FALSE according to N-term limit
sub fNterm {
    my ( $nterm_lim, $nterm_var, $strand, $length, $start, $end ) = @_;

    my ( $p, $l, $v );

    if ( $strand eq '+' ) {

```

```

        ( $p, $l, $v ) = ( $start, $nterm_lim, $nterm_var );
    }
    elsif ( $strand eq '-' ) {
        ( $p, $l, $v ) = ( $end, $length - $nterm_lim, $nterm_var );
    }

    return ( ( $p >= ( $l - $v ) ) and ( $p <= ( $l + $v ) ) ) ? 1 : 0;
}

# Return TRUE/FALSE according to C-term limit
sub fCterm {
    my ( $cterm_lim, $cterm_var, $strand, $length, $start, $end ) = @_;

    my ( $p, $l, $v );

    if ( $strand eq '+' ) {
        ( $p, $l, $v ) = ( $end, $length - $cterm_lim, $cterm_var );
    }
    elsif ( $strand eq '-' ) {
        ( $p, $l, $v ) = ( $start, $cterm_lim, $cterm_var );
    }

    return ( ( $p >= ( $l - $v ) ) and ( $p <= ( $l + $v ) ) ) ? 1 : 0;
}

# Return TRUE if all test are successfully passed
# else FALSE
sub filter {
    my (
        $cov_lim, $nterm_lim, $nterm_var, $cterm_lim, $cterm_var,
        $strand, $length, $start, $end,
    )
    = @_;

    my ( $cov_ok, $nterm_ok, $cterm_ok );
    if ( $cov_lim ) {
        $cov_ok = fCoverage( $cov_lim, $length, $start, $end );
    }
    else {
        $cov_ok = 1;
    }
    if ( $nterm_lim and $nterm_var ) {
        $nterm_ok =
            fNterm( $nterm_lim, $nterm_var, $strand, $length, $start, $end );
    }
    else {
        $nterm_ok = 1;
    }
    if ( $cterm_lim and $cterm_var ) {
        $cterm_ok =
            fCterm( $cterm_lim, $cterm_var, $strand, $length, $start, $end );
    }
    else {
        $cterm_ok = 1;
    }
}

```



```
} return ( $cov_ok and $nterm_ok and $cterm_ok ) ? 1 : 0;
```

INDEX

- 1D SDS-PAGE. *Voir* Électrophorèse sur gel SDS-polyacrylamide à une dimension
- 2D PAGE. *Voir* Électrophorèse bidimensionnelle en gel de polyacrylamide
- A Reconfigurable Memory for Indexing Mass of Data*. *Voir* ReMIX
- Adénosine 5'-diphosphate, 1, 22, 33, 39, 141
- Adénosine 5'-triphosphate, 1, 21–22, 33, 39, 41, 52–54, 60–62, 141, 164
- Amplification en chaîne par polymérase, 13, 31
- ATPase 6, 4
- ATPase 8, 4
- Atrophie optique autosomique dominante, 43–67, 103–40
- Atrophie optique autosomique dominante et cataracte, 45
- Atrophie optique autosomique dominante et surdit , 50, 53–54, 60
- Atrophie optique autosomique r cessive, 44
- Atrophie optique li e   l'X, 44
- Base de donn es sp cifique d'un g ne, 104–105, 139, 148–150
- BLAST (*Basic Local Alignment Search Tool*), 84–86, 169, 173, 176
- Brin l ger, 3
- Brin lourd, 2
- Cadre de lecture ouvert, 31, 55
- Cellule ganglionnaire de la r tine, 43, 45–47, 50–51, 55, 60–61, 65–66
- Centres fer-soufre, 21–22
- Chromatographie en phase liquide coupl e   une spectrom trie de masse   transform e de Fourier, 27, 30, 31
- Chromatographie liquide multidimensionnelle, 24
- Couche des fibres nerveuses r tiniennes, 43
- CSB (*Conserved sequence box*), 3, 5
- Cytochrome *b*, 3–4, 153
- Cytochrome *c*, 3–4, 22–23, 54, 58–59, 65, 153, 166–168, 174
- Cytochrome *c* oxydase, 4, 54
- DDP1 (Deafness dystonia peptide 1)*, 41, 160
- D gradation des ARNm, 60
- D riv s r actifs de l'oxyg ne, 8, 19, 41, 61, 66, 139
- Domaine GED, 55–56
-  lectrophor se bidimensionnelle en gel de polyacrylamide, 23–24, 27, 30, 31
-  lectrophor se sur gel SDS-polyacrylamide   une dimension, 24, 27, 29
- electrospray ionization mass spectrometry*. *Voir* Spectrom trie de masse   ionisation par  lectrospray
- eOPA1, 81
-  pilepsie myoclonique avec fibres rouges d chiquet es, 40
- ESI-MS. *Voir* Spectrom trie de masse   ionisation par  lectrospray
- Espace intermembranaire, 57–58

ETAS (*Extended termination associated sequences*), 3, 5

Fe/S. *Voir* Centres fer-soufre

Fibres rouges déchiquetées, 50–51, 54

FISH. *Voir* Hybridation *in situ* fluorescente

Forme oxydée du nicotinamide adénine dinucléotide, 1

Forme réduite du nicotinamide adénine dinucléotide, 1, 3–4, 45, 157

GED. *Voir* Domaine GED

GFP. *Voir* Protéine fluorescente verte

Green fluorescent protein. *Voir* Protéine fluorescente verte

GTPase effector domain. *Voir* Domaine GED

Haplogroupes, 11–20, 130–38

Hybridation *in situ* fluorescente, 60

Imagerie par résonance magnétique, 40

IRM. *Voir* Imagerie par résonance magnétique

LC-FTICR-MS. *Voir* Chromatographie en phase liquide couplée à une spectrométrie de masse à transformée de Fourier

Liquid chromatography fourier transformation cyclotron resonance mass spectrometry. *Voir* Chromatographie en phase liquide couplée à une spectrométrie de masse à transformée de Fourier

Liquide céphalorachidien, 50–51

Locus Specific Databases. *Voir* Base de données spécifique d'un gène

LSDB. *Voir* Base de données spécifique d'un gène

MDLC. *Voir* Chromatographie liquide multidimensionnelle

MELAS. *Voir* Myopathie mitochondriale, encéphalopathie, acidose lactique et tableaux neurologiques aigus ressemblant à des accidents ischémiques cérébraux

Membrane externe mitochondriale, 1, 22, 33, 58–59, 148

Membrane interne mitochondriale, 1, 22, 33, 39, 55, 58–60, 83, 148, 218

MERRF. *Voir* Épilepsie myoclonique avec fibres rouges déchiquetées

Microscopie électronique en transmission, 65

MLPA (*Multiplex ligation probe amplification*), 57

Multidimensional liquid chromatography. *Voir* Chromatographie liquide multidimensionnelle

Myopathie mitochondriale, encéphalopathie, acidose lactique et tableaux neurologiques aigus ressemblant à des accidents ischémiques cérébraux, 40

NAD⁺. *Voir* Forme oxydée du nicotinamide adénine dinucléotide

NADH. *Voir* Forme réduite du nicotinamide adénine dinucléotide

NADH déshydrogénase, 3–4, 157

Nano chromatographie liquide couplée à une spectrométrie de masse en tandem, 24, 27

Nanoscale liquid chromatography tandem mass spectrometry. *Voir* Nano chromatographie liquide couplée à une spectrométrie de masse en tandem

- Neuropathie optique héréditaire de Leber, 18, 40, 44–45, 48, 50–53, 60, 62, 105, 139
- n-LC-MS/MS. *Voir* Nano chromatographie liquide couplée à une spectrométrie de masse en tandem
- NMD. *Voir* Dégradation des ARNm
- Nonsense-mediated mRNA decay*. *Voir* Dégradation des ARNm
- Ophtalmoplégie chronique progressive externe, 50–51, 54, 165
- Phosphorylation oxydative, 50–51, 62–63
- Polymorphisme de taille des fragments de restriction, 11, 16, 139
- Pression intracrânienne, 61
- Pression intraoculaire, 61
- Promoteur du brin léger, 3, 5
- Promoteur du brin lourd, 3, 5
- Protéine fluorescente verte, 28, 31, 34
- Ragged Red Fibers*. *Voir* Fibres rouges déchiquetées
- Reactive oxygen species*. *Voir* Dérivés réactifs de l'oxygène
- Récessive liée à l'X, 45
- ReMIX, 83–86, 88, 89, 95, 173–176, 182, 193, 201, 209–211
- Restriction fragment length polymorphisms*. *Voir* Polymorphisme de taille des fragments de restriction
- RFLP. *Voir* Polymorphisme de taille des fragments de restriction
- ROS. *Voir* Dérivés réactifs de l'oxygène
- RRF. *Voir* Fibres rouges déchiquetées
- Sclérose en plaques, 53
- Spectrométrie de masse à ionisation par électrospray, 24
- SQL (*Simple Query Language*), 91, 92
- Syndrome de Kearns-Sayre, 18
- Système nerveux central, 50
- Tomographie de cohérence optique, 47
- Ubiquinone, 1

RÉSUMÉ

Les mitochondries sont impliquées dans de nombreux processus cellulaires essentiels tels que le catabolisme des nutriments, la phosphorylation oxydative, l'apoptose et la régulation des flux calciques. Elles ont une structure dynamique, qui s'adapte en permanence aux besoins cellulaires, et sont sous le contrôle de réseaux de régulation coordonnant leur masse, leur structure et leurs fonctions. Le protéome mitochondrial est ainsi composé d'une très grande diversité de protéines qui dérive en partie de l'ancêtre procaryote des mitochondries et résulte majoritairement d'une information codée par le génome nucléaire, mais aussi d'une information plus restreinte portée par le génome mitochondrial. Sept cents protéines mitochondriales ont été caractérisées chez l'homme, grâce à diverses approches de protéomique, de génomique et de bio-informatique, mais il est probable que les mitochondries en comportent un nombre bien supérieur. La caractérisation de ces protéines est essentielle à la compréhension des nombreuses maladies génétiques et communes associées à des dysfonctionnements mitochondriaux. Au cours de ce travail de thèse nous avons comparé *in silico* les séquences des protéines mitochondriales humaines avec celles des procaryotes, mettant en évidence que les protéines impliquées dans les pathologies sont majoritairement homologues à des protéines procaryotes. Nous avons ensuite développé une stratégie de recherche bio-informatique de nouvelles protéines mitochondriales basée sur leur origine procaryote et la présence d'une extension N-terminale caractéristique. L'ensemble des protéomes procaryotes connus a été comparé au génome humain et différents outils de filtrage ont été développés pour identifier de nouvelles protéines. Parallèlement à cette stratégie globale de criblage, nous nous sommes focalisés sur l'étude d'une des protéines participant à la fusion mitochondriale qui est associée à l'atrophie optique autosomique dominante, la protéine Opa1. Cette dynamine GTPase est impliquée dans le remodelage de la membrane interne mitochondriale, l'apoptose, la maintenance de l'ADN mitochondrial et le métabolisme énergétique. Nous avons développé une base de données internationale répertoriant les différents variants affectant Opa1 afin de caractériser son spectre mutationnel. Cet outil a secondairement servi à une étude clinique multicentrique portant sur près de mille patients porteurs d'une neuropathie optique. À travers ces deux approches, nous avons pu développer de nouveaux outils bio-informatiques qui devraient contribuer à une meilleure compréhension de la physiopathologie mitochondriale.

ABSTRACT

Bioinformatics analysis of the mitochondrial proteome and the mutational spectrum of the Opa1 protein. Mitochondria are involved in several essential cellular processes including nutrient catabolism, oxidative phosphorylation, apoptosis and regulation of the calcium flux. The dynamic structure of mitochondria is constantly adapted to cellular needs by regulatory networks that determine the mass, structure and function of the organelles. The mitochondrial proteome is composed of a wide variety of proteins inherited in part from the prokaryotic ancestors of mitochondria and mainly encoded by the nuclear genome, as well as a smaller number of proteins encoded by the mitochondrial genome. So far, about 700 mitochondrial proteins have been characterized in humans by means of proteomics, genomics and bioinformatics. However, the actual number of mitochondrial proteins may be much higher. The characterization of these proteins is essential to the understanding of many common genetic diseases associated with mitochondrial dysfunction. We first compared *in silico* sequences of human mitochondrial proteins with those of prokaryotes and showed that most of the proteins involved in such diseases are homologous to prokaryotic proteins. We then developed a bioinformatics research strategy to identify new mitochondrial proteins on the basis of their prokaryotic origin and the presence of a characteristic N-terminal extension. All known prokaryotic proteomes were compared to the human genome and various filtering tools were developed to identify new proteins. In parallel to this overall strategy of screening, we focused on the study of the Opa1 protein, one of the proteins associated with autosomal dominant optic atrophy, which is involved in mitochondrial fusion. Opa1, a dynamin GTPase, is involved in the remodeling of the mitochondrial inner membrane, apoptosis, maintenance of mitochondrial DNA, and energy metabolism. We finally developed an international database listing the variations of Opa1 so as to characterize its mutational spectrum. This tool was used as a complement to a multicentric clinical study involving nearly a thousand patients with optic neuropathy. Our work has led to the development of novel bioinformatics tools that should contribute to a better understanding of mitochondrial pathophysiology.

Mots-clés : mitochondries ; maladies mitochondriales ; protéome ; évolution ; atrophie optique autosomique dominante ; AOAD ; optic atrophy 1 ; OPA1 ; base de données

Keywords: mitochondria; mitochondrial diseases; protéome; évolution; autosomal dominant optic atrophy; ADOA; optic atrophy 1; OPA1; database