

Université de Sherbrooke

**La coopération fonctionnelle entre deux protéines codées par l'ARNm du gène
*SMCR7L/MiD51***

Par
Annie Roy
Programmes de biochimie

Mémoire présenté à la Faculté de médecine et des sciences de la santé
en vue de l'obtention du grade de maître ès sciences (M. Sc.)
en biochimie

Sherbrooke, Québec, Canada
Janvier, 2017

Membres du jury d'évaluation
Pr. Xavier Roucou, Biochimie
Pr. Guylain Boissonneult, Biochimie
Pr. François-Michel Boisvert, Anatomie et biologie cellulaire

©Annie Roy, 2017

RÉSUMÉ

La coopération fonctionnelle entre deux protéines codées par l'ARNm du gène *SMCR7L/MiD51*

Par
Annie Roy
Programmes de biochimie

Mémoire présenté à la Faculté de médecine et des sciences de la santé en vue de l'obtention du diplôme de maître ès sciences (M. Sc.) en biochimie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

Il a été établi que l'ARN messenger (ARNm) eucaryote ne contient qu'un cadre de lecture ouvert (ORF, *open reading frame* en anglais) unique appelé CDS (pour *coding sequence* en anglais) entouré de deux régions non codantes (UTRs, *untranslated regions* en anglais). Pourtant, il est possible de retrouver dans l'ARNm d'autres ORFs appelés cadres de lecture ouverts alternatifs (altORFs). Ces altORFs peuvent être traduits en protéines fonctionnelles et contribuer à l'enrichissement du protéome.

Pour démontrer le potentiel fonctionnel des altORFs dans le génome de plusieurs espèces, nous avons effectué des analyses bio-informatiques. Nos analyses mettent en lumière le potentiel multicodeur des ARNm pour des petites protéines pouvant coopérer avec la protéine de référence dans une même voie biologique. Pour illustrer la coopération fonctionnelle entre une protéine alternative et de référence, je me suis intéressée à un altORF et un refORF retrouvés dans le gène *SMCR7L*. La protéine de référence MiD51 est connue pour son implication dans la fission mitochondriale. J'ai démontré que la protéine alternative altMiD51 est elle aussi impliquée dans la fission mitochondriale, car sa surexpression induit la fragmentation des mitochondries.

En conclusion, mes travaux de maîtrise ont permis de consolider le concept de l'existence des altORFs codant pour des protéines fonctionnelles par la caractérisation d'un altORF codant pour altMiD51. J'ai aussi apporté un exemple concret de coopération fonctionnelle dans le dynamisme mitochondrial entre une protéine de référence et une protéine alternative codées à partir du même ARNm. Ces résultats remettent en question le potentiel unicodeur de l'ARNm. L'ARNm eucaryote aurait plutôt un potentiel multicodeur pour des protéines coopérant de près ou de loin à une même finalité biologique.

Mots-clés: cadres ouverts de lecture alternatif, MiD51, AltMiD51, fission mitochondriale.

TABLE DES MATIÈRES

1	Introduction	1
1.1	Un ARN messager mature et une protéine	1
1.1.1	Les ARNs non codants	2
1.2	L'annotation des ORFs du génome humain	3
1.2.1	Les petits ORFs humains sous la loupe	4
1.3	Les cadres de lecture ouverts alternatifs (AltORFs)	4
1.3.1	Les altORFs ^{5'}	6
1.3.2	Les altORFs ^{CDS}	9
1.3.3	Les altORFs ^{nc}	10
1.3.4	L'importance de l'étude du protéome alternatif	11
1.4	Outils de prédiction et de détection des protéines alternatives traduites et fonctionnelles	12
1.4.1	Les bases de données de protéines alternatives	12
1.4.2	Les évidences de traduction	13
1.4.3	La détection par spectrométrie de masse	14
1.4.4	La conservation	16
1.4.5	Les signatures moléculaires et les domaines protéiques	17
1.5	La découverte de la protéine alternative altMiD51	17
1.5.1	Les évidences de traduction d'altMiD51	18
1.5.2	Détection par spectrométrie de masse d'altMiD51	20
1.5.3	La conservation d'altMiD51	20
1.5.4	La signature moléculaire d'altMiD51	21
1.6	La protéine de référence MiD51	22
1.6.1	La découverte du gène <i>SMCR7L</i>	22
1.6.2	La morphologie mitochondriale et les protéines associées	23
1.6.3	MiD51 est impliquée dans la fission mitochondriale	25
1.7	Hypothèse de recherche : Deux protéines, encodées par un même ARNm, impliquées dans la fission mitochondriale	29
2	Article	31
3	Discussion	67
3.1	Le potentiel inexploré de l'ARNm eucaryote	67
3.2	Le potentiel multicodeur de l'ARNm du gène <i>SMCR7L</i>	69
3.2.1	Un uORF codant pour une protéine fonctionnelle	70
3.2.2	AltMiD51 est une protéine à domaine LYR : LYRM12	71
3.3	La coopération fonctionnelle entre les protéines alternatives et leurs références	73

3.4	Deux protéines encodées par le même ARNm sont impliquées dans la fission mitochondriale	74
3.4.1	La fission mitochondriale Drp1 dépendante ou indépendante . . .	74
3.4.2	La fission engendrée par la surexpression d'altMiD51 n'affecte pas le métabolisme mitochondrial	78
4	Conclusions	79
5	Annexes	93
5.1	Figures supplémentaires de l'article	93
5.2	Tableaux supplémentaires de l'article	120
5.3	Matériels et méthodes de la figure 3.2	124
5.4	Matériels et méthodes de la figure 3.3	124

LISTE DES TABLEAUX

3.1	Les altORFs contenus dans le gène <i>SMCR7L</i> humain.	69
-----	---	----

LISTE DES FIGURES

1.1	ARNm mature eucaryote selon le dogme de la biologie moléculaire. . . .	2
1.2	Les ARNms matures eucaryotes possèdent un CDS et des altORFs dans les différents cadres de lecture.	5
1.3	Une base de données pour la détection par spectrométrie de masse des protéines alternatives dans les expériences biologiques.	15
1.4	AltMiD51 est une protéine alternative provenant d'un altORF ^{5'} dans l'ARNm du gène <i>SMCR7L</i>	18
1.5	Profilage ribosomal de l'initiation et l'élongation de traduction de l'altORF d'altMiD51.	19
1.6	Alignement multiple de l'altORF conservé d'altMiD51 provenant du génome de 100 vertébrés.	21
1.7	Première image de localisation subcellulaire de MiD51 dans LIFEdb. . .	23
1.8	Le dynamisme mitochondrial : régulation entre la fusion et la fission des mitochondries.	24
1.9	La surexpression de MiD51 entraîne la fusion des mitochondries et une agglomération du réseau mitochondrial près du noyau.	26
3.1	L'ARNm du gène <i>SMCR7L</i> est multicodeant pour MiD51 et plusieurs altORFs.	69
3.2	Localisation subcellulaire de l'interaction entre altMiD51 et ACPm par complémentation bimoléculaire fluorescente (BiFC).	72
3.3	L'influence de la surexpression d'altMiD51 sur les GTPases impliquées dans le dynamisme mitochondrial.	76

LISTE DES ABRÉVIATIONS

- CDS** Région codante *Coding sequence*
- ORF** Cadre ouvert de lecture *open reading frame*
- sORF** Petit cadre ouvert de lecture *small open reading frame*
- UTR** Région non-traduite *Untranslated region*
- AltORF** Cadre ouvert de lecture alternatif *Alternative open reading frame*
- AltORF^{5'}** AltORF dans le 5'UTR *AltORF in the 5'UTR*
- AltORF^{3'}** AltORF dans le 3'UTR *AltORF in the 3'UTR*
- AltORF^{CDS}** AltORF à l'intérieur du CDS *AltORF nested in the CDS*
- AltORF^{nc}** AltORF dans un ARN non codant *AltORF in a non coding RNA*
- uORF** Cadre ouvert de lecture en amont *Upstream open reading frame*
- oORF** Cadre ouvert de lecture en amont chevauchant le CDS *overlapping open reading frame*
- GFP** Protéine fluorescente verte *Green fluorescent protein*
- EST** marqueur de séquence exprimée *Expressed Sequence Tag*
- SMS** syndrome Smith-Magenis *Smith-Magenis syndrome*
- SMCR7** Région critique 7 du SMS *Smith-Magenis syndrome critical region 7*
- SMCR7L** Candidat semblable à SMCR7 *Smith-Magenis syndrome candidate 7 protein-like*
- MiD49/51** Protéine du dynamisme mitochondrial de 49/51kDa *Mitochondrial Dynamics proteins of 49/51 kDa*
- CCCP** Carbonyl cyanide m-chlorophenyl hydrazone *carbonyl cyanide m-chlorophenylhydrazone*
- ATP** Adénosine triphosphate *Adenosine triphosphate*

- ADP** Adénosine diphosphate *Adenosine diphosphate*
- GTP** Guanosine triphosphate *Guanosine triphosphate*
- ROS** Dérivé réactif de l'oxygène *Reactive oxygen species*
- ACPm** Protéine mitochondriale porteuse d'acyle *Mitochondrial acyl carrier protein*
- Drp1** Protéine reliée aux dynamines 1 *dynamain-related protein 1*
- Opa1** Atrophie optique 1 *Optic atrophy 1*
- Mfn1/2** Mitofusine-1/2 *Mitofusin-1/2*
- ARNm** Acide ribonucléique messenger *Messenger ribonucleic acid*
- ARNnc** ARN non codant *non codant RNA*
- ARNlnc** ARN long non codant *Lon non codant RNA*
- ARNr** ARN ribosomal *Ribosomal RNA*
- ARNt** ARN de transfert *Transfer RNA*
- siARN** Petit ARN interférent *small interfering RNA*
- miARN** Micro ARN *Micro RNA*
- snoARN** Petit ARN nucléolaire *Small nucleolar RNA*

*À mes parents, Julie et Guy, qui m'ont appris à toujours persévérer dans tout ce que
j'entreprends dans ma vie.*

«Que vous pensiez en être capable ou ne pas en être capable, dans les deux cas, vous avez raison.» -Henry Ford

REMERCIEMENTS

Pour commencer, j'aimerais remercier mon directeur de recherche Xavier Roucou. Il m'a accueilli dans son laboratoire et m'a confié un projet très stimulant, voire novateur. J'aimerais te remercier de la confiance et la grande estime envers ma personne et mon travail dont tu m'as fait part. J'aimerais remercier également François-Michel Boisvert pour son mentorat et l'évaluation de ce mémoire, ainsi que Guylain Boissonneault pour l'évaluation de ce mémoire.

Ensuite, j'aimerais grandement remercier les assistants de recherche Julie Motard et Jean-François Jacques pour avoir été des «parents scientifiques» dévoués et exemplaires lors de mon cheminement professionnel et académique. Ils ont grandement contribué à ma réussite, mais aussi à la qualité de mes expériences, de mon esprit scientifique et de mes connaissances. J'aimerais aussi remercier l'assistante de recherche Mylène Brunelle pour son temps, ses conseils et sa rigueur scientifique.

De plus, j'aimerais remercier Marie Brunet pour les corrections apportées à mon mémoire, son amitié et surtout pour m'avoir donné un modèle de grande scientifique pour l'avenir. Tes connaissances variées, ta rigueur, ta créativité et ton esprit scientifique sont édifiants. Un grand merci à Sondos Samandi pour avoir partagé avec moi ces deux ans de travail pour créer un article dont nous pouvons être très fières. Un grand merci à Vivian Delcourt pour son précieux temps pour toutes les analyses de spectrométrie de masse et ses explications diverses sur la chimie. Un autre merci spécial à Maxime Beaudoin pour avoir partagé avec moi les hauts et les bas de cette aventure altMiD51. Également, j'aimerais remercier tous les collègues de travail pour l'ambiance agréable au laboratoire, les soirées ensemble à travailler tard et l'entraide.

Enfin, j'aimerais remercier tous ces gens dans ma vie personnelle qui ont grandement contribué à ma réussite scolaire et professionnelle et qui m'ont soutenue tout au long de ma maîtrise. Un énorme merci à mes parents pour leur support moral et financier lors de toutes mes études. J'ai une chance inouïe d'avoir des parents qui sont fiers de moi et qui m'aident à réaliser mes rêves. Un merci spécial à Martin Cousineau pour son aide avec Latex, sa présence, son soutien moral et son amour. Tu m'as permis de persévérer dans les moments difficiles. Finalement, j'aimerais remercier infiniment tous mes amis : Mathilde

Chevin, Andréanne Loiselle, Audrey Daigneault, Ariane Brault, Jean-Michel Garant, Joël D'Astous-Pagé, Patrick Asselin-Mullen et Simon Boudreault. Vous avez tous et toutes contribué de près ou de loin à mon épanouissement personnel et professionnel par votre amitié, votre soutien, votre bienveillance et votre passion de la science. Vous êtes tous et toutes des amis en or et de grands scientifiques.

1 INTRODUCTION

1.1 Un ARN messenger mature et une protéine

L'ensemble de tous les gènes ou génome contient l'information nécessaire au fonctionnement de la cellule. Il est finement régulé pour exprimer les gènes au moment opportun afin de répondre aux besoins de la cellule. Globalement, le dogme de la biologie moléculaire stipule que l'information génétique véhiculée par l'ADN est transcrite en un ou plusieurs ARN messagers (ARNms) et ceux-ci possèdent une séquence codante traduite en une protéine particulière. Voyons plus en détail ce dogme afin de mieux comprendre ce qu'il constitue.

Spécifiquement, l'ADN est transcrit en ARNms immatures via les ARN polymérases et ces ARNms subiront plusieurs modifications pour devenir des ARNms matures prêts pour la traduction en protéines. La maturation implique notamment l'épissage par le spliceosome. L'épissage est un mécanisme permettant aux régions introniques non codantes (introns) d'être excisées pour ne conserver que les régions exoniques (exons) codantes pour des protéines. Le mécanisme d'épissage permet aussi de contrôler l'expression ou l'activité d'une protéine par l'agencement des exons entre eux. Cet agencement crée une diversité protéique par la création d'isoformes (Pan *et al.*, 2008). Aussi, l'ARNm sera modifié par l'ajout de la coiffe 5' et de la polyadénylation en 3' de sa séquence. Ces modifications permettent la stabilité de l'ARNm et sa prise en charge par la machinerie cellulaire. Une fois mature, un ARNm contient deux régions non codantes (UTR pour *untranslated region* en anglais) en 5' et en 3' et une région codante (CDS pour *coding sequence en anglais*) aussi appelée cadre de lecture ouvert (ORF pour *open reading frame* en anglais). Pour éviter les confusions dans ce mémoire, cet ORF codant pour une protéine est appelé CDS et code pour une «protéine de référence» (Figure 1.1). Chez les eucaryotes, un ARN messenger mature a été défini comme monocistronique, c'est-à-dire qu'une seule protéine est traduite à partir de celui-ci (Blumenthal, 2004; Kozak, 1999).

Cette conception de l'ARN messenger mature monocistronique découle en partie du mécanisme établi de traduction par le ribosome : la sous-unité ribosomale 43S du complexe pré initiateur reconnaît la coiffe en 5', puis glisse sur l'ARN messenger dans la région 5' jusqu'à la reconnaissance d'un codon initiateur, le plus courant étant le codon AUG (Kozak, 1978, 1989). La portion en amont de ce codon est annotée comme étant le 5' UTR. Une

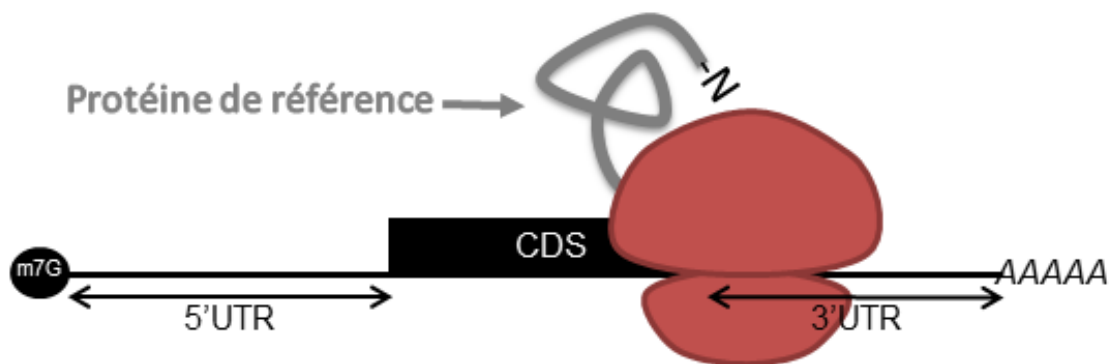


FIGURE 1.1 – ARNm mature eucaryote selon le dogme de la biologie moléculaire. Le dogme de la biologie moléculaire explique que l'information génétique véhiculé par l'ADN est transcrit en ARNm puis en protéine. Les ARNms matures eucaryotes sont dotés d'une coiffe en 5' par laquelle le ribosome commence son balayage en vue de traduire le CDS en protéine de référence.

fois le codon initiateur reconnu, la grande sous-unité 60S s'associe à la petite sous-unité 43S pour commencer la traduction par le ribosome 80S maintenant complet et fonctionnel (Kozak, 1999). Plusieurs AUG peuvent être présents dans un ARN messenger, et certains facteurs ont été identifiés pour expliquer lequel engendrera le début de la traduction d'une séquence codante pour une protéine. Il s'agit de la position du codon initiateur, la longueur du 5' UTR, les facteurs d'initiation présents, les structures secondaires en aval ou en amont du codon AUG et les nucléotides avoisinants que l'on regroupe sous le terme de contexte (Pain, 1986; Kozak, 1999). Dans le contexte d'un codon AUG, il est possible de retrouver une séquence consensus (GCCA/GCCAUGG) appelée Kozak (Kozak, 1997), contenant une purine en position -3 et une guanosine en position +4 du codon AUG (Kozak, 1987), qui facilite le choix du codon initiateur et le début de la traduction. Finalement, la fin de la traduction se produit lorsque le ribosome rencontre un des codons terminateurs : TAA, TGA ou TAG. La région en aval du codon terminateur constitue la région 3' UTR de l'ARN messenger.

1.1.1 Les ARNs non codants

Le dogme de la biologie moléculaire, comme discuté précédemment, prend en compte seulement l'ADN codant pour des protéines. En fait, la communauté scientifique a découvert que l'ADN peut aussi être transcrit en ARN non codant (ARNnc) et que ces ARNncs sont importants pour le fonctionnement de la cellule. Il existe à ce jour plusieurs types ARNncs ayant leurs fonctions propres. Notamment, les ARN ribosomiaux (ARNrs) forment les ribosomes (Attardi et Amaldi, 1970), les ARN de transfert (ARNts) per-

mettent la synthèse de protéines, les petits ARN interférents (siARNs) et les micros ARNs (miARNs) régulent l'expression des ARNms et les petits ARN nucléolaires (SnoARNs) modifient d'autres ARN non codants (Willingham et Gingeras, 2006; Kiss, 2002; Zamore et Haley, 2005; Mattick et Makunin, 2006). Il existe aussi un autre type d'ARN provenant des régions introniques et intergéniques du génome : les longs ARN non codants (ARNlnc) (Okazaki *et al.*, 2002). Les critères définissant un ARNlnc sont les suivants : une localisation dans les régions non codantes du génome et une longueur minimale de 200 nucléotides (Abouleish, 1987; Kapranov *et al.*, 2007).

1.2 L'annotation des ORFs du génome humain

Les régions codantes pour des protéines et non codantes ont été déterminées lors de l'annotation du génome humain par plusieurs groupes de recherche à travers le monde (Lander *et al.*, 2001; Venter *et al.*, 2001). Du fait de la complexité du génome humain, des analyses bioinformatiques et expérimentales à grande échelle ont été mises en commun dans le but de trouver le maximum de régions codantes. La prédiction des gènes a été réalisée avec trois approches : une évidence de transcription démontrée par les ARNms et les marqueurs de séquences exprimées (ESTs pour *Expressed Sequence Tag* en anglais), les similarités de séquences avec les gènes et protéines déjà connus et la reconnaissance *ab initio* des groupes d'exons par les sites d'épissages, le biais de codon, et la longueur des exons et introns (Lander *et al.*, 2001). Cette manière de procéder a contribué à la vision actuelle d'une séquence codante unique par locus ou transcrit (Okazaki *et al.*, 2002).

Après l'annotation du premier génome eucaryote, celui de l'organisme modèle *Saccharomyces cerevisiae*, Munira A. Basrai *et al.* ont remarqué l'ampleur des petits ORFs de moins de 100 codons que pouvait contenir le génome de la levure. Par ailleurs, les ORFs de moins de 100 codons (260 000) représentent 40 fois plus d'ORFs que le nombre d'ORF plus grands que 100 codons (6275). Les auteurs n'écartent pas la possibilité que certains petits ORFs puissent être traduits et fonctionnels, mais que ceux-ci soient noyés dans la quantité de petits ORFs faux positifs présents dans le génome (Basrai *et al.*, 1997). Ainsi, les génomes encore plus volumineux que celui de la levure ont eux aussi des milliers de petits ORFs et une limite de longueur pour les ORFs a été fixée pour faciliter les annotations et les prédictions de séquences codantes. Toutefois, il s'avère que certains laboratoires se sont déjà intéressés à l'analyse bioinformatique des petits ORFs (smORF ou sORF pour *small open reading frame* en anglais). Chacune des études ont leurs critères de sélection pour leurs analyses et ces études seront discutées dans la prochaine section.

1.2.1 Les petits ORFs humains sous la loupe

Une étude s'est intéressée aux sORFs contenus à l'intérieur du CDS avec un minimum de 50 acides aminés tout en regardant la conservation entre la souris, le rat et l'humain (Ribrioux *et al.*, 2008). Une autre étude s'est penchée sur la conservation entre la souris et l'humain des petits ORFs possibles à l'extérieur des CDSs, mais avec une longueur minimale de 500 nucléotides. Les auteurs avaient l'idée de trouver des gènes codants pour deux protéines plutôt qu'une seule (Chung *et al.*, 2007). Aussi, l'épissage alternatif à l'origine de changements de cadre de lecture entraînant l'apparition d'ORF a été exploré (Kovacs *et al.*, 2010; Liang et Landweber, 2006). Par contre, leurs prédictions sont passées à côté de beaucoup de candidats potentiels, autant du fait de leur étude ciblée sur certaines localisations de sORFs dans l'ARNm ou de leurs critères de longueur en acides aminés. L'étude de Vanderperre *et al* du laboratoire du Pr. Xavier Roucou dénote l'importance des altORFs et leur traduction en protéines alternatives pouvant enrichir significativement le protéome humain (Vanderperre *et al.*, 2013). Le concept d'altORFs est semblable à celui des sORFs, mais cette terminologie englobe les petits ORFs trouvés lorsqu'on exploite sous tous ses angles le potentiel multicodant de l'ARNm. Dans la prochaine section, le concept des altORFs tel qu'établi par le laboratoire du Pr Xavier Roucou sera expliqué et détaillé afin de saisir l'importance de leur étude ultérieure.

1.3 Les cadres de lecture ouverts alternatifs (AltORFs)

L'information génétique contenue dans les ARNms se décode par triplet de nucléotides pour la conversion en acides aminés par le ribosome via les ARN de transferts. Cette caractéristique du code génétique entraîne trois manières possibles pour les ribosomes de traduire l'ARNm, c'est-à-dire qu'il y a trois différents cadres de lecture possibles (Figure 1.2A). En analysant les trois cadres de lecture existant dans l'entièreté de l'ARNm, il est possible de trouver des codons initiateurs et terminateurs formant des ORFs différents des CDSs. Ces nouveaux ORFs sont appelés ORFs alternatifs (altORFs) et ceux-ci peuvent être traduits en protéines dites alternatives (Vanderperre *et al.*, 2013). Les altORF peuvent se loger dans les UTRs (altORF^{5'} et altORF^{3'}) dans les trois cadres de lecture possibles ou à l'intérieur même du CDS (altORF^{CDS}) à condition que ce ne soit pas dans le même cadre de lecture (Figure 1.2B). Ils peuvent aussi se loger dans des ARNncs (altORFs^{nc}) (Figure 1.2C). Ainsi, les protéines alternatives issues des altORFs ne sont pas des isoformes de la protéine de référence, mais de nouvelles protéines dont la séquence en acide aminés est unique. Parmi les quatre types d'altORFs possibles listés précédemment, il existe des exemples altORFs^{5'}, altORF^{CDS} et d'altORF^{nc}. Ceux-ci seront détaillés afin de mieux

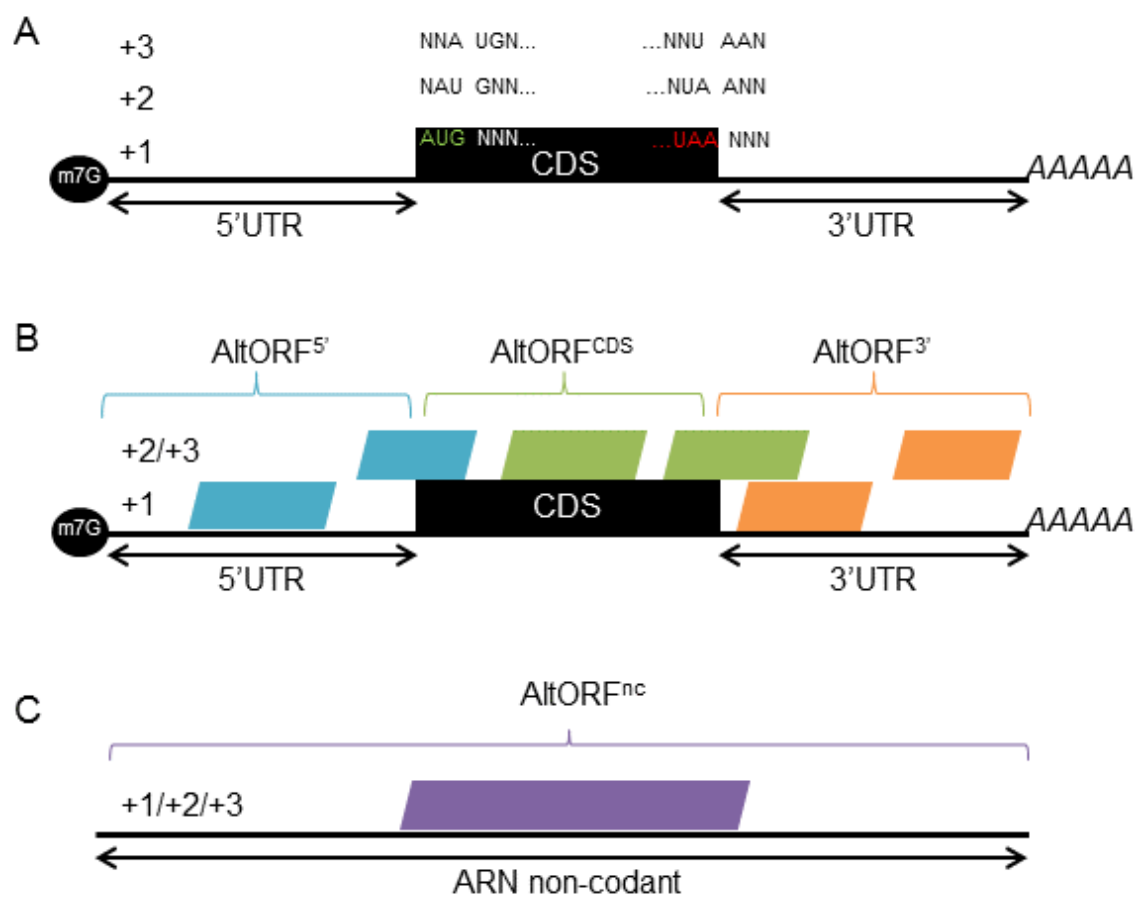


FIGURE 1.2 – Les ARNms matures eucaryotes possèdent un CDS et des altORFs dans les différents cadres de lecture.

(A) L'ARNm mature contient un CDS annoté dans les bases de données actuelles. Ce CDS se situe dans un des trois cadres de lecture pouvant être traduit par les ribosomes. (B) L'ARNm mature possède plusieurs codons initiateurs et terminateurs dans les trois cadres de lecture formant les altORFs. Ces altORFs sont complètement différents du CDS et codent pour des protéines alternatives uniques. (C) Les altORFs peuvent également se retrouver dans les ARNncs dans les trois cadres de lecture.

comprendre leurs particularités et leurs implications biologiques.

1.3.1 Les altORFs^{5'}

Comme discuté dans le paragraphe précédent, les altORFs^{5'} sont des altORFs contenus dans le 5'UTR des ARNms. Dans la littérature, ces altORFs sont souvent de très petits ORFs contenus en amont du CDS appelés uORFs pour *upstream reading frame* en anglais. Aussi, les uORFs peuvent être nommés oORFs pour *overlapping open reading frame* en anglais si le codon terminateur se retrouve à l'intérieur du CDS dans un autre cadre de lecture. Entre autres, ces petites séquences en amont du CDS sont présentes dans une grande majorité des ARNms et sont conservées chez l'humain, la souris et le poisson-zèbre (Chew *et al.*, 2016). Les uORFs sont connus majoritairement pour leur potentiel de répression de la traduction des CDSs.

1.3.1.1 les uORFs : régulateurs de la traduction

Les uORFs peuvent diminuer la traduction du CDS. D'une part, cette diminution peut être expliquée par les ribosomes qui commencent leur balayage de l'ARNm en 5' et qui rencontrent le codon initiateur de l'uORF avant celui du CDS (Chew *et al.*, 2016; Calvo *et al.*, 2009). À ceci s'ajoute que le uORF peut entraîner la rétention des facteurs d'initiation de la traduction (Young et Wek, 2016), puisque ces facteurs prennent un certain temps à s'associer et se dissocier du complexe d'initiation de la traduction. Ainsi, le uORF peut diminuer la disponibilité des facteurs pour l'initiation en aval. D'autre part, il a été démontré que la longueur d'un uORF influence l'efficacité de la traduction d'un CDS (Kozak, 2001). Un long uORF favorise le retardement de la formation du complexe d'initiation pour la traduction en aval et produit un «leaky scanning», c'est-à-dire que la sous-unité 40S ne s'arrêtera pas sur le codon initiateur et continuera son chemin jusqu'à la prochaine méthionine. Inversement, un uORF très court permettra une reprise du balayage et le temps nécessaire pour l'assemblage du complexe d'initiation de la traduction (Kozak, 2001). Outre la longueur de l'uORF, un autre point important à considérer est la longueur de la séquence séparant la fin de l'uORF et le début du CDS, car elle influence la capacité de la machinerie de traduction à reprendre la traduction en aval de l'uORF (Somers *et al.*, 2013). En fait, un uORF peut avoir un effet positif sur la traduction du CDS en augmentant la quantité de sous-unité 40S sur l'ARNm et la proximité des facteurs nécessaires à la réinitiation de la traduction (Young et Wek, 2016).

Par ailleurs, une structure secondaire ou un biais de codon engendré par le uORF peut entraîner le décrochage des ribosomes de l'ARNm. Par conséquent, le nombre de ribosomes

pouvant atteindre le CDS est diminué (Somers *et al.*, 2013). De plus, la présence d'un uORF dans un cadre de lecture différent du CDS diminue la traduction de celui-ci, car les ribosomes cherchent les AUG dans le cadre de lecture de l'uORF plutôt que celui du CDS (Young et Wek, 2016). Un phénomène un peu moins connu est la diminution de l'efficacité de la machinerie de traduction par la synthèse du peptide provenant de l'uORF (Wei *et al.*, 2012). En fait, le peptide obtenu par la traduction de l'uORF peut interférer avec la machinerie de traduction et empêcher la traduction en aval (Wei *et al.*, 2012). Finalement, les uORFs peuvent avoir un lien de répression en condition normale, mais permettre la traduction d'un CDS en condition de stress cellulaire (Andreev *et al.*, 2015; Young et Wek, 2016; Starck *et al.*, 2016).

Ce phénomène est expliqué par la phosphorylation de la protéine eIF2 dans des conditions de stress cellulaire. Cette protéine est essentielle à l'initiation de la traduction, mais sa phosphorylation diminue l'initiation de la traduction pour la plupart des ARNm. Ainsi, dans des conditions de stress, la disponibilité de la protéine eIF2 non phosphorylée (active) est moindre. Par contre, la présence d'un uORF stimule le recrutement de la machinerie cellulaire et la protéine eIF2 active, malgré qu'elle soit moins abondante, arrivera à temps pour la réinitiation de la traduction au CDS. Par conséquent, certains CDS ayant un uORF en amont sont davantage traduits en condition de stress (Andreev *et al.*, 2015; Young et Wek, 2016; Starck *et al.*, 2016).

La fonction régulatrice des uORFs met en lumière l'importance des ORFs non annotés et non explorés jusqu'à maintenant pour la cellule eucaryote et leurs implications dans des pathologies. Plusieurs uORFs sont responsables de la régulation étroite de la traduction de certaines protéines dont les niveaux en protéines peuvent entraîner une pathologie (Barbosa *et al.*, 2013; Ghilardi *et al.*, 1998). En 2016, une mutation dans le gène AR (*androgen receptor* en anglais) a été trouvée chez des patients atteints du syndrome de l'insensibilité à l'androgène. Cette mutation présente dans le 5'UTR de l'ARNm de ce gène crée un codon initiateur (AUG), donnant naissance à un uORF. Cet uORF diminue l'expression du CDS en aval codant pour le récepteur à l'androgène (Hornig *et al.*, 2016). Que les uORFs/altORFs^{5'} soient traduits ou non, il n'en demeure pas moins qu'ils ont des rôles régulateurs importants impliqués dans le fonctionnement cellulaire.

Parmi tous les altORFs^{5'}/uORFs/oORFs existants, certains sont traduits et fonctionnels et figurent dans la liste des protéines alternatives telle que définies par Vanderperre et ses collègues (Vanderperre *et al.*, 2013). Certains peuvent même avoir le rôle régulateur d'un uORF et être traduits en protéines fonctionnelles (Pueyo *et al.*, 2016; Ji *et al.*, 2015). Voyons plus en détail deux cas dont à la fois le CDS et l'altORF sont traduits en plus de

coopérer dans la même signalisation cellulaire : A2AR/uORF5 ; rAT1aR/Pep7

1.3.1.2 les uORFs/altORF^{5'} traduits et fonctionnels

Une protéine alternative humaine provenant d'un altORF^{5'} a été découverte dans un gène codant pour un récepteur couplé aux protéines G (*G protein coupled receptors* (GPCR) en anglais). De manière générale, les GPCRs sont des récepteurs membranaires essentiels aux cellules pour la réponse aux hormones et aux changements de l'environnement cellulaire (Rosenbaum *et al.*, 2009). Le gène *Adora2a* est donc un gène codant pour deux protéines : un uORF/altORF^{5'} nommé uORF5 codant pour une protéine alternative de 134 acides aminés et le CDS codant pour un récepteur adénosine couplé aux protéines G ou *A2A adenosine receptor* (A2AR) en anglais. Ce récepteur est une cible de la caféine, mais aussi de molécules thérapeutiques utilisées pour traiter des troubles du cerveau (Lee *et al.*, 2014). Ce récepteur active la voie dépendante de la PKA associée à l'AMPC. La stimulation du récepteur A2A par la forskolin augmente l'expression de l'uORF5 de manière PKA-dépendante. De plus, la protéine alternative régule l'expression de certains gènes impliqués dans la voie des MAPK (*Spag9*, *Lrrn3*, *Rasal2* et *Rshl1*). Lee *et al.* suggèrent que la protéine issue de l'uORF5 pourrait être impliquée dans la diminution des effets pro-inflammatoires suite à la stimulation du récepteur A2A.

Un autre duo a été découvert dans le gène codant pour le récepteur angiotensine de type 1 chez le rat (rAT1aR), un GPCR. Après stimulation à l'angiotensine II, rAT1aR active la voie de la phospholipase C (PLC)-inositol 1,4,5-triphosphate (IP3) et des protéines Erk1/2 (*extracellular signal-regulated protein kinases 1 and 2* en anglais) (Liu *et al.*, 2014). La protéine alternative PEP7 provient d'un altORF^{5'} codant pour seulement 7 acides aminés et ce peptide est conservé chez l'humain et la souris. Ce peptide diminue grandement l'activation des Erk1/2 par le récepteur (Liu *et al.*, 2014). En revanche, il n'influence pas les autres voies stimulées par celui-ci comme celle de la phospholipase C, si on le compare avec un peptide contrôle dont les acides aminés ont été mélangés aléatoirement (Liu *et al.*, 2014; Yosten *et al.*, 2015). Pep7 diminue l'activation des Erk par la voie des β -arrestines (Liu *et al.*, 2014; Yosten *et al.*, 2015). Pep7 est donc un uORF très petit, impliqué dans la régulation de sa protéine de référence, et qui influence des signaux cellulaires très importants dans la cellule.

Ainsi, les altORFs^{5'}, même de petite taille, peuvent être traduits en peptides et contribuent à la régulation de certaines protéines telles que des GPCRs. Cela démontre la contribution des altORFs^{5'} en tant que régulateurs de la traduction, mais aussi la contribution possible de ceux-ci en tant que protéines du protéome actuel (Tautz, 2009).

1.3.2 Les altORFs^{CDS}

La particularité des altORFs^{CDS} est leur présence dans les séquences codantes déjà connues dans les ARNm. Ainsi, lors d'une transfection de routine d'une quelconque séquence codante d'un gène d'intérêt, à la fois l'altORF^{CDS} et le CDS sont exprimés (Vanderperre *et al.*, 2011). La surexpression des deux protéines peut entraîner un biais dans les résultats expérimentaux (Li *et al.*, 2009). Il y a deux exemples dans la littérature d'altORFs^{CDS} traduits en protéines alternatives qui démontrent ou suggèrent une coopération entre la protéine de référence et la protéine alternative dans un même processus cellulaire : XL α s/Alex et ATXN1/altATXN1.

Pour commencer, l'un des cas les plus connus de la coopération entre une protéine alternative et sa référence chez l'humain est le duo XL α s et Alex (Klemke *et al.*, 2001). Il est notamment connu pour la découverte d'un gène humain codant pour un ARNm contenant deux ORFs codants. Ces protéines ont été détectées pour la première fois dans des cellules neuroendocrines (Klemke *et al.*, 2001). La protéine de référence XL α s est une protéine G(s) de sous-unité alpha de 92 kDa responsable de la transduction du signal de certains GPCRs par l'activation de l'adénylate cyclase. L'adénylate cyclase est une enzyme qui convertit l'ATP en AMPc dans la cellule. En raison de leur importance cellulaire, plusieurs GPCRs dysfonctionnels causent des pathologies comme le cancer (Dorsam et Gutkind, 2007). La protéine alternative Alex de 38 kDa provient d'un altORF^{CDS}. La séquence en acides aminés de Alex est conservée chez l'humain, le rat et la souris et peut lier le domaine XL de XL α s. Un polymorphisme causant une mutation dans l'altORF codant pour Alex a été découvert chez des patients. Ce polymorphisme change un acide aminé qui empêche la protéine alternative Alex d'interagir avec la protéine de référence XL α s. L'abolition de leur interaction entraîne un surplus d'AMPc dans la cellule par l'augmentation de l'activation de l'adénylate cyclase. Il en a été déduit qu'Alex régule l'activité de XL α s dans l'activation de l'adénylate cyclase. Cette dérégulation de XL α s entraîne un état pathologique, notamment un retard mental ou de croissance et de la brachydactylie (Freson *et al.*, 2003). En plus d'être un exemple étroit de coopération entre une protéine alternative et une protéine de référence provenant d'un même ARNm, il démontre l'implication biologique et physiologique que peuvent avoir les protéines alternatives.

Finalement, le laboratoire de Pr Xavier Roucou a exploré la protéine humaine ATXN1 et altATXN1 (Bergeron *et al.*, 2013). ATXN1 est une protéine impliquée dans l'ataxie spinocérébelleuse de type 1 (SCA1), une maladie neurodégénérative. Cette protéine a la particularité de posséder un domaine polyglutamine (polyQ) qui acquiert des propriétés

pathologiques au-delà de 30 répétitions de CAG (glutamine). Elle se localise au noyau dans des conditions physiologiques. La protéine alternative altATXN1 est aussi localisée au noyau dans des inclusions nucléaires et interagit avec sa protéine de référence normale ou pathologique (Bergeron *et al.*, 2013). Lorsqu'ATXN1 est absente, altATXN1 a une localisation nucléoplasmique dépendante de la transcription d'ARN et lie les queues poly(A) des ARNs (Bergeron *et al.*, 2013). À ce propos, Bergeron et collègues ont émis l'hypothèse que altATXN1 pourrait séquestrer ATXN1 dans les inclusions nucléaires pour réguler la fonction d'ATXN1 (Bergeron *et al.*, 2013).

1.3.3 Les altORFs^{nc}

Par ailleurs, le concept des altORFs peut s'appliquer sur les ARNncs (Figure 1.2C). Il faut comprendre que les ARNncs ont été répertoriés comme non codants, car ils ne possédaient pas d'ORF supérieurs à 100 codons lors de l'annotation du génome. Cependant, il est possible que ces ARNs non codants, plus particulièrement les ARNlncs, soient codants s'ils possèdent un ou des ORFs plus petits que le critère préétabli de longueur. En effet, de plus en plus d'ARNlncs dans les dernières années se sont avérés codants. Ils ont été découverts par l'utilisation de la technique du profilage ribosomal (Ji *et al.*, 2015; Bazzini *et al.*, 2014) et/ou par détection de la protéine qui en découle par spectrométrie de masse (Prabakaran *et al.*, 2014; O'Carroll *et al.*, 2013), des techniques expérimentales que je discuterai en détail dans la section 1.4.

Dans la littérature, il existe quelques exemples d'ARNncs qui sont devenus des ARNs codants (ARNm) par la découverte d'un ou plusieurs ORFs traduits. C'est le cas d'ELA, une hormone peptidique de 54 acides aminés, provenant d'un gène considéré non codant, et présente dans les cellules souches embryonnaires durant l'embryogenèse. ELA est le premier ligand parmi ceux déjà connus à lier le récepteur Aplyn pour entamer la différenciation de l'endoderme menant à la cardiogenèse (Chng *et al.*, 2013). Cette séquence génomique est passée d'ARNlnc à ARNm par la découverte de sa fonction.

Un autre exemple est celui du gène *mlpt* (*mille-patte*). Il a été découvert dans l'insecte *Tribolium castaneum*, un organisme modèle dans l'embryogenèse et est exprimé dans la région abdominale de l'insecte (Savard *et al.*, 2006). La séquence d'ARN découlant du gène *mlpt* s'apparentait beaucoup à un ARNm, mais possédait de trop petits ORFs pour être considéré un ARNm codant. En fait, il est un autre exemple d'ARNlnc devenu un ARNm codant. Cet ARNm est polycistronique, car il code pour quatre petits peptides de moins de 30 acides aminés (Savard *et al.*, 2006). Les trois premiers peptides possèdent un motif Kozak et ont une séquence consensus (LDPTGXY). Le quatrième peptide a une

séquence consensus (ETSSGRRRR). Ces séquences consensus sont retrouvées chez des homologues dans le gène *mlpt* de la *Drosophila* (Savard *et al.*, 2006). L'absence du gène engendre un surplus de pattes chez *Tribolium castaneum*, ce qui suggère qu'il est impliqué dans la segmentation abdominale et thoracique lors du développement (Savard *et al.*, 2006). Toutefois, la fonction exacte de chaque peptide et leur implication dans la segmentation embryonnaire sont encore inconnus (Savard *et al.*, 2006).

1.3.4 L'importance de l'étude du protéome alternatif

1.3.4.1 Un enrichissement du protéome actuel

Même si nous connaissons l'existence des petits ORFs dans les eucaryotes depuis environ une vingtaine d'années, ces altORFs ont presque été ignorés jusqu'à présent (Kastenmayer *et al.*, 2006). Les défis liés à l'étude des petites protéines (détection/étude de fonction) sont l'une des raisons pour laquelle ces protéines n'ont pas suscitées beaucoup de curiosité de la part des scientifiques. Pourtant, comme nous avons vu dans les sections précédentes, de plus en plus de petites protéines découvertes ont des fonctions importantes.

En 2013, Vanderperre et collègues ont démontré l'ampleur du protéome alternatif avec l'ajout d'un potentiel de 83 886 protéines alternatives uniques avec une longueur minimum de 40 acides aminés s'ajoutant aux 68 000 protéines de référence déjà connues. Même si seulement 10% du protéome alternatif est fonctionnel, cela représente 12% de protéines additionnelles dans le protéome actuel. Ces nouvelles protéines peuvent être impliquées, de façon directe ou indirecte, dans des pathologies dont la cause est encore inconnue. Ces petites protéines représentent également de potentiels biomarqueurs pour dépister certaines pathologies le plus précocement possible ou pour suivre l'évolution d'un traitement (Vanderperre *et al.*, 2013). Elles représentent aussi des cibles thérapeutiques potentielles, au même titre que les protéines de référence (Steen et Mann, 2004). Ce sont avant tout d'autres protéines de la cellule impliquées dans des processus biologiques et contribuant à la diversité du vivant. Leur découverte et leur visibilité dans la littérature piquent la curiosité de la communauté scientifique mondiale (Slavoff *et al.*, 2014; Olexiouk et Menschaert, 2016; Cabrera-Quio *et al.*, 2016; Cassidy *et al.*, 2016; Nam *et al.*, 2016). Pour pouvoir étudier les protéines alternatives, il faut avant tout les identifier. Il est possible de traduire *in silico* les altORFs et de les regrouper dans une base de données.

Comme mentionné précédemment, une quantité importante d'altORFs se logent dans le petit génome eucaryote de levure, donc le génome humain n'échappe pas au phénomène. Pour pouvoir identifier les altORFs traduits et fonctionnels parmi la masse d'altORFs po-

tentiellement sans fonctions (faux positifs), il est possible d'utiliser des techniques expérimentales telles que le profilage ribosomal et la spectrométrie de masse et des logiciels bio-informatiques pouvant analyser leurs séquences : conservation et prédiction de signature moléculaire.

1.4 Outils de prédiction et de détection des protéines alternatives traduites et fonctionnelles

Ces petites séquences codantes issues des altORFs sont sous-estimées par rapport à leur potentiel fonctionnel ou même leur capacité de traduction du fait du peu d'acides aminés qu'elles contiennent. Les bases de données actuelles disponibles et libres d'accès comme UniProt ne prennent pas en compte l'existence de ces protéines alternatives, qui pourtant, passent de plus en plus sous la loupe de la communauté scientifique (Andrews et Rothnagel, 2014). Il existe pourtant des bases de données disponibles de protéines alternatives.

1.4.1 Les bases de données de protéines alternatives

La première base de données établie pour les protéines alternatives regroupe les altORFs^{CDS} traduit *in silico* de 24 codons minimums commençant par un AUG dans les ARNm du transcriptome humain (Vanderperre *et al.*, 2012). Cette même base de données prend en compte les altORFs^{CDS} possédant un contexte Kozak. Ensuite, Vanderperre et collègues en 2013 ont enrichi cette base de données en incluant toutes les localisations d'altORFs : ceux contenus dans le CDS, ceux dans les UTRs, et ceux chevauchant le CDS (Vanderperre *et al.*, 2013). De plus, ce groupe de recherche a regardé la conservation entre les espèces, la présence ou non d'un motif Kozak et une longueur de 40 codons minimum pour un altORFs (Vanderperre *et al.*, 2013). Elle a permis au groupe de recherche du Pr Xavier Roucou de détecter, par spectrométrie de masse, plusieurs protéines alternatives endogènes dans des lignées de cellules et des tissus humains sains ou cancéreux, révélant la capacité de ces altORFs à pouvoir être traduits par les ribosomes (Vanderperre *et al.*, 2013). Certaines de ces protéines alternatives détectées ont été localisées par immunofluorescence *in cellulo* et leur co-expression avec la protéine de référence a été démontrée (Vanderperre *et al.*, 2013). Il existe aussi d'autres bases de données qui s'intéressent aux altORFs. Étant donnée l'émergence de la découverte d'altORFs^{5'}, une base de données s'est ajoutée en janvier 2014 au nom de uORFdb Wethmar *et al.* (2014). Elle regroupe la littérature concernant les uORFs contenus dans la base de données Pubmed. Elle permet de trouver facilement la fonction découverte d'un uORF, le gène associé, le taxon et le type d'étude conduit sur l'uORF. Une autre base de données est apparue en 2014 concernant encore une fois les uORF : uPEPPERoni (Skarshewski *et al.*, 2014). C'est une base de données qui contient les sORFs de la région 5'UTR des ARNm. Elle a été créée dans

le but d'identifier des peptides potentiellement fonctionnels provenant des uORF et qui sont conservés chez les eucaryotes (Skarszewski *et al.*, 2014).

Il est difficile de déterminer lesquelles parmi toutes ces petites protéines alternatives passées inaperçues sont réellement traduites, stables et fonctionnelles. Il est toutefois possible d'appliquer plusieurs outils bio-informatiques tels que ceux utilisés pour les bases de données du groupe de recherche de Xavier Roucou pour prédire les protéines alternatives potentiellement traduites, exprimées, conservées et fonctionnelles.

1.4.2 Les évidences de traduction

Il est possible de faire des analyses pour évaluer si de nouvelles séquences sont potentiellement traduites en regardant la présence du motif Kozak (Pop *et al.*, 2014). Comme discuté plus tôt, ce motif est un indice de traduction par les ribosomes. Par conséquent, il est possible par bio-informatique de regarder à grande échelle si les altORFs contiennent un motif Kozak parfait ou partiel. Néanmoins, seulement une faible proportion des séquences codantes déjà connues possèdent un motif Kozak (Smith *et al.*, 2005).

De même, il est possible de regarder le potentiel traductionnel d'un ORF par la technique de profilage ribosomal qui permet de voir la position des ribosomes en traduction sur les ARNms (Ingolia *et al.*, 2009). Cela est possible par l'ajout de cycloheximide, une drogue qui inhibe l'élongation de la traduction et fige les ribosomes sur l'ARNm (Ingolia, 2010). Plus spécifiquement, il est possible d'immobiliser seulement les ribosomes en initiation de traduction aux sites d'initiation sans affecter les ribosomes en élongation par l'ajout de harringtonine ou de lactimidomycine (Ingolia, 2014). Ensuite, les ARNs non protégés par des ribosomes sont dégradés par une nucléase et les régions protégées sont délogées des ribosomes puis séquencées (Ingolia *et al.*, 2009). Le nombre de fragments protégés par les ribosomes provenant d'un même gène permet une quantification de l'expression de ce gène, à condition que le tout soit normalisé par rapport à la longueur du gène (Ingolia *et al.*, 2009). Cette technique permet d'étudier le potentiel multicodant des ARNms dans les trois cadres de lecture (Michel *et al.*, 2012).

Avec l'augmentation grandissante du nombre de groupes de recherche qui génèrent des données de profilage ribosomal est venu le besoin de compiler et visualiser l'ensemble de ces données à l'échelle génomique. Pour ce faire, Michel et collègues ont créé l'outil bio-informatique GWIPS-viz (Michel *et al.*, 2014). L'outil permet de regarder les résultats de séquençages transposés sur le génome entier et de regrouper les résultats d'un ensemble d'études (Michel *et al.*, 2014). Il permet de voir au nucléotide près le nombre de ribosomes présents à un endroit précis du génome. Les études sont classées selon le trai-

tement que les cellules reçoivent, c'est-à-dire selon le type de ribosomes figés : ceux en élongation ou en initiation de traduction. GWIPS-viz permet notamment de visualiser des évidences de traduction en amont ou en aval du CDS pour la découverte de protéoformes (Michel *et al.*, 2015). Récemment, une autre base de données permettant de regrouper les données de profilage ribosomal s'est ajouté : RPFdb (*Ribosome Protected Fragments* en anglais) (Xie *et al.*, 2015). La particularité de cette base de données est qu'elle permet une vue d'ensemble sur une étude de profilage ribosomale et la manière dont les données ont été traitées (Xie *et al.*, 2015). Elle permet aussi de voir les résultats pour chacun des réplicats d'une étude. Également, elle permet de voir une liste des ARNm les plus traduits pour chaque échantillon d'une étude. Finalement, RPFdb permet de visualiser sur le génome les comptes normalisés de ribosomes pour chaque ARNm traduit.

1.4.3 La détection par spectrométrie de masse

Pour identifier des protéines alternatives exprimées dans des cellules ou des tissus, Benoît Vanderperre a eu recours à la spectrométrie de masse, une technique expérimentale bien connue et sensible. La spectrométrie de masse est une technique qui permet de mettre en évidence l'expression des protéines, mais aussi de mettre en lumière leur localisation et leur interactôme (Aebersold et Mann, 2003). L'identification des protéines peut se faire à partir d'échantillons de cellules ou de tissus préalablement digérés par une enzyme pour générer de petits peptides uniques à celles-ci. L'analyse de ces peptides se fait sur un appareil appelé spectromètre de masse. Ce dernier possède une source d'ionisation pour charger les peptides. Les peptides sont ensuite séparés par leur ratio masse/charge et détectés. L'appareil conçoit un spectre de ces peptides avec leur ratio masse/charge. Ultérieurement, on identifie la protéine qui contient ces peptides à l'aide des bases de données disponibles des protéines connues et d'un logiciel d'analyse de données de spectrométrie de masse (Aebersold et Mann, 2003). Ce logiciel simule *in silico* une digestion enzymatique des séquences des protéines contenues dans la base de données pour générer des peptides uniques et les apparier aux peptides détectés dans les expériences expérimentales (Figure 1.3).

Grâce à la base de données contenant les protéines alternatives, il est possible d'identifier les peptides uniques des protéines alternatives dans des échantillons de cellules ou de tissus ou encore de réanalyser des données publiées générées par d'autres laboratoires (Vanderperre *et al.*, 2013). Plus récemment, des logiciels comme PeptideShaker (Vaudel *et al.*, 2015) permettent l'analyse d'ensemble de données de spectrométrie de masse en combinant l'expertise de plusieurs logiciels d'analyse pour maximiser la détection de protéines.

Bio-informatique

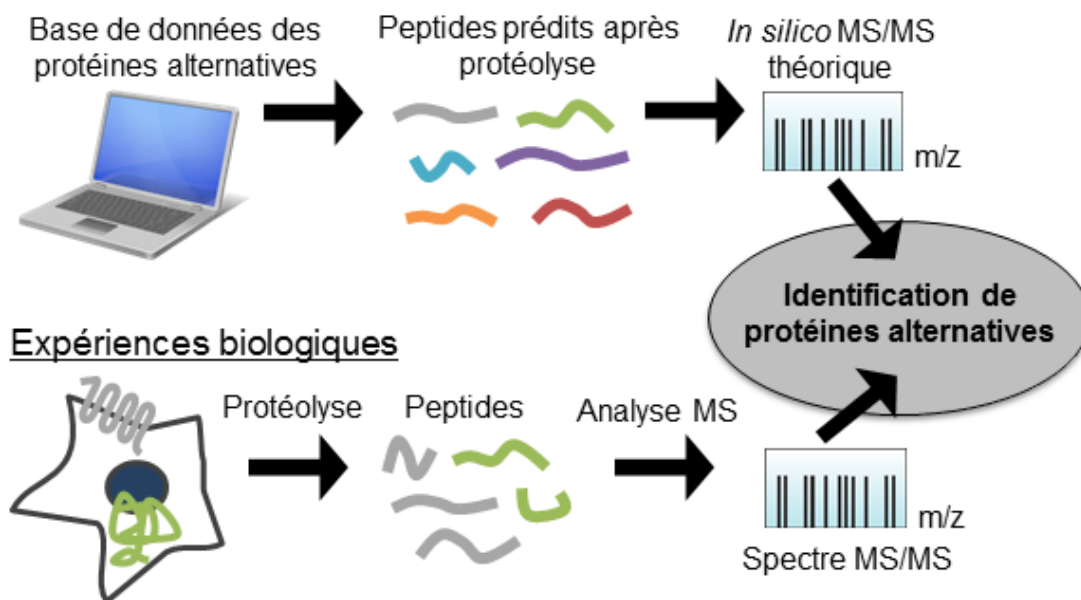


FIGURE 1.3 – Une base de données pour la détection par spectrométrie de masse des protéines alternatives dans les expériences biologiques.

Adapté de (Duncan *et al.*, 2010). La base de données du laboratoire du Pr. Xavier Roucou regroupe toutes les protéines alternatives possibles à partir des ARNm contenus dans NCBI. Chaque transcrit (ARNm) est traduit *in silico*. On peut ensuite simuler une digestion de ces protéines par la trypsine et générer des spectres théoriques *in silico* de ces peptides. En parallèle, un échantillon biologique peut être digéré à la trypsine et analysé par un spectromètre de masse pour générer les spectres des peptides. Un logiciel permet de comparer les spectres théoriques prédits et ceux obtenus expérimentalement à partir de l'échantillon biologique. Ceci permet d'identifier les peptides des protéines alternatives présentes dans l'échantillon biologique.

En fait, il a été démontré que l'analyse de données de spectrométrie de masse par plusieurs logiciels d'analyse permettait d'augmenter le nombre de protéines identifiées avec un niveau de confiance plus élevée (Vaudel *et al.*, 2015; Shteynberg *et al.*, 2013; Nesvizhskii, 2010).

1.4.4 La conservation

La conservation évolutive est l'étude de la similarité des séquences d'ADN, d'ARN ou d'acides aminés comparés entre les espèces ou à l'intérieur d'un même organisme. Elle est un outil de comparaison pour relier la fonction d'un gène trouvé, par exemple, dans un organisme modèle et un gène dans une autre espèce comme l'humain (Remm *et al.*, 2001). On parle alors d'étude de similarité (homologie) des gènes. Deux gènes sont homologues s'ils descendent du même gène ancestral. Il existe deux catégories de gènes homologues : les orthologues et les paralogues.

Deux gènes sont orthologues s'ils se sont séparés par spéciation. Généralement, ils sont dans deux espèces différentes. Alors que deux gènes paralogues proviennent d'un phénomène de duplication dans une même espèce.

L'analyse de la conservation des séquences a été utilisée ultérieurement pour prédire les séquences codantes du génome humain lors de son annotation dans les années 2000 (Lander *et al.*, 2001; Venter *et al.*, 2001). Une séquence non conservée à travers les espèces ne justifie pas que cette séquence est non-codante ou non-fonctionnelle, puisqu'elle peut être spécifique/propre à l'espèce (Clamp *et al.*, 2007).

Toutefois, il est possible d'étudier l'homologie des gènes par l'outil InParanoid qui permet d'identifier les orthologues et paralogues en comparant les meilleurs alignements réciproques de séquences entre deux espèces (Sonnhammer et Östlund, 2015). En identifiant les gènes homologues présents chez l'espèce humaine et une autre espèce, cela permet de relier les fonctions moléculaires retrouvées dans l'une ou l'autre (Goodstadt et Ponting, 2006). Ainsi, la découverte d'orthologues ou de paralogues d'altORFs humains dans d'autres espèces peut donner un indice du potentiel codant et même fonctionnel des altORFs (Tatusov *et al.*, 1997).

Également, il est possible d'identifier des éléments fonctionnels d'une séquence en regardant la vitesse de substitution de leurs nucléotides à travers plusieurs espèces. L'outil bio-informatique PhyloP permet cette analyse. Il peut calculer un score ou un P-value de conservation ou d'accélération en se basant sur les données d'alignements multiples de séquences et un modèle évolutif neutre. (Pollard *et al.*, 2010).

1.4.5 Les signatures moléculaires et les domaines protéiques

Pour démontrer qu'une protéine peut avoir une fonction, il est possible d'analyser la séquence protéique pour y trouver des domaines ou des motifs. Ces signatures moléculaires sont des régions consensus reliées soit à une fonction, une famille de protéines connues, une localisation cellulaire, une activité enzymatique et/ou une voie de signalisation particulière (Quevillon *et al.*, 2005). Une protéine possédant une région consensus est susceptible d'avoir la caractéristique fonctionnelle rattachée à cette séquence. InterProScan est un logiciel permettant la prédiction de signatures moléculaires existantes au sein des protéines ou de séquences d'ADN inconnues (Jones *et al.*, 2014). C'est un logiciel permettant de trouver les signatures moléculaires à partir de plusieurs bases de données jumelées. Cela permet d'éviter d'analyser une séquence avec plusieurs bases de données une par une (Jones *et al.*, 2014). Une autre particularité d'InterProScan est l'obtention de l'ontologie du gène (GO) (*gene ontology* en anglais), un terme regroupant la fonction moléculaire, le compartiment cellulaire et le processus biologique associé à la fonction d'un gène (Na *et al.*, 2014). Ce type d'information nous indique des pistes de fonctions possibles pour de nouvelles séquences protéiques comme celles encodées par les altORFs.

Finalement, l'utilisation de plusieurs outils bio-informatiques tel que l'étude de la conservation ou de la prédiction de signatures moléculaires peut permettre de cibler les protéines alternatives à caractériser. De plus, les analyses expérimentales comme le profilage ribosomal et la spectrométrie de masse renforcent ce choix de candidats, car elles apportent des résultats concrets de potentiel de traduction et d'expression d'une protéine. Toutes ces informations bio-informatiques et expérimentales combinées concernant les altORFs et leurs protéines alternatives peuvent servir à identifier leurs fonctions biologiques. Parmi les candidates les plus intéressants se retrouvent la protéine alternative altMiD51.

1.5 La découverte de la protéine alternative altMiD51

Dans l'article de Benoît Vanderperre et collègues en 2013, les auteurs ont détecté dans leur analyse de spectrométrie de masse « une protéine alternative chez l'humain (cellules HeLa) par la détection de 4 peptides, avec une couverture de séquence de 38% » (Vanderperre, 2013). Cette protéine alternative est traduite à partir d'un altORF contenu dans le deuxième exon du gène *SMCR7L* (Figure 1.4)(Vanderperre, 2013). Le deuxième exon se retrouve dans le 5'UTR de l'ARNm du gène *SMCR7L*. Cette protéine alternative, alt-MiD51, a 70 acides aminés (Vanderperre, 2013). Ce gène est transcrit en ARNm dont le CDS code pour la protéine MiD51 (Vanderperre, 2013). Dans sa thèse doctorale, Benoît Vanderperre explique que de coupler la détection par spectrométrie de masse à une ana-

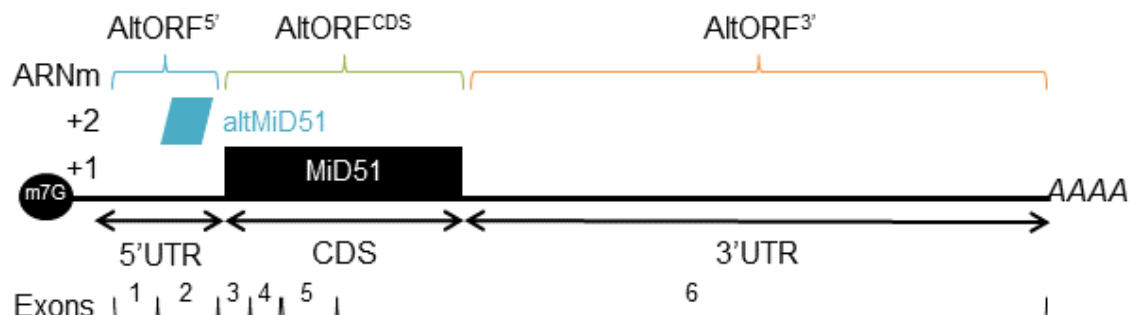


FIGURE 1.4 – AltMiD51 est une protéine alternative provenant d'un altORF^{5'} dans l'ARNm du gène *SMCR7L*.

Représentation de l'ARNm du gène *SMCR7L* (NM_019008.4). Dans l'exon 3 à 6 dans le cadre de lecture +1 se retrouve le CDS codant pour la protéine MiD51 de 463 acides aminés. Dans l'exon du 5'UTR dans le cadre +2, l'on retrouve un altORF de 210 nucléotides codant pour une protéine alternative de 70 acides aminés : altMiD51.

lyse de conservation évolutive augmente la confiance de sélection de candidats de protéine alternative à étudier (Vanderperre, 2013). AltMiD51 est à la fois une protéine conservée et détectée par spectrométrie de masse, donc elle est une bonne candidate d'étude (Vanderperre, 2013). Voyons voir plus en détail les études sur cette candidate.

1.5.1 Les évidences de traduction d'altMiD51

Les résultats générés de profilage ribosomal et représentés sur les différents génomes sont des outils pour déterminer si un altORFs présente des indices de traduction par les ribosomes. Dans le cas de la protéine alternative altMiD51, plusieurs études démontrent la présence de ribosomes au site d'initiation (codon initiateur AUG) de l'altORF codant pour altMiD51 (Figure 1.5A) (Stern-Ginossar *et al.*, 2012; Lee *et al.*, 2012; Fritsch *et al.*, 2012; Gao *et al.*, 2015). Ces études ont été réalisées à l'aide de drogues telles que la lactimidomycine, harringtonine permettant de figer les ribosomes uniquement lorsqu'ils initient la traduction d'un ORF. Certaines utilisent de la puromycine pour dégager les ribosomes en élongation (Gao *et al.*, 2015; Fritsch *et al.*, 2012). Notamment, altMiD51 peut être traduite dans les différents types cellulaires utilisés lors de ces études de profilage ribosomal : HEK293, fibroblastes, monocytes, épithéliales embryonnaires.

De plus, GWIPS-viz regroupe autant les études des ribosomes en initiation de la traduction que les études montrant la densité des ribosomes en élongation par l'ajout de drogues spécifiques à l'élongation (Michel *et al.*, 2014). En effet, plusieurs études de profilage ribosomal démontrent une densité de ribosomes importante à l'intérieur de l'altORF d'alt-

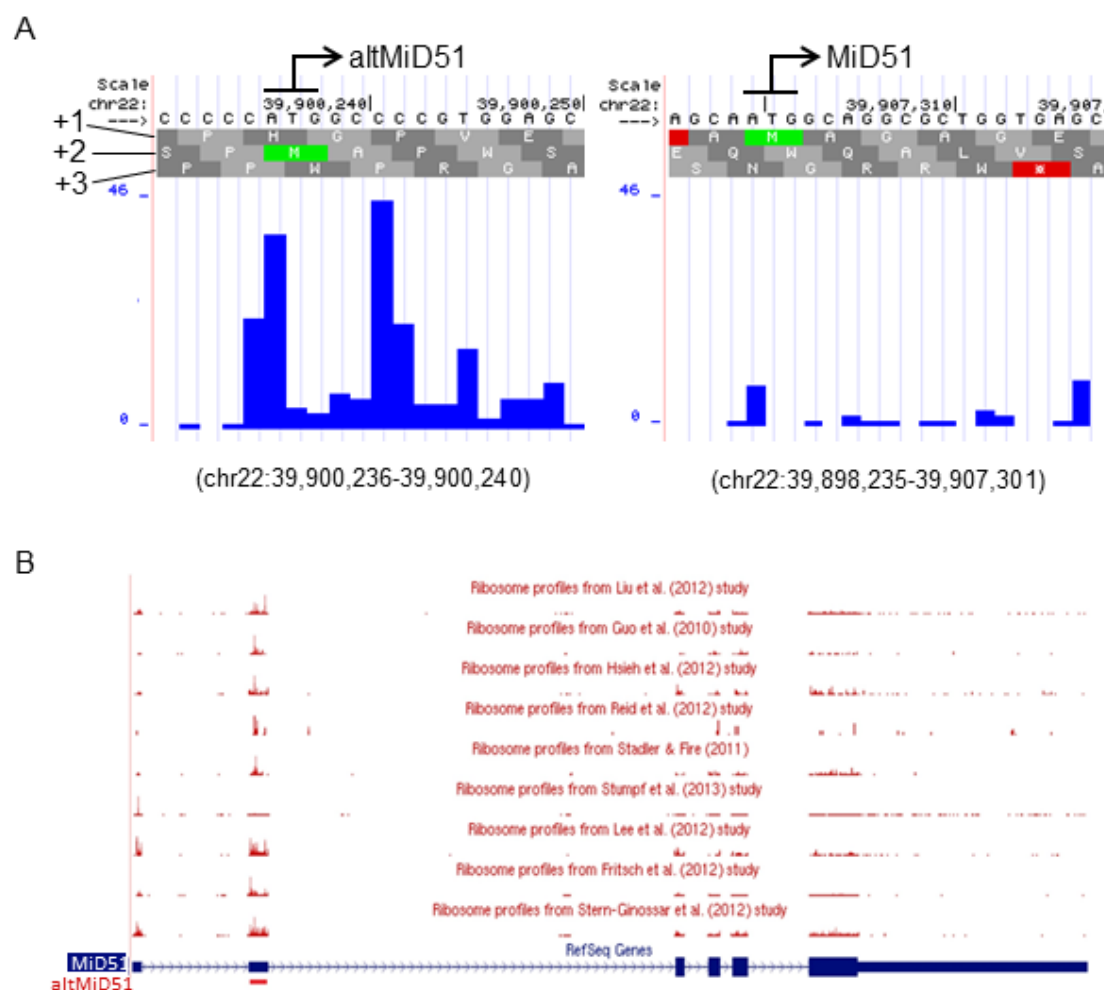


FIGURE 1.5 – Profilage ribosomal de l’initiation et l’élargissement de traduction de l’altORF d’altMiD51.

(A) inspiré de (Mouilleron *et al.*, 2016). Représentation de la somme des ribosomes détectés sur la méthionine (ATG de l’ARNm) d’altMiD51 à gauche ou de MiD51 à droite. Le compte total de ribosomes provient de la somme des ribosomes détectés dans les quatre études de profilage ribosomal d’initiation de traduction disponible sur GWIPS-viz (Michel *et al.*, 2014). (B) Adapté de (Andreev *et al.*, 2015). Vue de l’ensemble de l’ARNm du gène de MiD51 (NM_019008.4) avec neuf différentes études de profilage ribosomal d’élargissement de traduction disponible sur GWIPS-viz. Cette vue d’ensemble permet de visualiser qu’il y a beaucoup de ribosomes en élargissement dans l’exon 2 où se trouve l’altORF d’altMiD51 comparé au reste de l’ARNm. Les pics correspondent au nombre total de ribosomes retrouvé au nucléotide près sur l’ARNm.

MiD51 (Andreev *et al.*, 2015). Notamment, Andreev et collègues ont découvert l'altORF d'altMiD51 lors de leur étude de profilage ribosomal en condition de stress par l'ajout d'arsenite de sodium. Ils ont voulu démontrer qu'ils n'étaient pas les seuls à observer la possible traduction de cet uORF en illustrant différentes études de profilage ribosomales (Figure 1.5B) (Andreev *et al.*, 2015). Ces études ont été effectuées avec plusieurs types de drogues différentes bloquant l'élongation des ribosomes sur des cellules épithéliales (HEK293), des cellules cancéreuses (HeLa), des fibroblastes et des monocytes (Liu *et al.*, 2013; Stadler et Fire, 2011). Par ailleurs, un autre laboratoire effectuant du profilage ribosomal a développé un outil, le RiboTaper, pour identifier les régions traduites à partir des données de profilage ribosomal dans les HEK293 (Calviello *et al.*, 2015). Cette étude a identifié altMiD51 comme étant un peptide traduit avec un traitement cycloheximide qui bloque les ribosomes en élongation (Calviello *et al.*, 2015).

Le profilage ribosomal est un indice de traduction d'un ORF, mais il ne confirme pas la traduction d'une protéine, sa stabilité dans la cellule et son potentiel fonctionnel. Pour confirmer l'expression d'une protéine, une des techniques les plus utilisées est la spectrométrie de masse.

1.5.2 Détection par spectrométrie de masse d'altMiD51

Cette protéine alternative est détectée aisément dans des analyses de spectrométrie de masse, non seulement dans notre laboratoire (Vanderperre *et al.*, 2013), mais aussi dans d'autres études publiées. En analysant des données de spectrométrie de masse sur le protéome de la matrice mitochondriale (Rhee *et al.*, 2013), Benoît Vanderperre a détecté altMiD51 et démontré sa localisation cellulaire à la mitochondrie par immunofluorescence (Vanderperre, 2013). De plus, dans les données d'une étude visant à établir un profil du protéome des tissus sains, elle a été détectée dans des tissus adultes du cœur, des glandes surrénales et des reins (Kim *et al.*, 2014). Plus tard en 2014, le même groupe de recherche de Dr. Akhilesh Pandey l'a détecté dans une étude des « missing proteins » du chromosome 22 (Pinto *et al.*, 2014).

1.5.3 La conservation d'altMiD51

Dans sa thèse doctorale, Benoît Vanderperre avait réalisé des alignements de séquences multiples avec ClustalX 2.1 avec la séquence d'altMiD51. Il avait déterminé que cette protéine alternative était conservée avec «ses orthologues dans d'autres espèces animales» (Vanderperre, 2013). Plus tard en 2015, Andreev et collègues ont découvert altMiD51 dans leur analyse de profilage ribosomal en condition de stress et se sont intéressés à savoir si

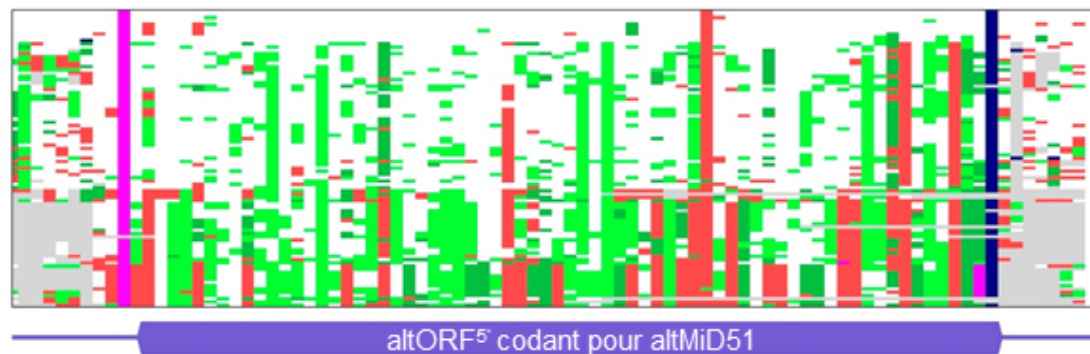


FIGURE 1.6 – Alignement multiple de l’altORF conservé d’altMiD51 provenant du génome de 100 vertébrés.

Adapté de (Andreev *et al.*, 2015). Les codons sont représentés par un rectangle coloré : rose=ATG et bleu=codon stop. Les autres couleurs représentent la substitution des acides aminés : blanc=aucune , vert pâle=synonyme, vert foncé=positive (BLOSUM62 > 0), rouge=négative (BLOSUM62 ≤ 0) et gris=délétion. Les régions vertes indiquent une évolution de la séquence protéique.

cet uORF régulateur de 70 codons potentiellement traduits était conservé (Andreev *et al.*, 2015). En effet, leur analyse de conservation à travers 100 vertébrés démontre qu’altMiD51 est un uORF conservé (Figure 1.6). Étant donné la conservation de cet uORF, ce même groupe de recherche suggère qu’altMiD51 est une protéine fonctionnelle (Andreev *et al.*, 2015).

1.5.4 La signature moléculaire d’altMiD51

Les signatures moléculaires contenues dans les protéines peuvent aider l’étude de la fonction d’une nouvelle protéine comme altMiD51. PFAM est une base de données regroupant des familles de protéines. Elle permet d’analyser les domaines dans la séquence protéique qu’on lui fournit par des alignements de séquence. En analysant la séquence d’altMiD51 avec PFAM, Benoît Vanderperre a déterminé qu’altMiD51 a un domaine protéique : le domaine LYR (Vanderperre, 2013). Tout comme Benoît Vanderperre, Andreev et collègues ont eux aussi analysé la séquence d’altMiD51 et ont trouvé qu’elle fait partie de la famille des protéines du complexe 1 (famille LYR) (Andreev *et al.*, 2015). Notamment, les protéines de cette famille sont retrouvées dans de gros complexes mitochondriaux comme les complexes de la chaîne respiratoire et le complexe de biogenèse des centres fer-soufre (Angerer, 2015). Plus de détails concernant le domaine LYR d’altMiD51 seront dans la discussion de mon mémoire.

1.6 La protéine de référence MiD51

1.6.1 La découverte du gène *SMCR7L*

En 1986, un groupe de recherche a trouvé une délétion de la région du chromosome 17p11.2 chez des patients ayant un syndrome inconnu qu'ils ont baptisé le syndrome Smith-Magenis (SMS) (Smith *et al.*, 1986). Ce syndrome est associé à un retard mental et des malformations congénitales (Smith *et al.*, 1986). Puis en 2002, l'équipe du Dr. James R. Lupski s'est intéressée à trouver les gènes responsables du syndrome Smith-Magenis (SMS) associé à cette délétion chromosomique (Bi *et al.*, 2002). Les gènes découverts par cette équipe de recherche ont été nommés *SMCR* pour « Smith-Magenis syndrome critical region » (Bi *et al.*, 2002). En fait, le gène *SMCR7* est codant pour une protéine transmembranaire de 353 acides aminés, *SMCR7*, dont aucun homologue a été trouvé chez l'humain. En 2003, *SMCR7* a été exclu des gènes responsables du syndrome (Slager *et al.*, 2003).

Entre temps, le 22e chromosome était en processus d'annotation et un certain ADNc a été attribué à une région de ce chromosome (Dunham *et al.*, 1999). En fait, cet ADNc encodait une protéine encore inconnue qui présente 45% d'homologie avec la protéine *SMCR7* (Palmer *et al.*, 2011b). La région codante du chromosome déterminée à partir de cet ADNc est devenue le gène *SMCR7L* pour (*Smith-Magenis syndrome candidate 7 protein-like* en anglais).

En 2000, Simpson et collègues ont eu comme objectif d'étudier une panoplie d'ADNcs à partir de cellules et de tissus humains (Simpson *et al.*, 2000). Ils ont voulu montrer les localisations cellulaires des protéines qui découlent de ces ADNcs (Simpson *et al.*, 2000). De plus, ils voulaient attribuer des informations bio-informatiques et une localisation cellulaire à ces potentielles nouvelles protéines en guise de piste vers une fonction et les regrouper dans une base de données : LIFEdb (Simpson *et al.*, 2000). Parmi les ADNcs étudiés se trouvaient celui qui encode MiD51, la protéine encore appelée *SMCR7L* comme le nom de son gène. Dans l'étude de Simpson et ses collègues, MiD51 fusionnée en c-terminal avec la protéine fluorescente GFP (*green fluorescent protein* en anglais), se localise à la mitochondrie (Figure 1.7). Il aura fallu attendre jusqu'en 2013 pour que les protéines découlant des gènes *SMCR7* et *SMCR7L* soient étudiées plus en profondeur. Celles-ci ont été rebaptisées MiD49 et MiD51 pour « Mitochondrial Dynamics proteins of 49/51 kDa » par Palmer et collègues. Ce groupe de recherche leur a attribué la fonction de récepteurs de la membrane externe des mitochondries impliqués dans la morphologie mitochondriale (Palmer *et al.*, 2011b). Dans ce mémoire, je parlerais uniquement en détail de MiD51.

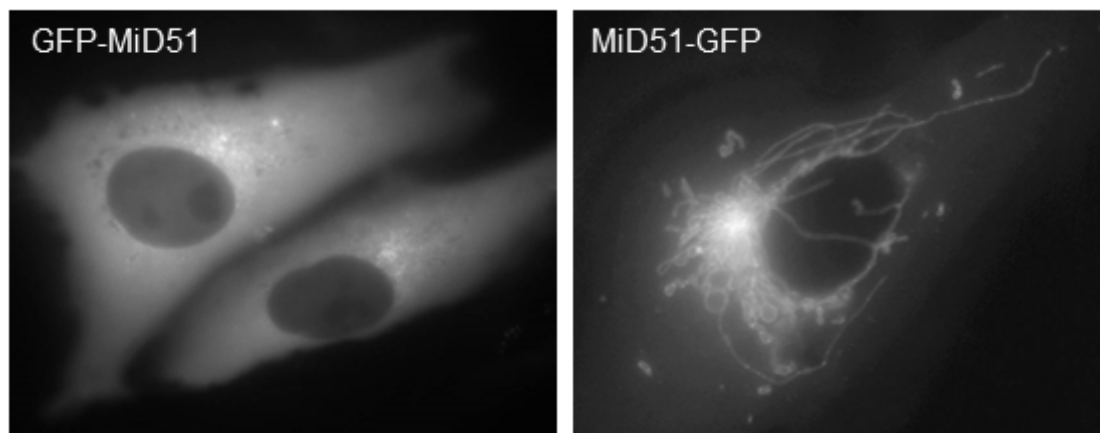


FIGURE 1.7 – Première image de localisation subcellulaire de MiD51 dans LIFEdb. Adapté de (Simpson *et al.*, 2000). Localisation cellulaire des fusions protéiques GFP-MiD51 et MiD51-GFP. La surexpression de GFP-MiD51 n'est pas bien localisée possiblement à cause de la présence du domaine d'ancrage à la membrane externe des mitochondries en N-terminal. La surexpression de MiD51-GFP se localise à la mitochondrie.

Pour bien comprendre la fonction de MiD51, il faut d'abord comprendre le concept de morphologie mitochondriale et les principales protéines impliquées dans ce mécanisme cellulaire.

1.6.2 La morphologie mitochondriale et les protéines associées

La morphologie mitochondriale est un processus dynamique et finement régulé qui s'adapte au besoin de la cellule (Figure 1.8A). La cellule contient des milliers de mitochondries et celles-ci se fusionnent les unes à la suite des autres puis se séparent. Plusieurs raisons peuvent influencer la morphologie mitochondriale, notamment : un stress, un problème dans la respiration cellulaire, une augmentation des «reactive oxygen species» (ROS) et/ou un changement métabolique (Liesa *et al.*, 2009). Les deux principaux phénomènes dynamiques sont la fusion et la fission des mitochondries (Figure 1.8B,C). Ces mécanismes sont exécutés majoritairement par des protéines de la famille des protéines GTPases, des protéines qui utilisent le GTP (guanosyl triphosphate) pour couper ou fusionner les mitochondries (Antonny *et al.*, 2016).

La fusion est nécessaire lorsque les mitochondries ont besoin de partager des ressources et se protéger d'un stress (Palmer *et al.*, 2011b). Pour connecter les mitochondries ensemble, il faut fusionner les membranes externes de deux mitochondries voisines pour terminer avec la fusion des membranes internes (Figure 1.8B). La membrane externe mitochondriale est fusionnée par l'entremise des mitofusine 1 et 2 et leur domaine GTPase

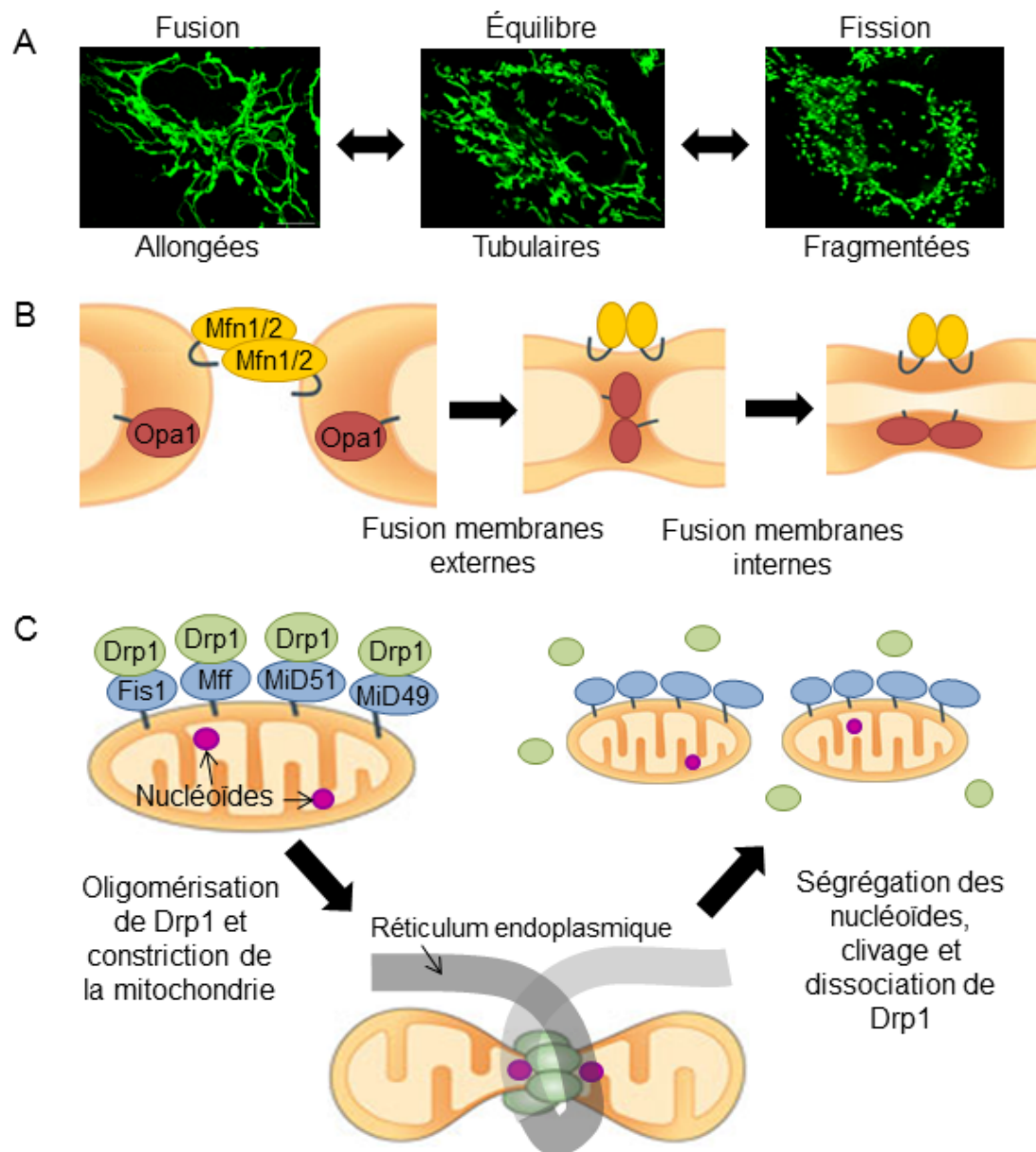


FIGURE 1.8 – Le dynamisme mitochondrial : régulation entre la fusion et la fission des mitochondries.

(A) Images de trois différentes morphologies observées dans les cellules HeLa. Lorsque la fusion domine la fission dans le dynamisme mitochondrial, les mitochondries apparaissent allongées et interconnectées (gauche). Quand la fission domine la fusion, les mitochondries sont très courtes ou fragmentées (image à droite). Lorsque le dynamisme mitochondrial est à l'équilibre, la morphologie apparente est tubulaire, c'est-à-dire qu'elle a un mélange de courtes et de plus longues mitochondries (centre). (B) Adapté de (Mishra et Chan, 2014). Schéma du mécanisme global de la fusion mitochondriale. (C) Adapté de (Mishra et Chan, 2014). Schéma du mécanisme global de fission mitochondriale par Drp1. Drp1 est recruté à partir du cytoplasme vers la mitochondrie. Il pourra ensuite oligomériser et former un anneau au site de fission.

(Chen *et al.*, 2003). Pour les membranes internes, la protéine responsable est Opa1 (Ishihara *et al.*, 2006). Opa1 est active sous forme longue, mais est sujette à être clivée par des protéases (Ishihara *et al.*, 2006). Sa forme courte ne peut pas fusionner les membranes (Ishihara *et al.*, 2006). Opa1 est également importante pour le remodelage des crêtes mitochondriales et facilite le relargage du cytochrome c lors de l'apoptose (MacVicar et Langer, 2016).

Par ailleurs, la fission mitochondriale est importante pour la mitose (Yamano et Youle, 2011). Cela permet de partager les mitochondries entre la cellule mère en division et la cellule fille. Récemment, il a été découvert que l'ADN mitochondrial en réplication se retrouve aux sites de fissions pour être séparé entre deux mitochondries, ce qui permet le partage d'ADN mitochondrial entre cellule mère et cellule fille (Lewis *et al.*, 2016) (Figure 1.8C). De plus, la fission mitochondriale peut s'enclencher pour se débarrasser des mitochondries défectueuses par la mitophagie (autophagie des mitochondries) (Kim *et al.*, 2007). Les mitochondries endommagées ont un potentiel membranaire diminué et ont moins d'Opa1 fonctionnel ce qui diminue leur capacité à fusionner (Twig *et al.*, 2008). Elles seront éliminées par les autophagosomes, des vésicules spécialisées à l'autophagie (Twig *et al.*, 2008). Aussi, la fission mitochondriale est importante pour le processus d'apoptose (Westermann, 2010). Les protéines pro-apoptotiques BAX et BCL-2 peuvent interagir avec Drp1, la protéine principale responsable de la fission mitochondriale (Westermann, 2010). La fission est possible grâce à la GTPase Drp1 recrutée à la surface des mitochondries par des récepteurs tels que Fis1, Mff, MiD51 et MiD49 (Losón *et al.*, 2013). Drp1 forme un anneau autour de la mitochondrie par oligomérisation et son activité GTPase clive la mitochondrie en deux (Smirnova *et al.*, 2001) (Figure 1.8C). Plusieurs modifications post-traductionnelles peuvent moduler son activité telles que la phosphorylation, l'ubiquitination, la sumoylation et la S-nitrosylation (Chang et Blackstone, 2010). Également, lors de la fission, le réticulum endoplasmique s'enroule autour des mitochondries pour serrer la mitochondrie aux sites de fission et faciliter le clivage par Drp1 (Friedman *et al.*, 2011; Elgass *et al.*, 2015) (Figure 1.8C). Finalement, MiD51 est un récepteur pour Drp1 et son implication dans le dynamisme mitochondrial est encore étudiée aujourd'hui.

1.6.3 MiD51 est impliquée dans la fission mitochondriale

Tout d'abord, il faut mentionner que la fonction de MiD51 est sujette à controverse. Est-ce que MiD51 est une protéine impliquée dans la fission ou la fusion mitochondriale ? En 2011, une première étude décrit MiD51 comme étant une protéine de 463 acides aminés

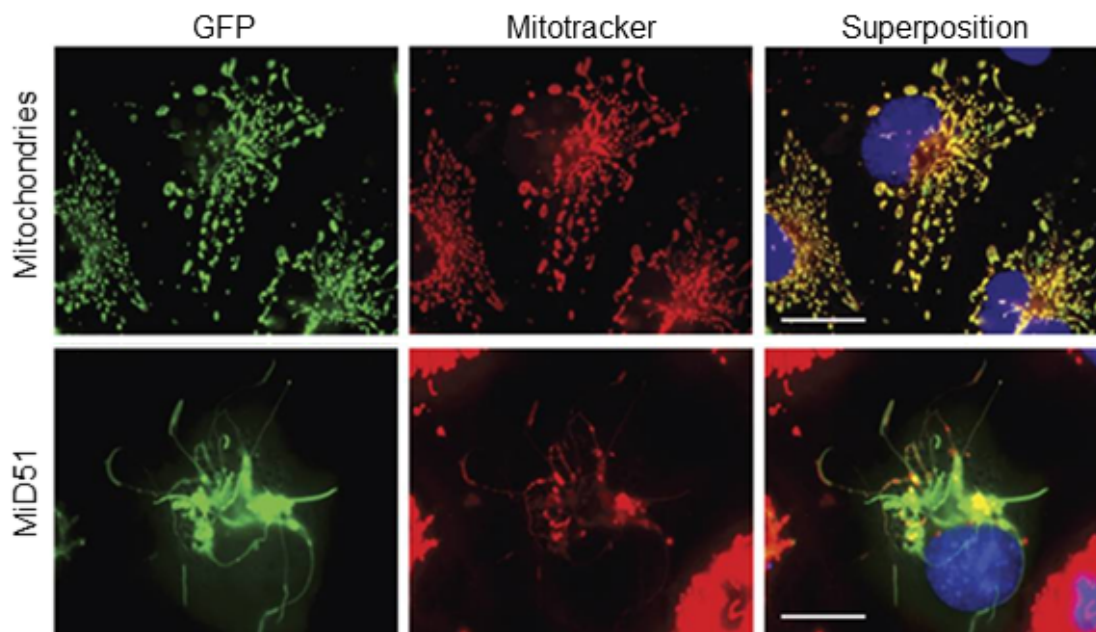


FIGURE 1.9 – La surexpression de MiD51 entraîne la fusion des mitochondries et une agglomération du réseau mitochondrial près du noyau.

Adapté de (Palmer *et al.*, 2011a). Des cellules COS7 ont été transfectées pour surexprimer MiD51-GFP ou une GFP avec un signal d'adressage à la mitochondrie (mito-GFP) (vert). Les mitochondries ont été marquées par le colorant MitoTracker (rouge) et le noyau par le colorant Hoechst (bleu). La surexpression de MiD51-GFP se localise à la mitochondrie et entraîne l'élongation des mitochondries et une agglomération des mitochondries autour du noyau. Le contrôle mitochondrie (mito-GFP) présente un réseau mitochondrial fragmenté.

avec un domaine d'ancrage membranaire en N-terminal de sa séquence (Palmer *et al.*, 2011a). Palmer et collègues ont déterminé que MiD51 est un récepteur à la surface mitochondriale qui recrute Drp1, un acteur clé dans la fission mitochondriale, aux sites de fissions (Palmer *et al.*, 2011a). D'une part, la surexpression de MiD51 allonge drastiquement les mitochondries et le réseau mitochondrial s'agglomère autour du noyau (Figure 1.9)(Palmer *et al.*, 2011a). Puisque MiD51 recrute Drp1 et que les mitochondries s'allongent plutôt que de fissionner, les auteurs suggèrent que MiD51 empêcherait le bon déroulement de l'assemblage de Drp1 pour exécuter la fission, mais que le mécanisme est encore inconnu (Palmer *et al.*, 2011a). D'autre part, l'inhibition de l'expression «knock-down» par des ARNs interférents de MiD51 diminue le recrutement de Drp1 à la mitochondrie et le réseau mitochondrial s'allonge également. Par conséquent, Palmer et collègues concluent que MiD51 est impliqué dans la fission mitochondriale en régulant Drp1 (Palmer *et al.*, 2011a).

Dans la même année que Palmer et collègues, un autre groupe de recherche s'est intéressé

au rôle de MiD51 dans la fission mitochondriale. Zhao et collègues ont aussi observé que la surexpression de MiD51 entraîne l'élongation et l'agrégation périnucléaire des mitochondries (Zhao *et al.*, 2011). Selon leur étude, MiD51 peut recruter Drp1 indépendamment des autres récepteurs connus (mff, fis1, MiD49) et entraîne la fusion mitochondriale en séquestrant Drp1 (Zhao *et al.*, 2011). En contradiction avec Palmer et collègues, l'inhibition de l'expression «knockdown» de MiD51 par ARNs interférents engendre la fragmentation des mitochondries plutôt que leur élongation (Zhao *et al.*, 2011). De plus, ils ont regardé l'effet de la surexpression de MiD51 sur des mitochondries fragmentées par l'inhibition par des ARNs interférents de Mfn2, une GTPase impliquée dans la fusion mitochondriale. La surexpression de MiD51 en plus de l'inhibition de Mfn2 entraîne l'allongement des mitochondries (Zhao *et al.*, 2011). Par conséquent, Zhao et ses collègues suggèrent que MiD51 est impliquée dans la fusion mitochondriale plutôt que la fission, car MiD51 peut entraîner la fusion des mitochondries malgré l'inhibition de Mfn2 (Zhao *et al.*, 2011).

En 2013, le rôle de MiD51 dans la fusion mitochondriale a été remis en question lorsque Palmer et collègues ont généré des cellules «knockout» pour les gènes codant pour les protéines Mfn1 et Mfn2, car toutes deux sont impliquées dans la fusion de la membrane externe des mitochondries (Palmer *et al.*, 2013). Cette fois-ci, la surexpression de MiD51 en absence des deux joueurs clés dans la fusion (Mfn1 et Mfn2) n'a pas provoqué l'élongation des mitochondries (Palmer *et al.*, 2013). Finalement, plusieurs études montrent que l'inhibition de MiD51 diminue le recrutement de Drp1 aux mitochondries et par le fait même diminue la fission mitochondriale (Palmer *et al.*, 2013; Koirala *et al.*, 2013; Otera *et al.*, 2013). Ces résultats supportent davantage le rôle de MiD51 dans la régulation de Drp1 dans la fission mitochondriale que dans la fusion.

Puisque Drp1 est recruté par MiD51, des études se sont intéressées à mieux comprendre le lien possible entre MiD51 et Drp1 dans la morphologie mitochondriale. Comme discuté précédemment, Drp1 peut subir plusieurs modifications post-traductionnelles qui modulent son activité (Chang and Blackstone 2010). Notamment, une phosphorylation à la serine 637 inhibe l'activité GTPase de Drp1 (Chang et Blackstone, 2007), tandis qu'une phosphorylation à la sérine 616 augmente son activité (Taguchi *et al.*, 2007). En 2011, Zhao et collègues ont démontré que le recrutement de Drp1 ne dépendait pas de son état de phosphorylation et que la surexpression de MiD51 diminue la liaison du GTP pour Drp1 (Zhao *et al.*, 2011). En fait, il a été démontré en 2013 que la surexpression de MiD51 entraîne le recrutement de Drp1 à la mitochondrie, mais qu'une majorité de ce Drp1 est phosphorylé à la serine 637, donc inactif (Losón *et al.*, 2013). Ce même groupe de recherche a utilisé un traitement au CCCP (Carbonyl cyanide m-chlorophenyl hydrazone),

un découplant qui perturbe le gradient électrochimique établi par la chaîne respiratoire et qui entraîne la fragmentation des mitochondries. Lors de ce traitement, la protéine Drp1 est déphosphorylée à la serine 637, donc Drp1 peut faire son activité de fission (Losón *et al.*, 2013). En combinant la surexpression de MiD51 avec le traitement CCCP, Losón et collègues ont obtenu une morphologie fragmentée (Losón *et al.*, 2013). Ainsi, lorsque MiD51 recrute Drp1 P-S637 et que le CCCP déphosphoryle Drp1, les mitochondries allongées par la surexpression de MiD51 sont clivées par la réactivation de Drp1 par le CCCP. Voilà pourquoi MiD51 participe au processus de fission plutôt que de fusion car il recrute le joueur clé de la fission mitochondriale, mais Drp1 est finement régulé pour induire la fission mitochondriale au moment opportun (Losón *et al.*, 2013).

L'élucidation de la structure 3D des protéines permet de comprendre les différents domaines d'une protéine et leur fonctionnement. En 2014, des études se sont penchées sur la caractérisation structurale de MiD51 pour comprendre le fonctionnement de la fission par le recrutement de Drp1. La première structure de MiD51 révèle qu'elle adopte un repliement caractéristique d'un site enzymatique de nucléotidyl transférase. Ce site enzymatique peut lier le GDP et l'ADP sans toutefois détenir d'activité catalytique permettant l'hydrolyse des nucléotides ou leur transfert (Richter *et al.*, 2014). La liaison des molécules au site enzymatique ne change pas la capacité de MiD51 à recruter Drp1 (Richter *et al.*, 2014). L'élément important dans la structure de MiD51 pour le recrutement de Drp1 est la formation d'une boucle par les acides aminés 210 à 258 de MiD51. Cette cristallisation de MiD51 délimite le site de recrutement pour Drp1 et les régions désordonnées (Richter *et al.*, 2014).

Un mois après la première cristallisation de MiD51 par Richter et ses collègues, un autre groupe de recherche travaillant sur MiD51 s'est intéressé aussi à sa structure 3D. Ils obtiennent les mêmes résultats sur la pochette nucléotidyl transférase et la boucle importante dans le recrutement de Drp1 dans la structure de MiD51 (Losón *et al.*, 2014). Par ailleurs, ils ont découvert que la pochette nucléotidyl transférase lie fortement l'ADP. Par contre, ils concluent que l'ADP n'est pas impliqué dans le recrutement de Drp1, mais que son absence empêche la fission par Drp1 (Losón *et al.*, 2014). Cette découverte de ce site de liaison sur MiD51 a permis de comprendre que l'inhibition de la fission lors du recrutement de Drp1 est régulée par la présence d'ADP (Losón *et al.*, 2014). L'ADP ôte l'inhibition de MiD51 sur l'oligomérisation et l'activité GTPase de Drp1 en plus d'augmenter son activité catalytique (Losón *et al.*, 2014). De manière intéressante, Losón et collègues en 2014 démontrent par des mutants de MiD51 que la dimérisation de MiD51 est possible par une liaison électrostatique en N-terminal qui n'est pas nécessaire au re-

crutement de Drp1. Toutefois, cette liaison est nécessaire, avec l'aide du cofacteur ADP, pour l'oligomérisation et pour l'activité de fission de Drp1 (Losón *et al.*, 2014).

En résumé, MiD51 est un récepteur pour Drp1 et est impliquée dans la fission mitochondriale. Elle possède un site de liaison à l'ADP un peu différent du site habituel nucléotidyl transférase liant les nucléotides triphosphate retrouvés dans d'autres protéines. Par son cofacteur ADP et sa dimérisation, celle-ci peut recruter et réguler l'activité de fragmentation de la GTPase Drp1 qui est essentielle à l'homéostasie des mitochondries.

1.7 Hypothèse de recherche : Deux protéines, encodées par un même ARNm, impliquées dans la fission mitochondriale

Dans un premier temps, les outils bio-informatiques et expérimentaux décrits dans ce mémoire comme le motif Kozak, le profilage ribosomal, la spectrométrie de masse, les études de conservations et les études de signatures moléculaires nous permettent d'identifier les protéines alternatives ayant des indices de fonctionnalité.

Ainsi, la protéine altMiD51 possède toutes les indications énumérées ci-dessus, à l'exception près du motif Kozak – toutefois présent seulement pour 37% des protéines de référence. Notamment, elle a été détectée dans plusieurs analyses de spectrométrie de masse, et dans plusieurs études de profilage ribosomal. Elle est conservée chez les vertébrés et possède le domaine LYR.

Par ailleurs, certaines protéines alternatives sont impliquées dans des processus cellulaires importants, parfois même en lien avec la fonction de leur protéine de référence. C'est le cas de XL α s/Alex, A2AR/uORF5 et de rAT1aR/Pep7, comme discuté précédemment (section 1.2 à 1.4). Toutefois, la coopération fonctionnelle entre protéines provenant d'un même ARNm demeure un mécanisme cellulaire à explorer. Or, les résultats préliminaires apportés par Benoît Vanderperre montrent une localisation mitochondriale pour altMiD51, tout comme sa protéine de référence, MiD51 (Vanderperre, 2013). Une observation qui pourrait suggérer une coopération entre les deux protéines codées par l'ARNm du gène *SMCR7L*.

De fait, ce mémoire s'intéresse dans un premier temps à la caractérisation de cette nouvelle protéine, altMiD51 ; puis à son éventuelle coopération fonctionnelle avec MiD51 dans la régulation de la dynamique mitochondriale, notamment la fission.

Tout d'abord, altMiD51 servira de preuve de concept quant à l'existence d'ARNms multi-codants chez les Eucaryotes. Pour ce faire, sa détection, son expression, et sa localisation seront démontrées.

Puis, l'importance d'altMiD51 dans la régulation de la dynamique mitochondriale sera mise en lumière. Notamment, son implication dans la morphologie mitochondriale et sa relation avec un acteur majeur de la fission, Drp1.

2 ARTICLE

Deep transcriptome annotation suggests cooperation between small and large proteins coded by the same genes

Auteurs de l'article: Sondos Samandi, Annie V. Roy, Vivian Delcourt, Jean-François Lucier, Jules Gagnon, Maxime C. Beaudoin, Benoît Vanderperre, Marc-André Breton, Julie Motard, Jean-François Jacques, Isabelle Gagnon-Arsenault, Isabelle Fournier, Aida Ouangaoua, Darel J. Hunting, Alan A. Cohen, Christian R. Landry, Michelle S. Scott, Xavier Roucou

Statut de l'article: Soumis dans Nature Communications

Avant-propos: J'ai planifié et réalisé la majorité des expériences biologiques dans l'article. Cela inclue la figure 3 a,c,d et les figures supplémentaires 17, 18, 19 et 20. J'ai contribué à l'écriture de l'article pour ces expériences (matériel et méthodes, figures et les légendes des figures, résultats et discussion).

Résumé: Les expériences fonctionnelles, protéomiques et de profilage ribosomal effectuées chez les eucaryotes ont démontré la traduction de cadres de lecture ouverts alternatifs (altORFs) en plus de la séquence codante (CDS) déjà existante dans les ARNm. Nous montrons que chez les eucaryotes, la majorité des protéines alternatives traduites à partir d'altORFs ont des orthologues dans plusieurs espèces. Des milliers de protéines alternatives sont détectées lorsque nous analysons des données publiques avec une base de données contenant les séquences des protéines alternatives. Les patrons de conservation entre le CDS et les altORFs du même ARNm indiquent qu'ils ont évolué ensemble. Nous avons identifié une nouvelle protéine mitochondriale de 70 acides aminés impliquée dans la fission mitochondriale traduite à partir d'un altORF dans le gène *MiD51/Mief1/SMCR7L*. Ce gène code une protéine annotée de 51kDa, MiD51, impliquée dans la fission mitochondriale. Les domaines protéiques connus retrouvés dans les protéines alternatives suggèrent une coopération possible avec les protéines déjà annotées. Nos résultats appuient une coopération fonctionnelle entre les protéines alternatives et celles annotées dans un même transcrit. Cela représente une couche supplémentaire de la régulation génique.

Deep transcriptome annotation suggests cooperation between small and large proteins coded by the same genes

Sondos Samandi^{1,7} †, Annie V. Roy^{1,7} †, Vivian Delcourt^{1,7,8}, Jean-François Lucier², Jules Gagnon², Maxime C. Beaudoin^{1,7}, Benoît Vanderperre¹, Marc-André Breton¹, Julie Motard^{1,7}, Jean-François Jacques^{1,7}, Isabelle Gagnon-Arsenault^{6,7}, Isabelle Fournier⁸, Aida Ouangraoua³, Darel J. Hunting⁴, Alan A. Cohen⁵, Christian R. Landry^{6,7}, Michelle S. Scott¹, Xavier Roucou^{1,7}*

¹Department of Biochemistry, ²Department of Biology and Center for Computational Science, ³Department of Computer Science, ⁴Department of Nuclear Medicine & Radiobiology, ⁵Department of Family Medicine, Université de Sherbrooke, Quebec, Canada; ⁶Department of Biology, Université Laval, Quebec, Canada; ⁷PROTEO, Quebec Network for Research on Protein Function, Structure, and Engineering, Quebec, Canada; ⁸Univ. Lille, INSERM U1192, Laboratoire Protéomique, Réponse Inflammatoire & Spectrométrie de Masse (PRISM) F-59000 Lille, France

†These authors contributed equally to this work

*Correspondance to Xavier Roucou: Department of Biochemistry (Z8-2001), Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, 3201 Jean Mignault, Sherbrooke, Quebec J1E 4K8, Canada, Tel. (819) 821-8000x72240; Fax. (819) 820 6831; E-Mail: xavier.roucou@usherbrooke.ca

Abstract

Functional, proteomic and ribosome profiling approaches in eukaryotes have concurrently demonstrated the translation of alternative open reading frames (altORFs) in addition to annotated protein coding sequences (CDSs). We show that the majority of small eukaryotic proteins are coded by altORFs and have orthologs in many species. Several thousand alternative proteins can be detected in publicly available large proteomic datasets by simple reanalysis with a database containing the sequence of alternative proteins. Conservation patterns of altORFs and their CDSs indicate that many are paired evolutionary units. We identify a new 70 amino acid mitochondrial fission factor encoded by an altORF in *MiD51/Mief1/SMCR7L*, a gene encoding an annotated 51kDa protein promoting mitochondrial fission. Known domains detected in alternative proteins suggest functional coupling with large proteins. Our results support a functional cooperation between small and large proteins encoded in the same transcripts, representing a new layer of regulation in genome function.

Introduction

Current protein databases are cornerstones of modern biology but are based on a number of assumptions. In particular, a mature mRNA is assumed to contain a single CDS; yet, ribosomes can select more than one translation initiation site (TIS)¹⁻³. Also, minimum size limits are imposed on the length of potential ORFs, resulting in many RNAs being mistakenly classified as non-coding (ncRNAs)⁴⁻¹¹. As a result of these assumptions, the size and complexity of most eukaryotic proteomes have probably been greatly underestimated¹²⁻¹⁵. In particular, few small proteins, defined here as proteins of 100 amino acids or less are annotated in current databases. The absence of annotation of small proteins is a major bottleneck in the study of their function, particularly since many critical regulatory molecules, including F0 subunits of the F0F1-ATP synthase¹⁶, the sarcoplasmic reticulum calcium ATPase regulator phospholamban¹⁷, and the key regulator of iron homeostasis hepcidin¹⁸, are small proteins.

Functional annotation of ORFs encoding small proteins is particularly challenging since an unknown fraction of small ORFs may occur by chance in the transcriptome, generating a significant level of noise. However, given that many small proteins have important functions, it is time to address the challenge of their functional annotations¹³.

We have systematically reanalyzed several eukaryotic transcriptomes to annotate alternative ORFs (altORFs), and we have annotated the corresponding hidden proteome. Here, altORFs are defined as potential protein-coding ORFs exterior to, or in different reading frames from annotated CDSs in mRNAs and ncRNAs (Fig. 1a). For clarity, proteins translated from altORFs are termed alternative proteins and proteins translated from annotated CDSs are termed reference proteins.

Here, our goal was to provide functional annotations of alternative proteins by (1) characterizing the relative size distributions of alternative and reference proteins; (2) analyzing relative patterns of evolutionary conservation between alternative and reference proteins and their coding sequences; (3) detection of functional signatures in alternative proteins; and (4) testing functional coupling between an alternative protein (altMiD51) and its reference protein (MiD51).

Results

Prediction and evidence of expression of alternative proteins.

We predicted a total of 551,380 altORFs with a size ranging from 30 codons to 1480, compared to 68,066 annotated CDSs in the human transcriptome (Fig. 1B, Supplementary Table 1). Identical ORFs can be present in different RNA isoforms, and thus the number of unique altORFs and CDSs were 183,191 and 51,818, respectively. AltORFs were also predicted in other organisms (Supplementary Table 1). By convention, only reference proteins are annotated in current protein databases. The median size for human alternative proteins is 45 amino acids compared to 460 for reference proteins (Supplementary Fig. 1), and 92.96 % of alternative proteins have less than 100 amino acids. Thus, the bulk of the translation products of altORFs would be small proteins. The majority of altORFs either overlap annotated CDSs in a different reading frame (35.98%), or are located in 3'UTRs (40.9%) (Fig. 1C). Only about 10% of altORFs are located in repeat sequences (Supplementary Fig. 2). To assess whether observed altORFs might be attributable solely to random occurrence, we determined the number of altORFs generated in 100 shuffled human transcriptomes.

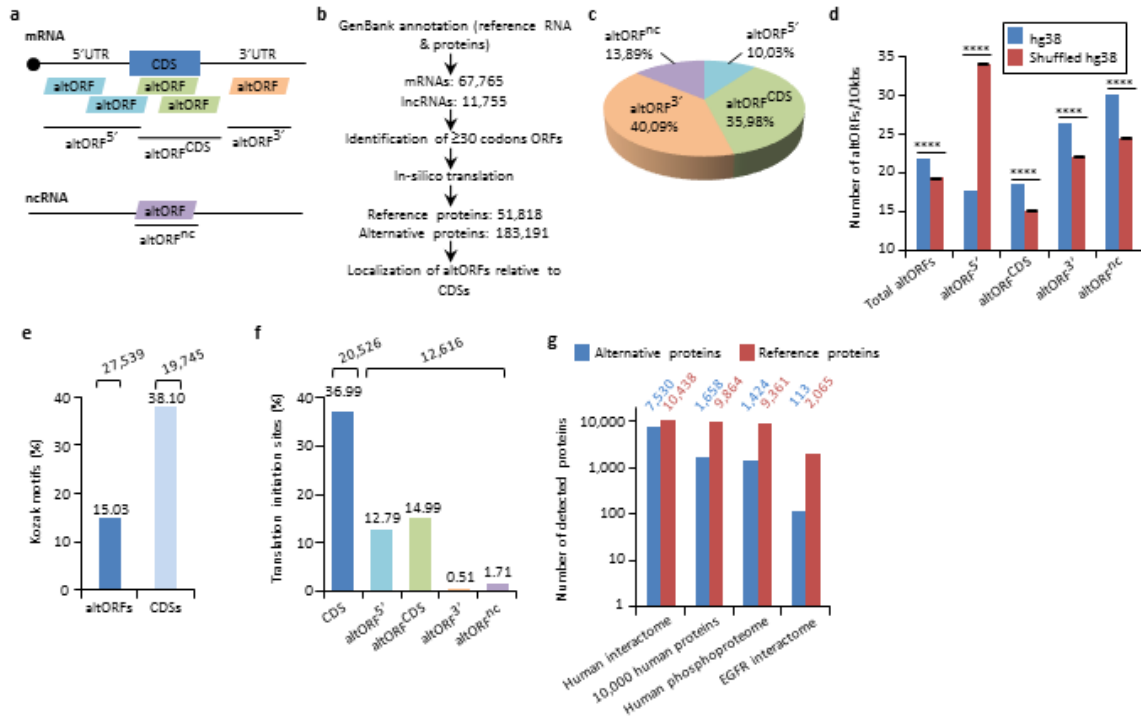


Fig. 1. Annotation and expression of human altORFs.

(a) AltORF nomenclature. AltORFs partially overlapping the CDS must be in a different reading frame. (b) Pipeline for the identification of altORFs. (c) Distribution of altORFs in the human hg38 transcriptome. (d) Number of total altORFs in hg38 compared to shuffled hg38. Means and standard deviations for 100 replicates obtained by sequence shuffling are shown. Statistical significance was determined by using one sample t-test with two-tailed p -values. **** $P < 0.0001$. (e) Number of CDSs and altORFs with an optimal Kozak motif. (f) Percentage of detected TISs in CDSs and altORFs by ribosomal profiling and footprinting of human cells²³. The total number of CDSs and altORFs with a detected TIS is indicated at the top. (g) Alternative and reference proteins detected in three large proteomic datasets: human interactome²⁷, 10,000 human proteins³⁰, human phosphoproteome²⁹, EGFR interactome²⁸. Numbers are indicated above each column.

Overall, we observed 62,307 more altORFs than would be expected from random occurrence (p -value < 0.0001 ; Supplementary Fig. 3). The density of altORFs observed in the CDSs, 3'UTRs and ncRNAs (Fig. 1d) was markedly higher than in the shuffled transcriptomes, suggesting a positive selection for altORFs in these sequences. In contrast, the density of altORFs observed in 5'UTRs was much lower than in the shuffled transcriptomes, confirming a negative selection^{19,20}. Although the majority of human annotated CDSs do not have a TIS with a Kozak motif (Fig. 1e)²¹, there is a correlation between a Kozak motif and

translation efficiency²². Around 15% of human altORFs encoding 27,539 alternative proteins have a Kozak motif, as compared to 38% for annotated CDSs encoding 19,745 reference proteins (Fig. 1e). The number of altORFs with kozak motifs is significantly higher in the human transcriptome compared to shuffled transcriptomes (Supplementary Fig. 4).

Two lines of evidence indicate that several thousand altORFs are translated into proteins. First, we re-analyzed detected TISs in publicly available ribosome profiling data^{23,24}, and found 20,526 TISs mapping to annotated CDSs and 12,616 mapping to altORFs (Fig. 1f, Supplementary Data 1). Although predicted altORFs^{3'} are more abundant than altORFs^{5'}, only a small fraction of TISs detected by ribosomal profiling mapped to altORFs^{3'}. This observation likely reflects a technical limitation of the method, which appears to have been resolved only recently²⁵. New methods to analyze ribosome profiling data are being developed and will likely uncover more translated altORFs²⁶. Second, we re-analyzed proteomic data using our composite database containing alternative proteins in addition to annotated reference proteins (Fig. 1g, Supplementary Data 2). We selected four studies representing different experimental paradigms and proteomic applications: large-scale²⁷ and targeted²⁸ protein/protein interactions, post-translational modifications²⁹, and a combination of targeted and non-targeted proteomics³⁰. We detected 7,530 predicted alternative proteins in the interactome of reference proteins²⁷, providing a framework to uncover the function of these proteins. A total of 1,658 alternative proteins were detected in a proteomic dataset containing about 10,000 reference human proteins³⁰. Using a phosphoproteomic large data set²⁹, we detected 1,424 alternative proteins. The biological function of these proteins is supported by the observation that some alternative proteins are specifically phosphorylated in cells stimulated by the epidermal growth factor, and others are specifically phosphorylated

during mitosis (Supplementary Fig 5-7, Supplementary Data 3). A fourth proteomic dataset contained 113 alternative proteins in the epidermal growth factor receptor interactome²⁸. A total of 10,362 different alternative proteins were detected in these proteomic data. Overall, by mining the proteomic and ribosomal profiling data, we detected the expression of 22,155 different alternative proteins and the expression of 823 alternative proteins was detected by both MS and ribosome profiling (Supplementary Fig. 8).

Conservation analyses. Next, we compared evolutionary conservation patterns of altORFs and CDSs. A large number of human alternative proteins have homologs in other species. In mammals, the number of homologous alternative proteins is higher than the number of homologous reference proteins (Fig. 2a), and a few are even conserved from humans to yeast, supporting a functional role (Supplementary Fig. 9a). As phylogenetic distance from humans increases, the number and percentage of genes encoding homologous alternative proteins decreases more rapidly than the percentage of genes encoding reference proteins (Supplementary Fig. 9b). This observation indicates either that altORFs have evolved more rapidly than CDSs or that distant homologies are less likely to be detected given the smaller sizes of alternative proteins. The first and second bases of a codon experience stronger selection (i.e., greater conservation or acceleration) than the third³¹. Thus, CDS regions overlapping altORFs, with a shifted reading frame, should be more selected at the third wobble nucleotide if the altORFs are functional since the wobble nucleotide in the CDS corresponds to the first or second nucleotides in the overlapping altORF.

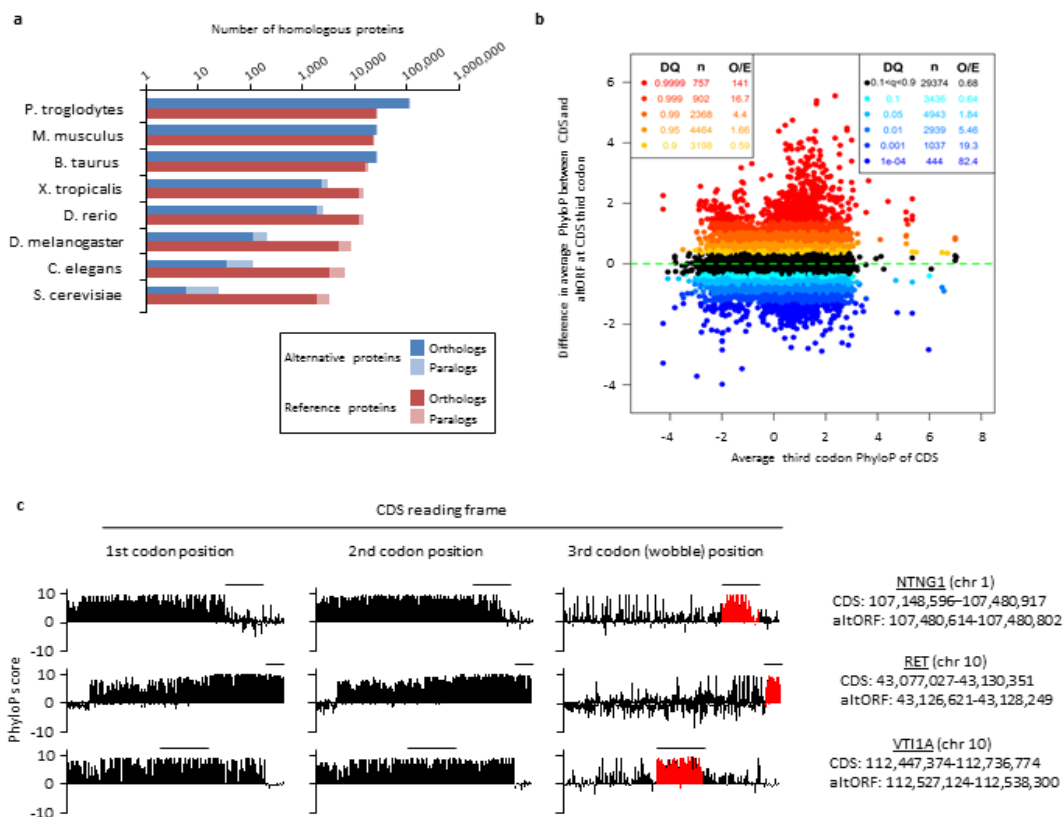


Fig. 2. Conservation analyses of alternative proteins and altORFs.

(a) Number of orthologous and paralogous alternative and reference proteins between *H. sapiens* and other species (pairwise study). (b) AltORFs show more extreme PhyloP values (more conserved or faster evolving) than their CDSs. Differences between altORF and CDS PhyloP scores (altORF PhyloP – CDS PhyloP, y -axis) are plotted against PhyloPs for their respective CDSs (x -axis). The plot contains all 20,814 CDSs containing at least one fully nested altORF, paired with one of its altORFs selected at random (to avoid problems with statistical non-independence). PhyloPs for both altORFs and CDSs are based on 3rd codons in the CDS reading frame, calculated across 100 vertebrate species. We compared these differences to those generated based on five shuffled versions of the PhyloP scores. Expected quantiles of the differences (“DQ” columns) were identified and compared to the observed differences. We show the absolute numbers (“n”) and observed-to-expected ratios (“O/E”) for each quantile. There are clearly substantial over-representations of extreme values (red signalling conservation and blue signalling accelerated evolution), more so as the values become more extreme. Conserved altORFs are more over-represented than altORFs under fast evolution, as apparent visually, in the table, and in the positive skewness (0.996) of the adjusted difference between altORF and CDS PhyloPs. (c) First, second and third codon nucleotides PhyloP scores for 100 vertebrate species for the CDS of the NTNG1, RET and VTI1A genes. Chromosomal coordinates for the different CDSs and altORFs are indicated on the left. The regions highlighted in red indicate the presence of an altORF characterized by a region with elevated PhyloP scores for wobble nucleotides. The region of the altORF is indicated by a black bar above each graph.

We analyzed conservation/acceleration at third nucleotides of CDSs for 100 vertebrate species for the 53,862 altORFs completely nested within the 20,814 CDSs from 14,677 genes. We observed that in regions coding for altORFs wobble nucleotides of the CDS were more selected than wobble nucleotides in the entire CDS when compared to shuffled controls (Fig. 2b), e.g. 141-fold for conservation at $p < 0.0001$. This is illustrated with three altORFs located within the CDS of NTNG1, RET and VTI1A genes (Fig. 2c, Supplementary Data 2). These three genes encode a protein promoting neurite outgrowth, the proto-oncogene tyrosine-protein kinase receptor Ret and a protein mediating vesicle transport to the cell surface, respectively. The phyloP scores for wobble nucleotides in the CDS region overlapping these 3 altORFs are remarkably elevated compared to the remaining region of the CDS. Additionally, the slope of the relationship between the PhyloP difference and CDS PhyloPs is greater than 0, indicating that as a CDS comes under fast evolution or conservation, its altORF comes under even faster evolution or stronger conservation (Fig. 2b). This slope gets greater and greater as the analysis is restricted to more extreme quantiles: 0.019 for all points ($p < 0.0001$), 0.031 excluding the middle 80% ($p < 0.0001$), 0.080 excluding the middle 99.8% ($p = 0.0017$), and 0.152 excluding the middle 99.98% ($p < 0.0004$).

Functional annotations of alternative proteins. An important goal of this study was to associate functional annotations to alternative proteins. Because the sequence similarities and the presence of particular signatures (families, domains, motifs, sites) are a good indicator of a protein's function, we analyzed the sequence of alternative proteins in several organisms with InterProScan, a powerful analysis and classification tool for characterizing

unknown protein sequences by predicting the presence of combined protein signatures from most main domain databases³² in several organisms (Supplementary Figs. 10-13). We found 41,511 (23%) human alternative proteins with at least one InterPro signature (Supplementary Fig. 10b). Of these, 37,739 (or 20.6%) are classified as small proteins. Interestingly, the reference proteome has only 840 (or 1.6%) small proteins with at least one InterPro signature. Similar to reference proteins, signatures linked to membrane proteins are abundant in the alternative proteome and represent more than 15,000 proteins (Supplementary Fig. 10c-e). With respect to the targeting of proteins to the secretory pathway or to cellular membranes, the main difference between the alternative and the reference proteomes lies in the very low number of proteins with both signal peptides and transmembrane domains. Most of the alternative proteins with a signal peptide do not have a transmembrane segment and are predicted to be secreted (Supplementary Fig. 10c, d), supporting the presence of large numbers of alternative proteins in plasma³³. The majority of alternative proteins with transmembrane domains have a single membrane spanning domain but some display up to 27 transmembrane regions, compared to a maximum of 33 for reference proteins (Supplementary Fig. 10e).

A total of 585 alternative proteins were assigned 419 different InterPro entries, and 343 of them were assigned 192 gene ontology terms (Supplementary Fig. 12). The most abundant class of alternative proteins with at least one InterPro entry are C2H2 zinc finger proteins with 110 alternative proteins containing 187 C2H2-type/integrase DNA-binding domains, 91 C2H2 domains and 23 C2H2-like domains (Supplementary Fig. 10f). Seventeen of these were detected in public proteomic and ribosome profiling datasets (Fig. 1f, g, Supplementary Table 2). Alternative proteins have between 1 and 23 zinc finger domains (Supplementary

Fig. 13). Zinc fingers mediate protein-DNA, protein-RNA and protein-protein interactions³⁴. The linker sequence separating adjacent finger motifs matches or resembles the consensus TGEK sequence in nearly half the annotated zinc finger proteins³⁵. This linker confers high affinity DNA binding and switches from a flexible to a rigid conformation to stabilize DNA binding. The consensus TGEK linker is present 46 times in 31 alternative zinc finger proteins (Supplementary Data 4).

Evidence of functional coupling between reference and alternative proteins coded by the same genes. In order to test if alternative proteins have a biological function, we investigated the activity of altMiD51. AltMiD51 is a 70 amino acid alternative protein conserved in vertebrates³⁶ and robustly detected by mass spectrometry (Supplementary Data 2: HEK293, HeLa, HeLa S3, LNCaP, NCI60 and U2OS cells; Supplementary Fig. 14) and ribosome profiling (Supplementary Data 1)^{33,36,37}. Its coding sequence is present in exon 2 of the *MiD51/MIEF1/SMCR7L* gene. This exon forms part of the 5'UTR for the canonical mRNA and is annotated as non-coding in current gene databases (Fig. 3a). Yet, we confirmed co-expression of altMiD51 and MiD51 from the same transcript (Fig. 3b). Importantly, the tripeptide LYR motif predicted with InterProScan and located in the N-terminal domain of altMiD51 (Fig. 3a), is a signature of mitochondrial proteins localized in the mitochondrial matrix³⁸. Since *MiD51/MIEF1/SMCR7L* encodes a mitochondrial protein, MiD51 which promotes mitochondrial fission by recruiting cytosolic Drp1, a

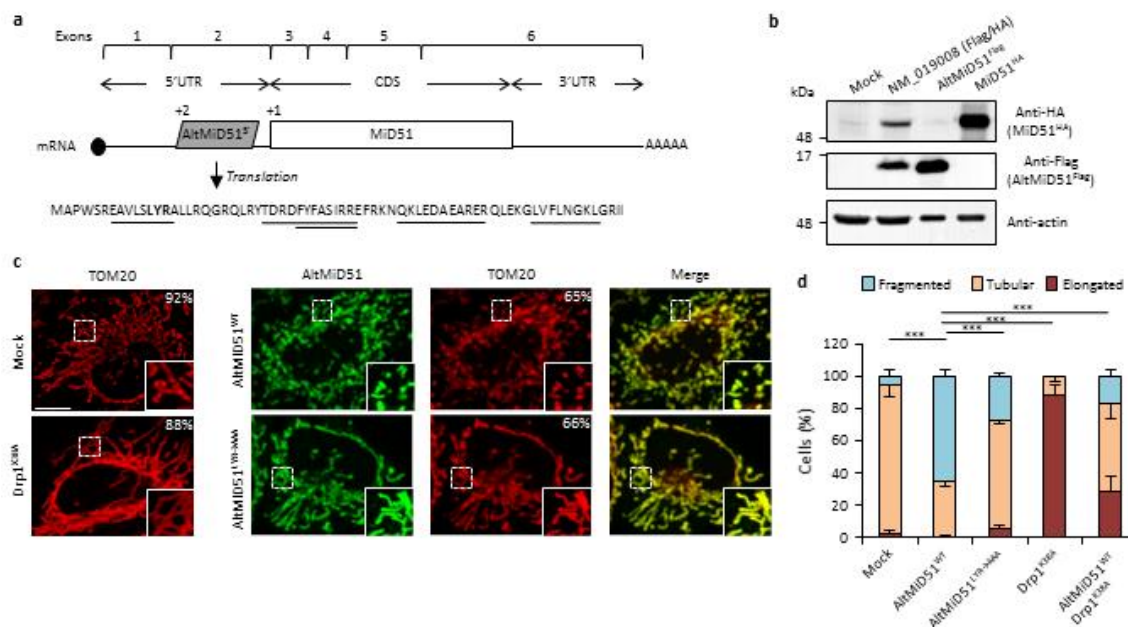


Fig. 3. AltMiD51^{5'} is a novel mitochondrial fission factor.

(a) AltMiD51^{5'} coding sequence is located in exon 2 of the MiD51/Mief1/SMCR7L gene and in the 5'UTR of the canonical mRNA (RefSeq NM_019008). +2 and +1 indicate reading frames. AltMiD51 amino acid sequence (LYR tripeptide is shown in bold). Underlined peptides were detected by MS (Fig. 1g) (b) Human HeLa cells transfected with empty vector (mock), a cDNA corresponding to the canonical MiD51 transcript with a Flag tag in frame with altMiD51 and an HA tag in frame with MiD51, altMiD51^{Flag} cDNA or MiD51^{HA} cDNA were lysed and analyzed by western blot with antibodies against Flag, HA or actin, as indicated. (c) Confocal microscopy of mock-transfected cells, cells transfected with altMiD51^{WT}, altMiD51^{LYR→AAA} or Drp1^{K38A} immunostained with anti-TOM20 (red channel) and anti-Flag (green channel) monoclonal antibodies. In each image, boxed areas are shown at higher magnification in the bottom right corner. % of cells with the most frequent morphology is indicated: mock (tubular), altMiD51^{WT} (fragmented), altMiD51(LYR→AAA) (tubular), Drp1(K38A) (elongated). Scale bar, 10 μ m. (e) Bar graphs show mitochondrial morphologies in HeLa cells. Means of three independent experiments per condition are shown. *** p <0.0005 (Fisher's exact test) for the three morphologies between altMiD51(WT) and the other experimental conditions.

member of the dynamin family of large GTPases, to mitochondria³⁹, we tested for a possible functional connection between these two proteins expressed from the same mRNA. We confirmed that MiD51 induces mitochondrial fission (Supplementary Fig. 15). Remarkably, altMiD51 localizes at the mitochondria (Fig. 3c, Supplementary Fig. 16) and its overexpression results in mitochondrial fission (Fig. 3c, d). This activity is unlikely to be

through perturbation of oxidative phosphorylation since altMiD51 overexpression did not change oxygen consumption, ATP and reactive oxygen species production (Supplementary Fig. 17). A significant decrease in spare respiratory capacity in altMiD51-expressing cells likely resulted from mitochondrial fission⁴⁰. A mutant of the LYR domain was unable to convert the mitochondrial morphology from tubular to fragmented. Drp1(K38A), a dominant negative mutant of Drp1⁴¹, largely prevented the ability of altMiD51 to induce mitochondrial fragmentation (Fig. 3d, Supplementary Fig. 18a). In a control experiment, co-expression of wild-type Drp1 and altMiD51 proteins resulted in mitochondrial fragmentation (Supplementary Fig. 18b). Expression of the different constructs used in these experiments was verified by western blot (Supplementary Fig. 19). Similarly, Drp1 knockdown interfered with altMiD51-induced mitochondrial fragmentation (Supplementary Fig. 20). It remains possible that altMiD51 promotes mitochondrial fission independently of Drp1 and is able to reverse the hyperfusion induced by Drp1 inactivation. However, Drp1 is the key player mediating mitochondrial fission and most likely mediates altMiD51-induced mitochondrial fragmentation.

Although altMiD51 and MiD51 do not share InterPro entries and could not have been predicted to function in the same biological pathway, these two proteins encoded in the same gene both localize in the mitochondria and promote mitochondrial fission. There are other examples of functional interactions between different proteins encoded in the same gene (Supplementary Table 3). In order to further test the hypothesis of functional interactions between reference and alternative proteins, we compared the functional annotations of the 585 alternative proteins with an InterPro entry with the reference proteins expressed from the same genes. Strikingly, 89 of 110 altORFs coding for zinc finger proteins (Supplementary

Fig. 10f) are present in transcripts in which the CDS also codes for a zinc finger protein. Overall, 138 alternative/reference protein pairs share at least one InterPro entry and many pairs share more than one entry (Fig. 4a).

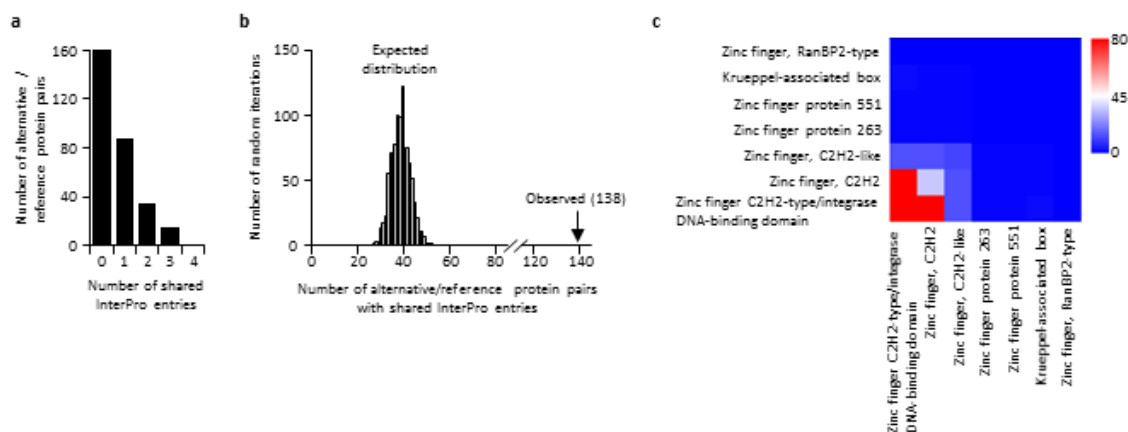


Fig. 4. Reference and alternative proteins share functional domains.

(a) Distribution of the number of shared InterPro entries between alternative and reference proteins coded by the same transcripts. 138 pairs of alternative and reference proteins share between 1 and 4 protein domains (InterPro entries). Only alternative/reference protein pairs that have at least one domain are considered ($n = 298$). (b) The number of reference/alternative protein pairs that share domains ($n = 138$) is higher than expected by chance alone. The distribution of expected pairs sharing domains and the observed number are shown. (c) Matrix of co-occurrence of domains related to zinc fingers. The entries correspond to the number of times entries co-occur in reference and alternative proteins. The full matrix is available in fig. S21.

The number of shared entries was much higher than expected by chance (Fig. 4b, $p < 0.0001$).

The correspondence between InterPro domains of alternative proteins and their corresponding reference proteins coded by the same transcripts also indicates that even when entries are not identical, the InterPro terms are functionally related (Fig. 4c, Supplementary Fig. 21).

Discussion

In light of the increasing evidence from approaches such as ribosome profiling and MS-based proteomics that the one mRNA-one canonical CDS assumption is outdated, our findings provide the first clear functional insight into a new layer of regulation in genome function. While many observed altORFs may be evolutionary accidents with no functional role, at least 8 independent lines of evidence support a functional role for thousands of alternative proteins: (1) overrepresentation of altORFs relative to shuffled sequences; (2) overrepresentation of altORF kozak sequences; (3) active altORF translation as detected via ribosomal profiling; (4) detection of thousand alternative proteins in multiple existing proteomic databases; (5) correlated altORF-CDS conservation, but with overrepresentation of highly conserved and fast-evolving altORFs; (6) underrepresentation of altORFs in repeat sequences; (7) overrepresentation of identical InterPro signatures between alternative and reference proteins encoded in the mRNA; and (8) presence of a clear, striking example in altMiD51, which plays a role in mitochondrial fission as does its reference protein. While 5 of these 8 lines of evidence support an unspecified functional altORF role, three of them (5, 7, and 8) independently support a specific functional/evolutionary interpretation of their role: that alternative proteins and reference proteins have paired functions. This makes sense: co-localization on a single mRNA is an efficient way to ensure proportional translation. Given that the regulation of protein function is known to evolve more rapidly than protein structure⁴², a reasonable interpretation is that altORF- and CDS-coded proteins often have related functions, but with alternative proteins often playing a regulatory role and thus responding more strongly to selective pressures than reference proteins.

Upstream ORFs here labeled altORFs^{5'} are important translational regulators of canonical

CDSs in vertebrates⁴³. The identification of altMiD51 as a new mitochondrial fission factor suggests that a fraction of altORFs⁵⁷ may have dual functions as translation regulators and functional proteins. Alternatively, altORFs⁵⁷ may belong to two independent classes, translational regulators and protein coding sequences for proteins with specific functions.

Detection is an important challenge in the study of small proteins. A translation initiation site detected by ribosome profiling does not necessarily imply that the protein is expressed as a stable molecule, and proteomic analyses more readily detect large proteins that generate several peptides after enzymatic digestion. Here, we have used a combination of five search engines and false discovery rate cut-offs were set at 1% for peptide-spectrum match, peptides and proteins, thus increasing the confidence and sensitivity of hits compared to single-search-engine processing^{44,45}. This strategy led to the detection of several thousand alternative proteins. However, ribosome profiling and MS have technical caveats and the comprehensive contribution of small proteins to the proteome will require more efforts, including the development of new tools such as specific antibodies.

In conclusion, our deep annotation of the transcriptome reveals that the vast majority of small eukaryotic proteins are still officially unannotated and that many small and large proteins coded by the same mRNA probably cooperate.

Methods

Generation of alternative open reading frames (altORFs) and alternative protein databases. Throughout this manuscript, annotated protein coding sequences and proteins in current databases are labelled annotated coding sequences or CDSs and reference proteins, respectively. For simplicity reasons, predicted alternative protein coding sequences are

labelled alternative open reading frames or altORFs.

To generate MySQL databases containing the sequences of all predicted alternative proteins translated from reference annotation of different organisms, a computational pipeline of Perl scripts was developed as previously described with some modifications³³. Genome annotations for *H. sapiens* (release hg38, Assembly: GCF_000001405.26), *P. troglodytes* (Pan_troglodytes-2.1.4, Assembly: GCF_000001515.6), *M. musculus* (GRCm38.p2, Assembly: GCF_000001635.22), *B. taurus* (release Btau 4.6.1, Assembly: GCF_000003205.5), *D. melanogaster* (release 6, Assembly: GCA_000705575.1), *C. elegans* (WBcel235, Assembly: GCF_000002985.6) and *S. cerevisiae* (Sc_YJM993_v1, Assembly: GCA_000662435.1) were downloaded from the NCBI website (<http://www.ncbi.nlm.nih.gov/genome>). For *X. tropicalis* (release JGI_4.2.78) and *D. rerio* (ZV9.79), genome annotations were downloaded from Ensembl (<http://www.ensembl.org/info/data/ftp/>). Each annotated transcript was translated *in silico* with Transeq⁴⁶. All ORFs starting with an AUG and ending with a stop codon different from the CDS, with a minimum length of 30 codons (including the stop codon) and identified in a distinct reading frame compared to the annotated CDS were defined as altORFs.

AltORF localization was assigned according to the position of the predicted translation initiation site (TIS): altORFs^{5'}, altORFs^{CDS} and altORFs^{3'} are altORFs with TISs located in 5'UTRs, CDSs and 3'UTRs, respectively. Non-coding RNAs (ncRNAs) have no annotated CDS and all ORFs located within ncRNAs are labelled altORFs^{nc}.

The presence of the simplified Kozak sequence (A/GNNATGG) known to be favorable for efficient translation initiation was also assessed for each predicted altORF⁴⁷.

Identification of TISs. The global aggregates of initiating ribosome profiles data were obtained from the initiating ribosomes tracks in the GWIPS-viz genome browser²³ with ribosome profiling data collected from five large scale studies^{2,9,48-50}. Sites were mapped to hg38 using a chain file from the UCSC genome browser (<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/liftOver/hg19ToHg38.over.chain.gz>) and CrossMap v0.1.6 (<http://crossmap.sourceforge.net/>). Similar to the methods used in these studies, an altORF is considered as having an active TIS if it is associated with at least 10 reads at one of the 7 nucleotide positions of the sequence NNNAUGN (AUG is the predicted altORF TIS). An additional recent study was also included in our analysis²⁴. Raw sequencing data for ribosome protected fragments in harringtonine treated cells was aligned to the human genome (GRCh38) using bowtie2 (2.2.8). Similar to the method used in this work, altORFs with at least 5 reads overlapping one position in the kozak region were considered as having an experimentally validated TIS.

Generation of shuffled transcriptomes. Each annotated transcript was shuffled using the Fisher-Yates shuffle algorithm. In CDS regions, all codons were shuffled except the initiation and stop codons. For mRNAs, we shuffled the 5'UTRs, CDSs and 3'UTRs independently to control for base composition. Non-coding regions were shuffled at the nucleotide level. The resulting shuffled transcriptome has the following features compared to hg38: same number of transcripts, same transcripts lengths, same nucleotide composition, and same amino-acid composition for the proteins translated from the CDSs. Shuffling was repeated 100 times and the results are presented with average values and standard deviations. The total number of altORFs is 551,380 for hg38, and an average of 489,073 for shuffled hg38. AltORFs and

kozak motifs in the 100 shuffled transcriptomes were detected as described above for hg38.

Identification of paralogs/orthologs in alternative proteomes. Both alternative and reference proteomes were investigated. Pairwise ortholog and paralog relationships between the human proteomes and the proteomes from other species, were calculated using an InParanoid-like approach⁵¹, as described below. The following BLAST procedure was used. Comparisons using our datasets of altORFs/CDS protein sequences in multiple FASTA formats from *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Danio rerio*, *Xenopus tropicalis*, *Bos taurus*, *Mus musculus*, *Pan troglodytes*, *Homo sapiens* were performed between each pair of species (human against the other species), involving four whole proteome runs per species pair: pairwise comparisons (organism A vs organism B, organism B vs organism A), plus two self-self runs (organism A vs organism A, organism B vs organism B). BLAST homology inference was accepted when the length of the aligned region between the query and the match sequence equalled or exceeded 50% of the length of the sequence, and when the bitscore reached a minimum of 40⁵². Orthologs were detected by finding the mutually best scoring pairwise hits (reciprocal best hits) between datasets A-B and B-A. The self-self runs were used to identify paralogy relationships as described⁵¹.

Analysis of third codon position (wobble) conservation. Basewise conservation scores for the alignment of 100 vertebrate genomes including *H. sapiens* were obtained from UCSC genome browser (<http://hgdownload.soe.ucsc.edu/goldenPath/hg38/phyloP100way/>). Conservation PhyloP scores relative to each nucleotide position within codons were

extracted using a custom Perl script and the Bio-BigFile module version 1.07. The PhyloP conservation score for the wobble nucleotide of each codon within the CDS was extracted. For the 53,862 altORFs completely nested inside 20,814 CDSs, the average PhyloP score for wobble nucleotides within the altORF region was compared to the average score for the complete CDS. To generate shuffled controls, PhyloP scores for wobble nucleotides were randomly shuffled for the complete CDSs. We compared the differences between altORFs and CDSs PhyloP scores (altORF PhyloP – CDS PhyloP) to those generated based on shuffled versions. We identified expected quantiles of the differences (“DQ” column in the table), and compared these to the observed differences. Because there was greater conservation of wobble nucleotides PhyloP scores within altORFs regions located farther from the center of their respective genes ($r = 0.08$, $p < 0.0001$), observed differences were adjusted using an 8-knot cubic basis spline of percent distance from center. These observed differences were also adjusted for site-specific signals as detected in the shuffled analyses. We assessed the slope of the PhyloP difference relative to CDS PhyloP by using a linear mixed model with a random term for the CDS to control for the potential non-independence of altORFs within a CDS (R, lme4 package, lmer function).

Human alternative protein classification and in silico functional annotation.

Repeat and transposable element annotation

RepeatMasker, a popular software to scan DNA sequences for identifying and classifying repetitive elements, was used to investigate the extent of altORFs derived from transposable elements⁵³. Version 3-3-0 was run with default settings.

Alternative protein analysis using InterProScan

InterProScan combines 15 different databases, most of which use Hidden Markov models for signature identification⁵⁴. Interpro merges the redundant predictions into a single entry and provides a common annotation. A recent local version of InterProScan 5.14-53.0 was run using default parameters to scan for known protein domains in alternative proteins. Gene ontology (GO) and pathway annotations were also reported if available with `-goterm` and `-pa` options. Only protein signatures with an E-value $\leq 10^{-3}$ were considered.

We classified the reported InterPro hits as belonging to one or several of three clusters; (1) alternative proteins with InterPro entries; (2) alternative proteins with signal peptides (SP) and/or transmembrane domains (TM) predicted by at least two of the three SignalP, PHOBIUS, TMHMM tools and (3) alternative proteins with other signatures.

The GO terms assigned to alternative proteins with InterPro entries were grouped and categorised into 13 classes within the three ontologies (cellular component, biological process, molecular function) using the CateGORizer tool⁵⁵.

Each unique alternative protein with InterPro entries and its corresponding reference protein (encoded in the same transcript) were retrieved from our InterProscan output. Alternative and reference proteins without any InterPro entries were ignored. The overlap in InterPro entries between alternative and reference proteins was estimated as follows. We went through the list of alternative/reference protein pairs and counted the overlap in the number of entries between the alternative and reference proteins as $100 \times \text{intersection} / \text{union}$. All reference proteins and the corresponding alternative proteins were combined together in each comparison so that all domains of all isoforms for a given reference protein were considered in each comparison. The random distribution of the number of alternative/reference protein

pairs that share at least one InterPro entry was computed by shuffling the alternative/reference protein pairs and calculating how many share at least one InterPro entry. This procedure was repeated 1,000 times. Finally, we compared the number and identity of shared InterPro entries in a two dimensional matrix to illustrate which Interpro entries are shared. In many instances, including for zinc-finger coding genes, InterPro entries in alternative/reference protein pairs tend to be related when they are not identical.

Mass Spectrometry identification parameters. Wrapper Perl scripts were developed for the use of SearchGUI v2.0.11⁵⁶ and PeptideShaker v1.1.0⁴⁴ on the Université de Sherbrooke's 39,168 core high-performance Mammouth Parallèle 2 computing cluster (<http://www.calculquebec.ca/en/resources/compute-servers/mammouth-parallele-ii>).

SearchGUI was configured to run the following proteomics identification search engines: X!Tandem⁵⁷, MS-GF+⁵⁸, MyriMatch⁵⁹, Comet⁶⁰, and OMSSA⁶¹. SearchGUI parameters were set as follow: maximum precursor charge, 5; maximum number of PTM per peptide, 5; X!Tandem minimal fragment m/z, 140; removal of initiator methionine for Comet, 1. A full list of parameters used for SearchGUI and PeptideShaker is available in Table S2. For PXD000953 dataset³⁰, precursor and fragment tolerance were set 0.006 Da and 0.1 Da respectively, with carbamidomethylation of C as a fixed modification and Nter-Acetylation and methionine oxidation as variable modifications. For PXD000788²⁸ and PXD000612²⁹ datasets, precursor and fragment tolerance were set to 4.5 ppm and 0.1 Da respectively with carbamidomethylation of cysteine as a fixed modification and Nter-Acetylation, methionine oxidation and phosphorylation of serine, threonine and tyrosine as variable modifications. For PXD002815 dataset²⁷, precursor and fragment tolerance were set to 4.5 ppm and 0.1 Da

respectively with carbamidomethylation of cysteine as a fixed modification and Nter-Acetylation and methionine oxidation as variable modifications. Datasets were searched using a target-decoy approach against a composite database composed of a target database [Uniprot canonical and isoform reference proteome (16 January 2015) for a total of 89,861 sequences + custom alternative proteome resulting from the in silico translation of all human altORFs (available to download at <https://www.roucoulab.com/p/downloads>)], and their reverse protein sequences from the target database used as decoys. False discovery rate cut-offs were set at 1% for PSM, peptides and proteins. Only alternative proteins identified with at least one unique and specific peptide, and with at least one confident PSM in the PeptideShaker Hierarchical Report were considered valid⁴⁴.

For spectral validation (Supplementary Fig. 14, 22-25), synthetic peptides were purchased from the peptide synthesis service at the Université de Sherbrooke. Peptides were solubilized in 10% acetonitrile, 1% formic acid and directly injected into a Q-Exactive mass spectrometer (Thermo Scientific) via an electro spray ionization source (Thermo Scientific). Spectra were acquired using Xcalibur 2.2 at 70000 resolution with an AGC target of 3e6 and HCD collision energy of 25. Peaks were assigned manually by comparing monoisotopic m/z theoretical fragments and experimental (PeptideShaker) spectra.

Cloning and antibodies. Human Flag-tagged altMiD51(WT) and altMiD51(LYR→AAA), and HA-tagged DrP1K38A were cloned into pcDNA3.1 (Invitrogen) using a Gibson assembly kit (New England Biolabs, E26115). The cDNA corresponding to human MiD51/MIEF1/SMCR7L transcript variant 1 (NM_019008) was also cloned into pcDNA3.1 by Gibson assembly. In this construct, altMiD51 and MiD51 were tagged with Flag and HA

tags, respectively. MiD51^{GFP} and altMiD51^{GFP} were also cloned into pcDNA3.1 by Gibson assembly. For MiD51^{GFP}, a LAP tag²⁷ was inserted between MiD51 and GFP. gBlocks were purchased from IDT.

For immunofluorescence, primary antibodies were diluted as follow: anti-Flag (Sigma, F1804) 1/1000, anti-TOM20 (Abcam, ab186734) 1/500. For western blots, primary antibodies were diluted as follow: anti-Flag (Sigma, F1804) 1/1000, anti-HA (BioLegend, 901515) 1/500, anti-actin (Sigma, A5441) 1/10000, anti-Drp1 (BD Transduction Laboratories, 611112) 1/500.

Cell culture, immunofluorescence, knockdown and western blots. Cell cultures, transfections, immunofluorescence, confocal analyses and western blots were carried out as previously described⁶². Mitochondrial morphology was analyzed as previously described⁶³. A minimum of 100 cells were counted (n=3 or 300 cells for each experimental condition). Three independent experiments were performed.

For Drp1 knockdown, 25,000 HeLa cells in 24-well plates were transfected with 25 nM Drp1 SMARTpool: siGENOME siRNA (Dharmacon, M-012092-01-0005) or ON-TARGET plus Non-targeting pool siRNAs (Dharmacon, D-001810-10-05) with DharmaFECT 1 transfection reagent (Dharmacon, T-2001-02) according to the manufacturer's protocol. After 24h, cells were transfected with pcDNA3.1 or altMiD51, incubated for 24h, and processed for immunofluorescence or western blot.

Trypan blue quenching experiment was performed as previously described⁶⁴.

Mitochondrial parameters and ROS production. A flux analyzer (XF96 Extracellular

Flux Analyzer; Seahorse Bioscience, Agilent technologies) was used to determine the mitochondrial function in HeLa cells overexpressing AltMiD51^{Flag}. Cells were plated in a XF96 plate (Seahorse Biosciences) at 1×10^4 cells per well in Dulbecco's modified Eagle's medium supplemented with 10% FBS with antibiotics. After 24 hours, cells were transfected for 24 hours with an empty vector (pcDNA3.1) or with the same vector expressing AltMiD51^{Flag} with GeneCellin transfection reagent according to the manufacturer's instructions. Cells were equilibrated in XF assay media supplemented with 25 mM glucose and 1 mM pyruvate and were incubated at 37°C in a CO₂-free incubator for 1 h. Baseline oxygen consumption rates (OCRs) of the cells were recorded with a mix/wait/measure times of 3/0/3 min respectively. Following these measurements, oligomycin (1 μM), FCCP (0.5 μM), and antimycin A/rotenone (1 μM) were sequentially injected, with oxygen consumption rate measurements recorded after each injection. Data were normalized to total protein in each well. For normalization, cells were lysed in the 96-well XF plates using 15 μl/well of RIPA lysis buffer (1% Triton X-100, 1% NaDeoxycholate, 0.1% SDS, 1mM EDTA, 50 mM Tris-HCl pH7.5). Protein concentration was measured using the BCA protein assay reagent (Pierce, Waltham, MA, USA).

Reactive oxygen species (ROS) levels were measured using Cellular ROS/Superoxide Detection Assay Kit (Abcam #139476). HeLa cells were seeded onto 96-well black/clear bottom plates at a density of 6,000 cells per well with 4 replicates for each condition. After 24 hours, cells were transfected for 24 hours with an empty vector (pcDNA3.1) or with the same vector expressing AltMiD51^{Flag} with GeneCellin according to the manufacturer's instruction. Cells were untreated or incubated with the ROS inhibitor (N-acetyl-L-cysteine) at 10mM for 1 hour. Following this, the cells were washed twice with the wash solution and

then labeled for 1 hour with the Oxidative Stress Detection Reagent (green) diluted 1:1000 in the wash solution with or without the positive control ROS Inducer Pyocyanin at 100 μ M. Fluorescence was monitored in real time. ROS accumulation rate was measured between 1 to 3 hours following induction. After the assay, total cellular protein content was measured using BCA protein assay reagent (Pierce, Waltham, MA, USA) after lysis with RIPA buffer. Data were normalised for initial fluorescence and protein concentration.

ATP synthesis was measured as previously described⁶⁵ in cells transfected for 24 hours with an empty vector (pcDNA3.1) or with the same vector expressing AltMiD51^{Flag}.

References

1. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
2. Lee, S. S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2424–2432 (2012).
3. Mouilleron, H., Delcourt, V. & Roucou, X. Death of a dogma: eukaryotic mRNAs can code for more than one protein. *Nucleic Acids Res.* **44**, 14–23 (2015).
4. Pauli, A. *et al.* Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors. *Science* **343**, 1248636–1248636 (2014).
5. Anderson, D. M. *et al.* A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell* **160**, 595–606 (2015).
6. Zanet, J. *et al.* Pri sORF peptides induce selective proteasome-mediated protein processing. *Science* **349**, 1356–1358 (2015).
7. Nelson, B. R. *et al.* A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* 271–275 (2016).
8. Bazzini, A. A. *et al.* Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* **33**, 981–993 (2014).
9. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* **4**, (2015).
10. Prabakaran, S. *et al.* Quantitative profiling of peptides from RNAs classified as noncoding. *Nat. Commun.* **5**, 5429 (2014).
11. Slavoff, S. a *et al.* Peptidomic discovery of short open reading frame-encoded peptides

- in human cells. *Nat. Chem. Biol.* **9**, 59–64 (2013).
12. Andrews, S. J. & Rothnagel, J. A. Emerging evidence for functional peptides encoded by short open reading frames. *Nat. Rev. Genet.* **15**, 193–204 (2014).
 13. Landry, C. R., Zhong, X., Nielly-Thibault, L. & Roucou, X. Found in translation: functions and evolution of a recently discovered alternative proteome. *Curr. Opin. Struct. Biol.* **32**, 74–80 (2015).
 14. Fields, A. P. *et al.* A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. *Mol. Cell* **60**, 816–827 (2015).
 15. Saghatelian, A. & Couso, J. P. Discovery and characterization of smORF-encoded bioactive polypeptides. *Nat. Chem. Biol.* **11**, 909–16 (2015).
 16. Stock, D., Leslie, A. G. & Walker, J. E. Molecular architecture of the rotary motor in ATP synthase. *Science* **286**, 1700–1705 (1999).
 17. Schmitt, J. P. *et al.* Dilated cardiomyopathy and heart failure caused by a mutation in phospholamban. *Science* **299**, 1410–1413 (2003).
 18. Nemeth, E. *et al.* Heparin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science* **306**, 2090–2093 (2004).
 19. Iacono, M., Mignone, F. & Pesole, G. uAUG and uORFs in human and rodent 5′ untranslated mRNAs. *Gene* **349**, 97–105 (2005).
 20. Neafsey, D. E. & Galagan, J. E. Dual modes of natural selection on upstream open reading frames. *Mol. Biol. Evol.* **24**, 1744–51 (2007).
 21. Smith, E. *et al.* Leaky ribosomal scanning in mammalian genomes: significance of histone H4 alternative translation in vivo. *Nucleic Acids Res.* **33**, 1298–1308 (2005).
 22. Pop, C. *et al.* Causal signals between codon bias, mRNA structure, and the efficiency

- of translation and elongation. *Mol. Syst. Biol.* **10**, 770 (2014).
23. Michel, A. M. *et al.* GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res.* **42**, D859–864 (2014).
 24. Raj, A. *et al.* Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *Elife* **5**, e13328 (2016).
 25. Miettinen, T. P. & Björklund, M. Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Res.* **43**, 1019–1034 (2015).
 26. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* **4**, e08890 (2015).
 27. Hein, M. Y. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163**, 712–723 (2015).
 28. Tong, J., Taylor, P. & Moran, M. F. Proteomic analysis of the epidermal growth factor receptor (EGFR) interactome and post-translational modifications associated with receptor endocytosis in response to EGF and stress. *Mol. Cell. Proteomics* **13**, 1644–1658 (2014).
 29. Sharma, K. *et al.* Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Rep.* **8**, 1583–1594 (2014).
 30. Rosenberger, G. *et al.* A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci. data* **1**, 140031 (2014).
 31. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

32. Mitchell, A. *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213–221 (2014).
33. Vanderperre, B. *et al.* Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS One* **8**, e70698 (2013).
34. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
35. Laity, J. H., Lee, B. M. & Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11**, 39–46 (2001).
36. Andreev, D. E. *et al.* Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife* **4**, e03971 (2015).
37. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
38. Angerer, H. Eukaryotic LYR Proteins Interact with Mitochondrial Protein Complexes. *Biology (Basel)*. **4**, 133–150 (2015).
39. Losón, O. C., Song, Z., Chen, H. & Chan, D. C. Fis1, Mff, MiD49, and MiD51 mediate Drp1 recruitment in mitochondrial fission. *Mol. Biol. Cell* **24**, 659–667 (2013).
40. Motori, E. *et al.* Inflammation-Induced Alteration of Astrocyte Mitochondrial Dynamics Requires Autophagy for Mitochondrial Network Maintenance. *Cell Metab.* **18**, 844–859 (2013).
41. Smirnova, E., Shurland, D. L., Ryazantsev, S. N. & van der Bliek, A. M. A human dynamin-related protein controls the distribution of mitochondria. *J. Cell Biol.* **143**, 351–358 (1998).
42. Rosenquist, M., Sehnke, P., Ferl, R. J., Sommarin, M. & Larsson, C. Evolution of the

- 14-3-3 protein family: does the large number of isoforms in multicellular organisms reflect functional specificity? *J. Mol. Evol.* **51**, 446–58 (2000).
43. Johnstone, T. G., Bazzini, A. A. & Giraldez, A. J. Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.* (2016). doi:10.15252/emj.201592759
 44. Vaudel, M. *et al.* PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* **33**, 22–24 (2015).
 45. Shteynberg, D. *et al.* iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell. Proteomics* **10**, M111.007690 (2011).
 46. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
 47. Kozak, M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299**, 1–34 (2002).
 48. Fritsch, C. *et al.* Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* **22**, 2208–2218 (2012).
 49. Stern-Ginossar, N. *et al.* Decoding human cytomegalovirus. *Science* **338**, 1088–93 (2012).
 50. Gao, X. *et al.* Quantitative profiling of initiating ribosomes in vivo. *Nat. Methods* **12**, 147–53 (2015).
 51. Sonnhammer, E. L. L. & Östlund, G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* **43**, D234–239 (2015).
 52. Remm, M., Storm, C. E. & Sonnhammer, E. L. Automatic clustering of orthologs and

- in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
53. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **Chapter 4**, Unit 4.10 (2009).
54. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
55. Na, D., Son, H. & Gsponer, J. Categorizer: a tool to categorize genes into user-defined biological groups based on semantic similarity. *BMC Genomics* **15**, 1091 (2014).
56. Vaudel, M., Barsnes, H., Berven, F. S., Sickmann, A. & Martens, L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* **11**, 996–999 (2011).
57. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467 (2004).
58. Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **5**, 5277 (2014).
59. Tabb, D. L., Fernando, C. G. & Chambers, M. C. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J. Proteome Res.* **6**, 654–661 (2007).
60. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24 (2013).
61. Geer, L. Y. *et al.* Open mass spectrometry search algorithm. *J. Proteome Res.* **3**, 958–64 (2004).
62. Vanderperre, B. *et al.* An overlapping reading frame in the PRNP gene encodes a novel polypeptide distinct from the prion protein. *FASEB J.* **25**, 2373–86 (2011).

63. Palmer, C. S. *et al.* MiD49 and MiD51, new components of the mitochondrial fission machinery. *EMBO Rep.* **12**, 565–573 (2011).
64. Vanderperre, B. *et al.* MPC1-like: a Placental Mammal-Specific Mitochondrial Pyruvate Carrier Subunit Expressed in Post-Meiotic Male Germ Cells. *J. Biol. Chem.* (2016). doi:10.1074/jbc.M116.733840
65. Vives-Bauza, C., Yang, L. & Manfredi, G. Assay of Mitochondrial ATP Synthesis in Animal Cells and Tissues. *Methods Cell Biol* **80**, 155–171 (2007).

Acknowledgements

This research was supported by CIHR grants MOP-137056 and MOP-136962 to X.R; MOP-299432 and MOP-324265 to C.L; a Université de Sherbrooke institutional research grant made possible through a generous donation by Merck Sharp & Dohme to X.R; a FRQNT team grant 2015-PR-181807 to C.L. and X.R; Canada Research Chairs in Functional Proteomics and Discovery of New Proteins to X.R, in Evolutionary Cell and Systems Biology to C.L and in Computational and Biological Complexity to A.O; A.A.C is supported by a Canadian Institutes of Health Research New Investigator Salary Award; M.S.S is a recipient of a Fonds de Recherche du Québec – Santé Research Scholar Junior 1 Career Award; V.D is supported in part by fellowships from Région Nord-Pas de Calais and PROTEO; A.A.C, D.J.H, M.S.S and X.R are members of the Fonds de Recherche du Québec Santé-supported Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke. We thank the staff from the Centre for Computational Science at the Université de Sherbrooke, Compute Canada and Compute Québec for access to the Mammoth supercomputer.

Author contributions

X.R., M.S., C.R.L. and A.A.C. conceived the study. S.S., A.V.R, J-F.L., J.G., V.D., M.C.B., J.M., J-F.J. C.R.L., I. A.-G. and A.A.C. performed the experiment. S.S., J-F.L., J.G., A.A.C. and C.R.L. performed the bioinformatics and conservation analyses. A.V.R and M.C.B. performed the functional characterization of altMiD51. X.R., M.S. and D.J.H supervised the project, and participated in experimental design and data analyses with A.A.C., C.R.L. A.O., J.M., I.F. and J-F.J. X.R. S.S. and A.V.R. wrote the manuscript with contributions from all co-authors.

3 DISCUSSION

Ce projet de maîtrise porte dans un premier temps sur la caractérisation d'une nouvelle protéine alternative, altMiD51. Pour ce faire, son expression et sa localisation ont été démontrées. En effet, nos résultats mettent en évidence la co-expression d'altMiD51 et MiD51 dans la cellule par immunobuvardage lorsque l'ADNc correspondant à l'ARNm du gène *SMCR7L* est transfecté dans des cellules HeLa. De plus, nous pouvons aisément détecter l'expression d'altMiD51 dans des cellules par spectrométrie de masse, aussi bien dans nos études que dans des études déjà publiées et réanalysées avec notre banque de données de protéines alternatives. Par immunofluorescence sur des cellules HeLa transfectées avec l'ADNc d'altMiD51, sa localisation mitochondriale a été démontrée, semblablement à sa protéine de référence. Plus précisément, elle se situe dans la matrice mitochondriale d'après nos résultats avec le bleu de trypan. Plus généralement, ces résultats sont en faveur de l'existence d'ARNm multicodants chez les eucaryotes, que je discuterai plus en profondeur plus loin dans cette discussion.

Dans un deuxième temps, la surexpression d'altMiD51 dans des cellules HeLa a permis l'observation par immunofluorescence d'une morphologie mitochondriale altérée. En effet, nos résultats démontrent que la surexpression d'altMiD51 induit la fragmentation des mitochondries via son domaine LYR. Le remplacement de seulement trois acides aminés par trois alanines (LYR→AAA) du domaine LYR (comprenant une cinquantaine d'acides aminés) abolit la fragmentation des mitochondries. Étant donné que la protéine de référence MiD51 est impliquée dans la fission mitochondriale par le recrutement et la régulation de Drp1, ce résultat suggère une coopération fonctionnelle avec MiD51 dans la régulation de la dynamique mitochondriale. Je discuterai plus loin de la coopération fonctionnelle entre les protéines alternatives et de référence. Je discuterai plus en détails de la coopération entre altMiD51 et MiD51 par l'implication d'altMiD51 dans la morphologie mitochondriale et sa relation avec l'acteur majeur de la fission, Drp1.

3.1 Le potentiel inexploré de l'ARNm eucaryote

Il est connu que le génome des virus est très petit, mais organisé et optimisé pour coder le plus d'information possible. Pour cela, les virus optimisent leur génome par la présence d'ORFs dans les différents cadres de lecture (Bolinger et Boris-Lawrie, 2009). En parallèle, les procaryotes possèdent des gènes organisés en opérons, permettant d'exprimer à

partir du même ARNm plusieurs protéines impliquées dans la même voie (Fortino *et al.*, 2014). Ce potentiel multicodant des ARNms permet une économie d'énergie de la cellule par une régulation stœchiométrique et temporelle des protéines impliquées dans la même voie (Wells *et al.*, 2016). Bien que d'autres organismes aient des mécanismes élaborés et différents de traduction, l'ARNm eucaryote ne possède, à de rares exceptions près, qu'un seul ORF annoté.

Toutefois, nos analyses, les différentes études portant sur les sORF et l'émergence des études de profilage ribosomal renforcent grandement le besoin de reconsidérer le dogme d'une seule séquence codante dans un ARNm eucaryote. Avec l'accumulation d'évidences de traduction d'altORF par les techniques de profilage ribosomal et de spectrométrie de masse, il s'avère que plusieurs des ARNms eucaryotes auraient un potentiel multicodant ajoutant un niveau supérieur de complexité aux génomes (Figure supplémentaire 8). Bien qu'il soit possible que la plupart des altORFs présents dans notre base de données ne soient pas traduits en protéines fonctionnelles, il n'en reste pas moins que même 1% des 183 191 altORFs uniques prédits serait un ajout notable au protéome actuel, soit 1832 nouvelles protéines. Plus particulièrement, nos études bio-informatiques à large échelle démontrent que les codons d'initiation d'environ 15% des altORFs (soit 27 539 protéines alternatives) peuvent avoir des séquences Kozak comme 38% des CDSs (soit 19,745 protéines de référence) (Figure 2.1E). Aussi, des milliers de protéines alternatives sont conservées à travers plusieurs espèces. Par exemple, il y a 25 930 protéines alternatives humaines conservées entre l'homme et la souris (Figure 2.2A). Nous avons pu observer une corrélation entre la conservation d'une protéine de référence et la conservation de sa protéine alternative et plus spécifiquement, un enrichissement de protéines alternatives hautement conservées et qui évoluent plus rapidement que leur protéine de référence (Figure 2.2B). De plus, certaines protéines alternatives contiennent des signatures moléculaires connues du protéome de référence (Figure supplémentaire 10). Plus particulièrement, certaines protéines alternatives contiennent la même signature moléculaire que leur protéine de référence (Figure 2.4).

Mais pourquoi n'avons-nous pas détecté les altORFs et leur protéines alternatives plus tôt? Une partie de la réponse s'explique par l'annotation du génome dans le début des années 2000. Étant déjà une tâche ardue et exploitant les ressources informatiques et technologiques de l'époque, il est possible que l'annotation par la vision simpliste de l'ARNm monocistronique et par le mécanisme de balayage cap-dépendant soit à l'origine de l'ignorance des petits ORFs alternatifs. Il existe à ce jour plus de 51 818 protéines de références annotées. Certaines d'entre-elles n'ont aucune fonction biologique attribuée et on ignore

leur rôle précis dans les maladies humaines. Cependant, l'ajout de 183 191 protéines supplémentaires de 30 acides aminés et plus à étudier semble un défi de grande envergure. Par contre, la simple reconnaissance de leur existence et leur considération dans les recherches futures ne feront qu'enrichir les découvertes. Il est à notre avantage d'exploiter le potentiel multicodeant de l'ARNm eucaryote. Pour ce faire, j'ai investigué l'ARNm du gène *SMCR7L* pour démontrer qu'un altORF et un CDS peuvent être traduits à partir du même ARNm.

3.2 Le potentiel multicodeant de l'ARNm du gène *SMCR7L*

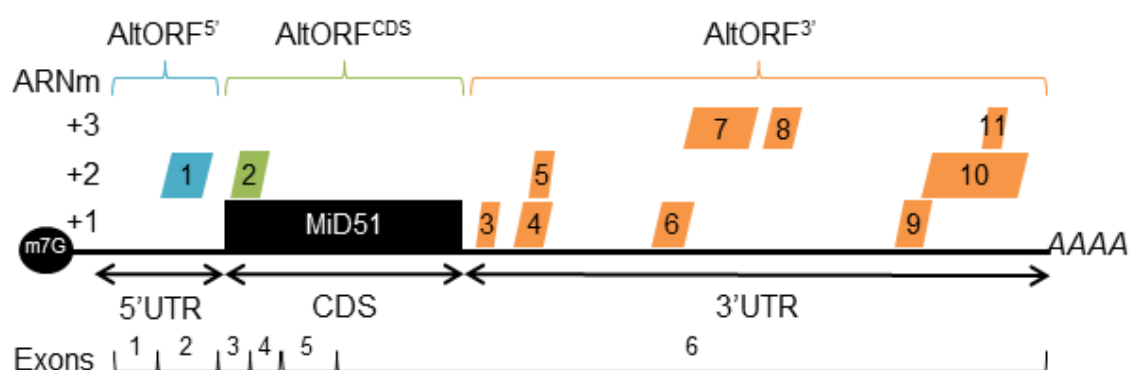


FIGURE 3.1 – L'ARNm du gène *SMCR7L* est multicodeant pour *MiD51* et plusieurs altORFs.

Représentation à l'échelle des différents ORFs contenus dans l'ARNm (NM_019008.4) du gène *SMCR7L* contenu dans le chromosome 22 chez l'humain. Chaque altORF est numéroté selon leur ordre 5' vers 3' et ceux-ci sont décrits dans le tableau 3.1.

#	Accession	Cadre	Exon	Kosak	Début	Fin	Longueur (a.a)
1	IP_294711.1	2	2	0	39504231	39504443	70
2	IP_294715.1	2	3-4	0	39511329	39511894	51
3	IP_294713.1	1	6	0	39514456	39514575	39
4	IP_294710.1	1	6	0	39514606	39514761	51
5	IP_294719.1	2	6	0	39514637	39514765	42
6	IP_294714.1	1	6	0	39515188	39515373	61
7	IP_294707.1	3	6	0	39515340	39515657	105
8	IP_294708.1	3	6	0	39515658	39515822	54
9	IP_294712.1	1	6	0	39517375	39517569	64
10	IP_294717.1	2	6	1	39517526	39518002	158
11	IP_294706.1	3	6	0	39517764	39517859	31

Tableau 3.1 – Les altORFs contenus dans le gène *SMCR7L* humain.

Répertoire des altORFs ayant >30 acide aminés contenu dans l'ARNm (NM_019008.4) du gène *SMCR7L*. L'altORf #1 correspond à altMiD51.

L'exemple du couple altMiD51 et MiD51 en plus des autres duos détaillés dans la section 1.3 met en avant l'hypothèse scientifique qu'un ARNm peut contenir plusieurs ORFs qui seront traduits. Ainsi, un gène est transcrit en un ARNm qui ne contiendrait pas qu'un seul message, mais plusieurs. C'est le cas du gène *SMCR7L* codant pour un ARNm qui, vraisemblablement, contient au moins deux messages pour le dynamisme mitochondrial des cellules : AltMiD51 et MiD51 (Figure 3.1).

Hormis les études ayant détectée altMiD51 dans les ARNms de la cellule, nous n'avons pas la certitude absolue que les deux protéines sont traduites à partir du même ARNm. Nous pouvons coexprimer en cellule à partir d'un plasmide les deux protéines à partir de la séquence de l'ARNm contenu dans les bases de données pour le gène *SMCR7L*. Néanmoins, nous ne pouvons pas ignorer la possibilité qu'il existerait un ARNm ne codant que pour altMiD51 et un ARNm codant pour MiD51 et qui contiendrait l'altORF d'AltMiD51 non traduit. Pour prouver une fois pour toutes que AltMiD51 et MiD51 proviennent du même ARNm, nous allons quantifier de manière absolue les deux protéines en spectrométrie de masse en présence ou non d'un siARN ciblant la région CDS de l'ARNm du gène *SMCR7L*. Si la quantité des deux protéines diminue de façon égale, nous aurons la preuve qu'il n'existe pas d'autre ARNm codant pour altMiD51, car sinon le niveau de altMiD51 ne serait pas autant diminué que celui de MiD51.

3.2.1 Un uORF codant pour une protéine fonctionnelle

Il n'est pas faux de dire que l'altORF codant pour altMiD51 est un uORF. Par définition, les uORFs sont des ORFs en amont des CDSs dans le 5'UTR des ARNms. Selon notre nomenclature, les uORFs sont des altORFs⁵. Ils diminuent l'expression du CDS par le ralentissement du ribosome, par la réinitiation du balayage du ribosome passé l'AUG du CDS, par le décrochage prématuré du ribosome et autres mécanismes discutés dans la section (1.3.1.1). AltMiD51 est située dans le 5'UTR en amont du CDS pour MiD51, ce qui *a priori* est la définition d'un uORF. Bien que nous n'ayons pas concentré nos recherches sur la régulation traductionnelle par le uORF d'altMiD51 sur le CDS de MiD51, Andreev et collègues ont observé des changements d'efficacité de traduction de MiD51 par profilage ribosomal entre une condition cellulaire normale et en présence d'un stress (Andreev *et al.*, 2015).

Toutefois, cet uORF présent dans le gène *SMCR7L* code pour une protéine de 70 acides aminés, possède un domaine LYR, est détectée en spectrométrie de masse et par la technique de profilage ribosomal. De plus, altMiD51 joue un rôle dans la fission mitochondriale. Ainsi, cet uORF n'est pas seulement un régulateur du CDS pour MID51, mais

aussi un message contenu dans l'ARNm qui code pour une protéine fonctionnelle. Il ne serait pas étonnant que plusieurs uORFs aient un double rôle : régulateur de traduction du CDS et codant pour une protéine fonctionnelle.

3.2.2 *AltMiD51 est une protéine à domaine LYR : LYRM12*

Bien que la surexpression d'altMiD51 induise la fission mitochondriale, aucune autre protéine à domaine LYR n'est connue à l'heure actuelle pour influencer le dynamisme mitochondrial. Il ne serait pas impossible que la fragmentation mitochondriale soit un effet secondaire à une autre fonction d'altMiD51. Il ne faut pas oublier qu'altMiD51 est avant toute la 12^e protéine à domaine LYR et que sa fonction primaire est peut-être reliée à ce domaine. Néanmoins, lorsque le domaine LYR d'altMiD51 est muté (LYR→AAA), la fission mitochondriale par altMiD51 est abolie. Un réseau mitochondrial tubulaire semblable aux cellules transfectées avec un vecteur vide est obtenu lorsqu'on surexprime ce mutant AAA.

Les protéines contenant le motif Leu-Tyr-Arg (LYR) sont connues pour se retrouver dans des complexes protéiques et être impliquées dans le transfert de centre fer-soufre (Rouault, 2015). Les centres fer-soufre sont des cofacteurs de certaines protéines mitochondriales. Ils peuvent accepter ou donner un électron pour effectuer les réactions d'oxydation et de réduction de la chaîne respiratoire (Rouault, 2015). La protéine à domaine LYR la plus connue se nomme LYRM4/ISD11 et est conservée de l'humain à la levure (Lim *et al.*, 2013). Elle est connue pour se loger dans un complexe important dans la formation des centres fer-soufre. D'autres protéines à domaine LYR sont aussi logées dans les complexes de la chaîne respiratoire et seraient importantes pour leur intégrité et leur bon fonctionnement (Angerer, 2013). C'est le cas de LYRM6 au niveau du complexe 1 de la chaîne respiratoire, dont le domaine LYR muté diminue le fonctionnement du complexe 1 (Angerer *et al.*, 2014). Pour déterminer si altMiD51 se loge aussi dans un des complexes de la chaîne respiratoire, deux expériences avec l'appareil Seahorse ont été réalisées avec l'ajout du mutant LYR→AAA qui ne peut pas fissionner. Aucune différence entre le mutant et altMiD51 sauvage (résultats non présentés) n'a été observée. Cependant, altMiD51 pourrait être dans un autre complexe mitochondrial nécessitant un centre fer-soufre. En somme, les protéines à domaine LYR sont peu connues et peu étudiées. Nous ne savons pas quels acides aminés sont impliqués dans leur localisation mitochondriale et elles n'ont pas de domaine de localisation à la mitochondrie prédit. Nous ne savons pas encore toutes leurs fonctions ni pourquoi elles ont en majorité un partenaire commun ; la protéine porteuse de chaînes d'acyl mitochondriales (ACPM, acyl carrier protein en anglais).

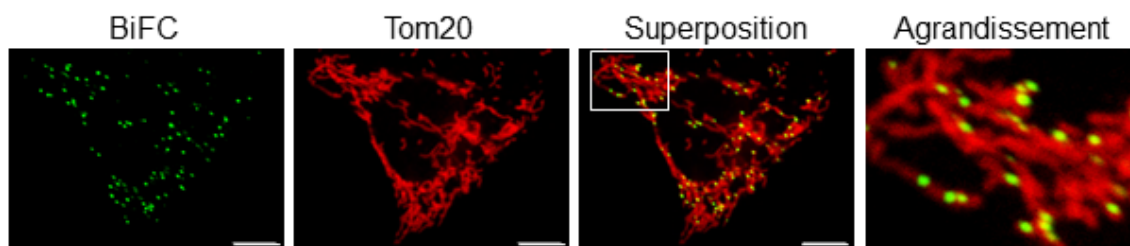


FIGURE 3.2 – Localisation subcellulaire de l'interaction entre altMiD51 et ACPm par complémentation bimoléculaire fluorescente (BiFC).

Des cellules HeLa transfectées avec une protéine fusion altMiD51(2flag)-Vénus(1-173) et une protéine fusion ACPm-Vénus(155-239) (vert). Les mitochondries ont été marquées avec un anticorps Tom20 par immunofluorescence soulignant les mitochondries (rouge). L'interaction entre altMiD51 et ACPm se retrouvent dans des *foci* mitochondriaux. Échelle de 10 μ M.

ACPm est une protéine adaptatrice permettant l'allongement des chaînes d'acyl dans le métabolisme mitochondrial des acides gras. ACPm est retrouvée soluble dans la matrice mitochondriale (Cronan *et al.*, 2005) ou associée au complexe 1 de la chaîne respiratoire en interagissant avec LYRM6 et LYRM3 (Runswick *et al.*, 1991). La mutation du domaine LYR de LYRM6 abolit l'interaction avec ACPm et alors diminue la présence d'ACPm dans le complexe, dérangeant l'activité du complexe 1. La raison d'être de l'interaction entre LYRM6 et LYRM3 avec ACPm au niveau du complexe 1 est encore inconnue. ACPm est connue pour interagir avec la plupart des protéines à domaine LYR dans des complexes, dont LYRM4, LYRM3, LYRM6 et FMC1.

La protéine ACPm a été détectée plusieurs fois dans des expériences d'immunoprécipitation d'altMiD51 par spectrométrie de masse (résultats non présentés), ce qui suggère qu'elle interagit avec altMiD51. Pour valider cette interaction, la technique du BiFC (*Bimolecular Fluorescence Complementation* en anglais) a été utilisée (Figure 3.2). Cette technique consiste à prendre la protéine fluorescente Vénus scindée en deux et ces moitiés sont fusionnées à une protéine d'intérêt (par exemple altMiD51 et ACPm). Lorsque les deux protéines d'intérêt interagissent, les moitiés de Vénus vont se retrouver et la protéine Vénus maintenant entière émettra de la fluorescence. Le résultat de cette expérience démontre que l'interaction entre altMiD51 et ACPm est possible et que celle-ci forme des foci au niveau de la mitochondrie. Cela implique que l'interaction n'est pas diffuse dans toute la mitochondrie, mais plutôt localisée à des endroits bien précis de la mitochondrie. Ce pourrait être une localisation dans un complexe mitochondrial connu ou inconnu. Nous avons regardé la colocalisation entre ces foci et les granules d'ARN, les nucléoïdes d'ADN et les sites de fission. La colocalisation fut très partielle ou nulle (résultats non présen-

tés). Nous ignorons encore où se retrouve leur interaction. De plus, lorsque la localisation d'altMiD51 ou de ACPm est obtenue par immunofluorescence grâce à leurs étiquettes respectives, un marquage diffus dans la matrice mitochondriale est obtenu pour les deux protéines. Par conséquent, il existe une population d'ACPM et d'altMiD51 libre (hors complexe) et une fraction se retrouvent en complexe. Ces résultats indiquent qu'altMiD51 agit comme une protéine à domaine LYR et interagit avec ACPm. Il faudrait confirmer l'interaction par co-immunoprécipitation et déterminer si les deux protéines sont logées dans un gros complexe comme d'autres protéines à domaine LYR dans la chaîne respiratoire.

3.3 La coopération fonctionnelle entre les protéines alternatives et leurs références

Au laboratoire, nous avançons l'hypothèse que les protéines codées à partir d'un même ARNm coopèrent dans la même voie cellulaire. Cela pourrait révéler chez les eucaryotes une régulation plus rigoureuse, une économie d'énergie et de ressources au niveau de la traduction pour le fonctionnement de la cellule, dans un même espace-temps et dans des proportions spécifiques (Vanderperre, 2013). Comme discuté précédemment, certains altORFs peuvent contenir la même signature moléculaire que le RefORF contenu dans le même ARNm, ce qui suggère que les deux protéines ont des fonctions connexes (Figure 2.4C, supplémentaire 21). C'est en fait le cas pour plusieurs ARNm codant pour des protéines à domaine doigt de zinc qui sont des facteurs de transcription.

Les protéines à doigt de zinc font partie des protéines les plus abondantes du protéome eucaryote présentement annoté, soit environ 9% chez l'homme (Figure supplémentaire 10F). Elles sont impliquées dans la reconnaissance, l'emballage et l'activation de la transcription de l'ADN (Laity *et al.*, 2001). Nos résultats montrent que 89 des 110 protéines alternatives possédant un domaine doigt de zinc sont codées dans un ARNm dont le CDS a également un domaine doigt de zinc (Figure supplémentaire 10F). Puisqu'elles partagent la même signature moléculaire, ces protéines alternatives et de références pourraient coopérer dans la régulation de la transcription de l'ADN. Cette découverte montre une fois de plus comment l'exploitation de la traduction dans les trois cadres de lecture enrichit le protéome. Ces ARNms multicotants pour des protéines à doigts de zinc pourraient être étudiés pour mettre de l'avant l'importance du potentiel multicotant d'un ARNm. Également, ils seraient des candidats d'étude de la coopération fonctionnelle entre les protéines alternatives et de références dans la transcription de l'ADN. Les facteurs de transcription à doigt de zinc sont connus pour dimériser entre eux (Laity *et al.*, 2001). Alors, regarder la dimérisation entre la protéine de référence et la protéine alternative serait une bonne

façon de commencer l'étude de ces candidats. Il est possible de regarder la dimérisation des protéines à doigt de zinc par le système double hybride dans la levure ou par co-immunoprécipitation.

Le deuxième objectif du projet fut de démontrer par un exemple biologique qu'il existe une coopération entre les protéines alternatives et de références. Cet objectif fut réalisé par l'étude d'altMiD51 et de MiD51. Puisque MiD51 est bien caractérisée et que l'on connaît son implication dans la morphologie mitochondriale, il fut possible de mesurer et d'évaluer la coopérativité entre les deux protéines

3.4 Deux protéines encodées par le même ARNm sont impliquées dans la fission mitochondriale

Le premier objectif du projet fut de caractériser une protéine alternative et démontrer que l'altORF contenu dans le même ARNm qu'un CDS peut être exprimé et avoir une fonction (Figure 2.3A et B). Il a été démontré que la protéine alternative altMiD51 est exprimée en contexte avec la protéine de référence MiD51. Aussi, les deux protéines ont été localisées à la mitochondrie. Le deuxième objectif du projet concernant la coopération fonctionnelle a débuté lorsque les résultats de surexpression d'altMiD51 dans les cellules HeLa démontraient un changement dans la morphologie des mitochondries (Figure 2.3C). Ainsi, altMiD51 avait un lien avec la morphologie mitochondriale, tout comme la protéine de référence MiD51. Des décomptes de la morphologie mitochondriale ont permis de découvrir que sa surexpression induit la fragmentation des mitochondries (Figure 2.3D). Parallèlement, MiD51 est un récepteur recrutant Drp1, une GTPase très bien connue pour cliver les mitochondries.

3.4.1 La fission mitochondriale Drp1 dépendante ou indépendante

La protéine MiD51 est connue pour recruter Drp1 à la surface des mitochondries et ainsi engendrer l'oligomérisation de Drp1 et la fission mitochondriale. Premièrement, il fallait déterminer si la surexpression d'altMiD51 engendre la fission mitochondriale via Drp1 (Drp1-dépendante). Pour ce faire, la morphologie des mitochondries a été évaluée lorsque l'expression de Drp1 est diminuée par de petits ARNs interférents (Figure supplémentaire 20) ou lorsque son activité est diminuée par un mutant dominant négatif (Figure 2.3D, supplémentaire 18). Le résultat montre que la fission mitochondriale résultant de la surexpression d'altMiD51 est diminuée par l'activité réduite de Drp1, mais ne l'est pas totalement. Il est probable qu'altMiD51 contribue aussi à la fission mitochondriale via un mécanisme Drp1-indépendant à l'intérieur de la mitochondrie.

Un changement métabolique ou une perturbation des protéines du dynamisme mitochon-

drial affecte la morphologie des mitochondries et ce paramètre est facilement mesurable par immunofluorescence avant d'entreprendre d'autres expériences plus complexes. Toutefois, le décompte mitochondrial réalisé présente un biais. Il est subjectif puisqu'il dépend du classement des mitochondries qu'effectue l'observateur selon sa perception des mitochondries. Il existe des logiciels comme microP (Peng *et al.*, 2011) qui permettent de quantifier les morphologies mitochondriales pour chaque cellule, mais aussi pour chaque expérience sur la population globale. Il serait pertinent de mettre au point l'utilisation de ce logiciel au laboratoire pour l'étude de protéines alternatives mitochondriales. Pour comprendre davantage l'implication de Drp1 dans la fission par altMiD51, le niveau de Drp1 phosphorylé à la sérine 616 lors de la surexpression d'altMiD51 a été quantifié par immunobuvardage de type *Western* par un anticorps spécifique à cette phosphorylation (Figure 3.3A et B). La phosphorylation de la sérine 616 (P-S616) est une modification post-traductionnelle de Drp1 qui augmente son activité GTPase (Taguchi *et al.*, 2007). Le résultat de l'immunobuvardage de type *Western* dont les bandes ont été quantifiées par le logiciel ImageJ indique que les niveaux de Drp1 P-S616 ne changent pas de manière significative entre la surexpression de altMiD51 et celle d'un vecteur vide (Figure 3.3A et B). Cependant, la surexpression d'altMiD51 n'augmente pas la capacité de Drp1 à fissionner les mitochondries. Toutefois, MiD51 est connue pour le recrutement de Drp1 P-S637 et réguler la fission via le cofacteur ADP liant MiD51. AltMiD51 pourrait être impliquée dans la disponibilité du cofacteur ADP dans la matrice mitochondriale afin que MiD51 induise la fission par Drp1. Bien que le niveau de Drp1 P-S616 ne change pas significativement lors de la surexpression d'altMiD51, les résultats montrent une augmentation dans le niveau total d'expression de Drp1 (Figure 3.3B). D'une part, on pourrait supposer qu'altMiD51 augmente indirectement le niveau d'expression de Drp1 pour faciliter la fission mitochondriale.

D'autre part, l'expression de Mfn2 et Opa1, deux protéines importantes dans la fusion mitochondriale a été investiguée par immunobuvardage de type *Western*. La fission mitochondriale induite par la surexpression d'altMiD51 pourrait provenir d'une diminution d'expression ou d'activité d'un de ces joueurs clés dans la fusion mitochondriale, le processus inverse de la fission. Le clivage d'Opa1 (sa forme courte) est connu pour empêcher la fusion des membranes internes des mitochondries (Ishihara *et al.*, 2006). Aussi, l'ubiquitination de la protéine Mfn2 est responsable de la diminution des événements de fusion de la membrane externe et entraîne une perte d'équilibre entre les événements de fusion/fission (Escobar-Henriques et Anton, 2013). Cela déplace la morphologie mitochondriale vers la fission. Concernant ces deux protéines, mes résultats indiquent qu'il n'y a pas de clivage d'Opa1 (Figure 3.3C) ni de diminution de l'expression de Mfn2

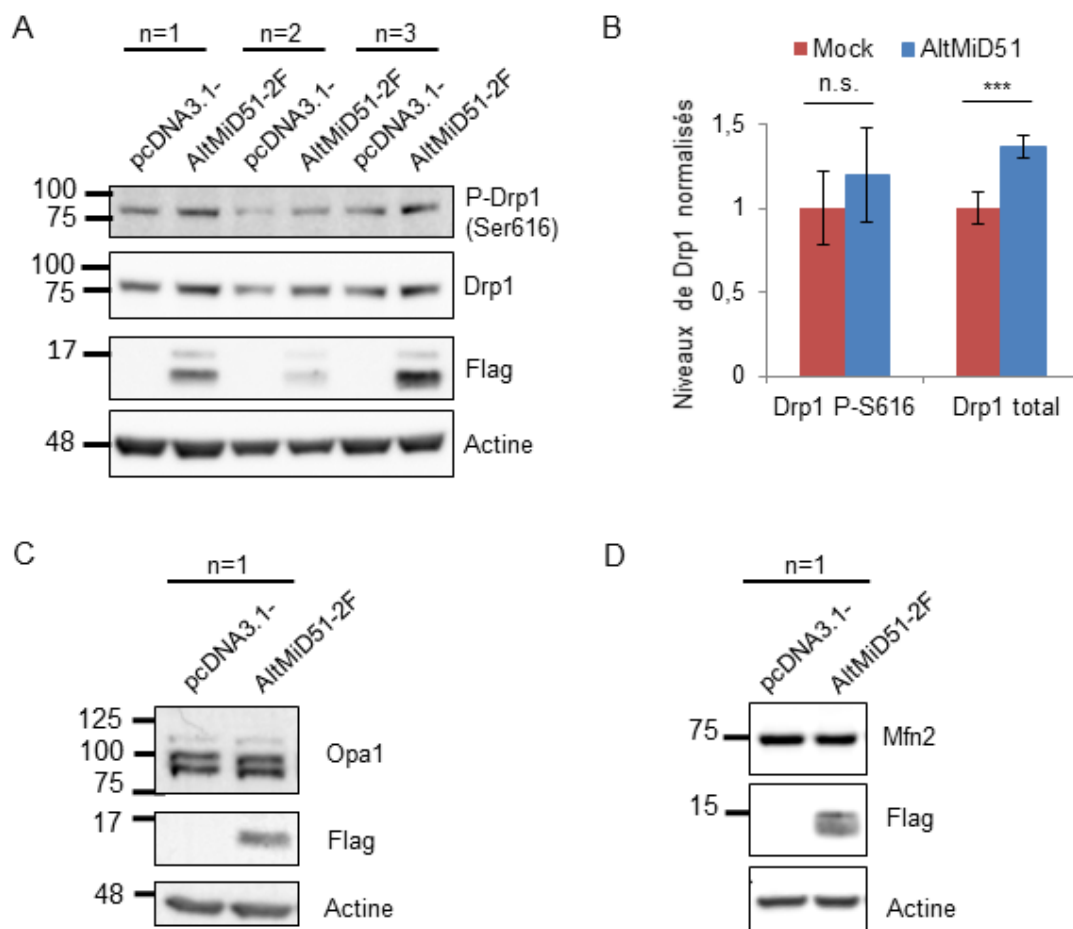


FIGURE 3.3 – L'influence de la surexpression d'altMiD51 sur les GTPases impliquées dans le dynamisme mitochondrial.

(A-D) Transfection 24h de cellules HeLa avec un vecteur vide ou altMiD51-2xflag (A) Immunobuvardage de type *Western* contre Drp1 total ou phosphorylé à la sérine 616 (P-S616). Le Drp1 total a été réalisé sur la même membrane que la forme P-S616 après «stripping» des anticorps. n=3. (B) Quantification avec le logiciel ImageJ de l'immunobuvardage fait en (A). Les niveaux de P-S616 ont été calculés comme suit : P-S616/Drp1 total/actine. Les niveaux de Drp1 total ont été normalisés sur l'actine. Les ratios ont été normalisés par rapport au vecteur vide. ***p=0.0003. (C et D) Immunobuvardage de type *Western* contre Opa1 et Mfn2 respectivement. Expérience fait une seule fois. (A,C,D) L'anticorps Flag révèle que la transfection d'altMiD51-2xflag a fonctionné et l'actine démontre que la quantité de protéine sur la membrane est semblable entre les différentes conditions.

(Figure 3.3D). Les niveaux d'expression entre la surexpression d'altMiD51 et le vecteur vide sont similaires à l'œil. Ainsi, la fission entraînée par la surexpression d'altMiD51 ne semble pas influencer les GTPases impliquées dans la fusion mitochondriale. La fission causée par altMiD51 pourrait provenir de l'augmentation de l'expression de Drp1 ou d'un mécanisme encore inconnu.

Avant de continuer dans des analyses plus poussées de la fission mitochondriale et de la coopération entre altMiD51 et MiD51, il serait important de déterminer si la fission par altMiD51 est dépendante de l'action de MiD51 par invalidation génique des deux protéines par la méthode CRISPR. Si la fission mitochondriale provoquée par la surexpression d'altMiD51 est possible lorsque MiD51 n'est pas présente dans les cellules HeLa, cela veut dire qu'altMiD51 peut induire la fission par elle-même sans nécessiter le recrutement de Drp1 par MiD51. De plus, il serait intéressant de voir si la perte d'altMiD51 dans les cellules déplace la morphologie mitochondriale vers la fusion, ce qui confirmerait l'implication d'altMiD51 dans la fission mitochondriale .

3.4.1.1 *Le CRISPR d'altMiD51/MiD51*

La surexpression est un moyen facile d'explorer des phénotypes cellulaires dans l'étude d'une nouvelle protéine. Pour confirmer un phénotype, il est en revanche primordial d'observer le phénotype inverse lors de l'invalidation génique de cette même protéine. Il est possible d'utiliser des petits ARN interférents visant l'ARNm entier pour étudier le phénotype lors d'une diminution d'expression d'une protéine alternative. Ensuite, Il suffit de restituer l'expression de la protéine de référence ou de la protéine alternative pour comprendre leurs rôles singuliers dans la cellule. La stratégie de l'ARN interférant n'est pas possible dans le cas de l'étude des protéines alternatives car la dégradation d'un ARNm inhibe simultanément l'expression des protéines de référence et de leur protéines alternatives (Vanderperre *et al.*, 2011; Bergeron *et al.*, 2013).

Pour bien étudier la fonction des protéines de référence et des protéines alternatives séparément, il est plus rigoureux de modifier le génome. Ainsi, on peut changer le codon initiateur de l'altORF par un codon terminateur dans le génome, ce qui inhibera l'expression de la protéine alternative sans affecter l'expression de la protéine de référence. Le laboratoire a récemment généré des cellules HeLa «Knockout (KO)» pour altMID51, MID51, et double KO par la méthode CRISPR. Cette méthode utilise l'endonucléase Cas9 pour induire des cassures double brin à un endroit défini de l'ADN pour y introduire un gène ou une mutation.

3.4.2 *La fission engendrée par la surexpression d'altMiD51 n'affecte pas le métabolisme mitochondrial*

Comme décrit précédemment, un lien étroit a été établi entre la morphologie mitochondriale et l'état bioénergétique de la mitochondrie (section 1.6.2). En effet, la mitochondrie change sa morphologie pour rétablir son homéostasie. Une perturbation des mitochondries peut engendrer la fragmentation du réseau mitochondrial. Pour déterminer si la surexpression d'altMiD51 a un impact sur la bioénergétique des mitochondries et pourrait ainsi causer leur fragmentation, j'ai mesuré différents paramètres mitochondriaux tels que le fonctionnement de la chaîne respiratoire, la production d'espèces d'oxygènes réactives (ROS) et la production d'ATP (Figure supplémentaire 17B et C). Aucune différence entre un vecteur vide et une surexpression d'altMiD51 pour ce qui concerne la production d'ATP et la production de ROS n'a été observée. Cela signifie que la fragmentation par altMiD51 ne provient pas d'un stress relié à l'augmentation de ROS ou d'une diminution de production d'énergie par la phosphorylation oxydative.

En ce qui concerne la respiration, altMiD51 n'influence pas la respiration basale, la respiration maximale, le flux de protons ni la production d'ATP (Figure supplémentaire 17A). En revanche, sa surexpression augmente la capacité respiratoire de réserve (reserve respiratory capacity en anglais). Ce paramètre représente l'ATP supplémentaire produite par phosphorylation oxydative dans des conditions où la cellule a un besoin plus important en énergie. Une diminution de cette capacité à produire de l'énergie supplémentaire est associée à la sensibilité des cellules à la sénescence et elle est compromise dans plusieurs pathologies comme des maladies cardiaques et neurodégénératives (Desler *et al.*, 2012). Il a été observé que des fibroblastes résistants à la mort cellulaire avaient une augmentation de leur capacité respiratoire de réserve (Nickens *et al.*, 2013). Donc, il semblerait qu'altMiD51 apporte un avantage adaptatif pour la cellule, sans affecter la condition générale de la mitochondrie. De plus, il s'avère que MiD51 régule le relâchement du cytochrome c durant la mort cellulaire (Otera *et al.*, 2016). Considérant la possible coopération entre altMiD51 et MiD51, on pourrait émettre l'hypothèse qu'altMiD51 apporte une résistance à la mort cellulaire par une augmentation de la capacité respiratoire de réserve.

4 CONCLUSIONS

En lien avec les objectifs présentés dans la partie 1.7 de l'introduction, les conclusions de mes travaux de recherche de maîtrise sont les suivantes :

- Le gène *SMCR7L* est traduit en ARNm multicodant pour la protéine de référence MiD51 et la protéine alternative altMiD51. Aussi, d'autres altORFs d'au moins 30 codons sont contenus dans l'ARNm et peuvent potentiellement être traduits. J'ai démontré qu'altMiD51 peut être co-exprimée avec MiD51 par un système hétérologue d'expression *in cellulo*. J'ai démontré qu'altMiD51 est une protéine à domaine LYR de la matrice mitochondriale conservée chez les vertébrés. L'expression endogène d'altMiD51 peut être détectée par spectrométrie de masse et profilage ribosomal.
- AltMiD51 peut avoir un lien fonctionnel avec MiD51 qui est un récepteur de Drp1 dans le processus de fission mitochondrial. La surexpression d'altMiD51 entraîne la fragmentation des mitochondries, mais pas totalement de manière Drp1-dépendante. Même si le mécanisme Drp1-indépendant est encore inconnu pour altMiD51, la fission causée par sa surexpression n'influence pas l'état bioénergétique de la mitochondrie, mis à part la capacité respiratoire de réserve.

Finalement, plusieurs perspectives pour ce projet ont été discutées :

- Prouver qu'AltMiD51 et MiD51 sont sur le même ARNm par quantification absolue par spectrométrie de masse.
- Déterminer si la fission mitochondriale dépend de l'action de MiD51 ou d'altMiD51 par l'observation de la morphologie mitochondriale des lignées de cellule HeLa *knockout* pour ces deux protéines individuellement.
- Confirmer par coimmunoprécipitation l'interaction entre altMiD51 et ACPm, puis déterminer si elles sont dans un complexe relié aux centres fer-souffre.
- Déterminer pourquoi altMiD51 entraîne la fission mitochondriale et identifier les acides aminés responsables de sa fonction.

BIBLIOGRAPHIE

- Abouleish, E. I. (1987) Are local anesthetics equal?. *Anesthesia & Analgesia*, 66(12) : 1339–1340.
- Aebersold, R. et Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, 422(6928) : 198–207.
- Andreev, D. E., O'Connor, P. B., Fahey, C., Kenny, E. M., Terenin, I. M., Dmitriev, S. E., Cormican, P., Morris, D. W., Shatsky, I. N., et Baranov, P. V. (2015) Translation of 5' leaders is pervasive in genes resistant to eif2 repression. *Elife*, 4 : e03971.
- Andrews, S. J. et Rothnagel, J. A. (2014) Emerging evidence for functional peptides encoded by short open reading frames. *Nature Reviews Genetics*, 15(3) : 193–204.
- Angerer, H. (2013) The superfamily of mitochondrial complex1_lyr motif-containing (lyrm) proteins. *Biochemical Society Transactions*, 41(5) : 1335–1341.
- Angerer, H. (2015) Eukaryotic lyr proteins interact with mitochondrial protein complexes. *Biology*, 4(1) : 133–150.
- Angerer, H., Radermacher, M., Mańkowska, M., Steger, M., Zwicker, K., Heide, H., Wittig, I., Brandt, U., et Zickermann, V. (2014) The lyr protein subunit nb4m/ndufa6 of mitochondrial complex i anchors an acyl carrier protein and is essential for catalytic activity. *Proceedings of the National Academy of Sciences*, 111(14) : 5207–5212.
- Antonny, B., Burd, C., De Camilli, P., Chen, E., Daumke, O., Faelber, K., Ford, M., Frolov, V. A., Frost, A., Hinshaw, J. E., et al. (2016) Membrane fission by dynamin : what we know and what we need to know. *The EMBO Journal*, page e201694613.
- Attardi, G. et Amaldi, F. (1970) Structure and synthesis of ribosomal rna. *Annual review of biochemistry*, 39(1) : 183–226.
- Barbosa, C., Peixeiro, I., et Romão, L. (2013) Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet*, 9(8) : e1003529.
- Basrai, M. A., Hieter, P., et Boeke, J. D. (1997) Small open reading frames : beautiful needles in the haystack. *Genome research*, 7(8) : 768–771.

- Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., Vejnar, C. E., Lee, M. T., Rajewsky, N., Walther, T. C., *et al.* (2014) Identification of small orfs in vertebrates using ribosome footprinting and evolutionary conservation. *The EMBO journal*, page e201488411.
- Bergeron, D., Lapointe, C., Bissonnette, C., Tremblay, G., Motard, J., et Roucou, X. (2013) An out-of-frame overlapping reading frame in the ataxin-1 coding sequence encodes a novel ataxin-1 interacting protein. *Journal of Biological Chemistry*, 288(30) : 21824–21835.
- Bi, W., Yan, J., Stankiewicz, P., Park, S.-S., Walz, K., Boerkoel, C. F., Potocki, L., Shaffer, L. G., Devriendt, K., Nowaczyk, M. J., *et al.* (2002) Genes in a refined smith-magenis syndrome critical deletion interval on chromosome 17p11. 2 and the syntenic region of the mouse. *Genome research*, 12(5) : 713–728.
- Blumenthal, T. (2004) Operons in eukaryotes. *Briefings in functional genomics & proteomics*, 3(3) : 199–211.
- Bolinger, C. et Boris-Lawrie, K. (2009) Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. *Retrovirology*, 6(1) : 1.
- Cabrera-Quio, L. E., Herberg, S., et Pauli, A. (2016) Decoding sorf translation—from small proteins to gene regulation. *RNA biology*, (just-accepted) : 00–00.
- Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler, M., Obermayer, B., et Ohler, U. (2015) Detecting actively translated open reading frames in ribosome profiling data. *Nature methods*.
- Calvo, S. E., Pagliarini, D. J., et Mootha, V. K. (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proceedings of the National Academy of Sciences*, 106(18) : 7507–7512.
- Cassidy, L., Prasse, D., Linke, D., Schmitz, R. A., et Tholey, A. (2016) Combination of bottom up 2d-lc-ms and semi-top-down gelfree-lc-ms enhances coverage of proteome and low molecular weight short open reading frame encoded peptides of the archaeon *methanosarcina mazei*. *Journal of Proteome Research*.
- Chang, C.-R. et Blackstone, C. (2007) Cyclic amp-dependent protein kinase phosphorylation of drp1 regulates its gtpase activity and mitochondrial morphology. *Journal of Biological Chemistry*, 282(30) : 21583–21587.
- Chang, C.-R. et Blackstone, C. (2010) Dynamic regulation of mitochondrial fission

- through modification of the dynamin-related protein drp1. *Annals of the New York Academy of Sciences*, 1201(1) : 34–39.
- Chen, H., Detmer, S. A., Ewald, A. J., Griffin, E. E., Fraser, S. E., et Chan, D. C. (2003) Mitofusins mfn1 and mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *The Journal of cell biology*, 160(2) : 189–200.
- Chew, G.-L., Pauli, A., et Schier, A. F. (2016) Conservation of uorf repressiveness and sequence features in mouse, human and zebrafish. *Nature communications*, 7.
- Chng, S. C., Ho, L., Tian, J., et Reversade, B. (2013) Elabela : a hormone essential for heart development signals via the apelin receptor. *Developmental cell*, 27(6) : 672–680.
- Chung, W.-Y., Wadhawan, S., Szklarczyk, R., Pond, S. K., et Nekrutenko, A. (2007) A first look at arfome : dual-coding genes in mammalian genomes. *PLoS Comput Biol*, 3(5) : e91.
- Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K., et Lander, E. S. (2007) Distinguishing protein-coding and noncoding genes in the human genome. *Proceedings of the National Academy of Sciences*, 104(49) : 19428–19433.
- Cronan, J. E., Fearnley, I. M., et Walker, J. E. (2005) Mammalian mitochondria contain a soluble acyl carrier protein. *FEBS letters*, 579(21) : 4892–4896.
- Desler, C., Hansen, T. L., Frederiksen, J. B., Marcker, M. L., Singh, K. K., et Juel Rasmussen, L. (2012) Is there a link between mitochondrial reserve respiratory capacity and aging? *Journal of aging research*, 2012.
- Dorsam, R. T. et Gutkind, J. S. (2007) G-protein-coupled receptors and cancer. *Nature reviews cancer*, 7(2) : 79–94.
- Duncan, M. W., Aebersold, R., et Caprioli, R. M. (2010) The pros and cons of peptide-centric proteomics [au : Ok?]. *Nature biotechnology*, 28(7) : 1.
- Dunham, I., Hunt, A., Collins, J., Bruskiewich, R., Beare, D., Clamp, M., Smink, L., Ainscough, R., Almeida, J., Babbage, A., et al. (1999) The dna sequence of human chromosome 22. *Nature*, 402(6761) : 489–495.
- Elgass, K. D., Smith, E. A., LeGros, M. A., Larabell, C. A., et Ryan, M. T. (2015) Analysis of er-mitochondria contacts using correlative fluorescence microscopy and soft x-ray tomography of mammalian cells. *J Cell Sci*, 128(15) : 2795–2804.

- Escobar-Henriques, M. et Anton, F. (2013) Mechanistic perspective of mitochondrial fusion : tubulation vs. fragmentation. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1833(1) : 162–175.
- Fortino, V., Smolander, O.-P., Auvinen, P., Tagliaferri, R., et Greco, D. (2014) Transcriptome dynamics-based operon prediction in prokaryotes. *BMC bioinformatics*, 15(1) : 1.
- Freson, K., Jaeken, J., Van Helvoirt, M., de Zegher, F., Wittevrongel, C., Thys, C., Hoylaerts, M. F., Vermylen, J., et Van Geet, C. (2003) Functional polymorphisms in the paternally expressed *xI α s* and its cofactor *alex* decrease their mutual interaction and enhance receptor-mediated camp formation. *Human molecular genetics*, 12(10) : 1121–1130.
- Friedman, J. R., Lackner, L. L., West, M., DiBenedetto, J. R., Nunnari, J., et Voeltz, G. K. (2011) Er tubules mark sites of mitochondrial division. *Science*, 334(6054) : 358–362.
- Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., Schreiber, S., Platzer, M., Krawczak, M., Hampe, J., *et al.* (2012) Genome-wide search for novel human uorfs and n-terminal protein extensions using ribosomal footprinting. *Genome research*, 22(11) : 2208–2218.
- Gao, X., Wan, J., Liu, B., Ma, M., Shen, B., et Qian, S.-B. (2015) Quantitative profiling of initiating ribosomes in vivo. *Nature methods*, 12(2) : 147–153.
- Ghilardi, N., Wiestner, A., et Skoda, R. C. (1998) Thrombopoietin production is inhibited by a translational mechanism. *Blood*, 92(11) : 4023–4030.
- Goodstadt, L. et Ponting, C. P. (2006) Phylogenetic reconstruction of orthology, paralogy, and conserved syntenies for dog and human. *PLoS Comput Biol*, 2(9) : e133.
- Hornig, N. C., De Beaufort, C., Denzer, F., Cools, M., Wabitsch, M., Ukat, M., Kulle, A. E., Schweikert, H.-U., Werner, R., Hiort, O., *et al.* (2016) A recurrent germline mutation in the 5'utr of the androgen receptor causes complete androgen insensitivity by activating aberrant uorf translation. *PloS one*, 11(4) : e0154158.
- Ingolia, N. T. (2010) Genome-wide translational profiling by ribosome footprinting. *Methods in enzymology*, 470 : 119–142.
- Ingolia, N. T. (2014) Ribosome profiling : new views of translation, from single codons to genome scale. *Nature Reviews Genetics*, 15(3) : 205–213.

- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., et Weissman, J. S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *science*, 324(5924) : 218–223.
- Ishihara, N., Fujita, Y., Oka, T., et Mihara, K. (2006) Regulation of mitochondrial morphology through proteolytic cleavage of opa1. *The EMBO journal*, 25(13) : 2966–2977.
- Ji, Z., Song, R., Regev, A., et Struhl, K. (2015) Many lncrnas, 5'utrs, and pseudogenes are translated and some are likely to express functional proteins. *Elife*, 4 : e08890.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014) Interproscan 5 : genome-scale protein function classification. *Bioinformatics*, 30(9) : 1236–1240.
- Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Dutttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermüller, J., Hofacker, I. L., et al. (2007) Rna maps reveal new rna classes and a possible function for pervasive transcription. *Science*, 316(5830) : 1484–1488.
- Kastenmayer, J. P., Ni, L., Chu, A., Kitchen, L. E., Au, W.-C., Yang, H., Carter, C. D., Wheeler, D., Davis, R. W., Boeke, J. D., et al. (2006) Functional genomics of genes with small open reading frames (sorfs) in *s. cerevisiae*. *Genome research*, 16(3) : 365–373.
- Kim, I., Rodriguez-Enriquez, S., et Lemasters, J. J. (2007) Selective degradation of mitochondria by mitophagy. *Archives of biochemistry and biophysics*, 462(2) : 245–253.
- Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., et al. (2014) A draft map of the human proteome. *Nature*, 509(7502) : 575–581.
- Kiss, T. (2002) Small nucleolar rnas : an abundant group of noncoding rnas with diverse cellular functions. *Cell*, 109(2) : 145–148.
- Klemke, M., Kehlenbach, R. H., et Huttner, W. B. (2001) Two overlapping reading frames in a single exon encode interacting proteins—a novel way of gene usage. *The EMBO journal*, 20(14) : 3849–3860.
- Koirala, S., Guo, Q., Kalia, R., Bui, H. T., Eckert, D. M., Frost, A., et Shaw, J. M. (2013) Interchangeable adaptors regulate mitochondrial dynamin assembly for membrane scission. *Proceedings of the National Academy of Sciences*, 110(15) : E1342–E1351.
- Kovacs, E., Tompa, P., Liliom, K., et Kalmar, L. (2010) Dual coding in alternative reading

- frames correlates with intrinsic protein disorder. *Proceedings of the National Academy of Sciences*, 107(12) : 5429–5434.
- Kozak, M. (1978) How do eucaryotic ribosomes select initiation regions in messenger rna? *Cell*, 15(4) : 1109–1123.
- Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger rnas. *Nucleic acids research*, 15(20) : 8125–8148.
- Kozak, M. (1989) The scanning model for translation : an update. *The Journal of cell biology*, 108(2) : 229–241.
- Kozak, M. (1997) Recognition of aug and alternative initiator codons is augmented by g in position+ 4 but is not generally affected by the nucleotides in positions+ 5 and+ 6. *The EMBO journal*, 16(9) : 2482–2492.
- Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene*, 234(2) : 187–208.
- Kozak, M. (2001) Constraints on reinitiation of translation in mammals. *Nucleic Acids Research*, 29(24) : 5226–5232.
- Laity, J. H., Lee, B. M., et Wright, P. E. (2001) Zinc finger proteins : new insights into structural and functional diversity. *Current opinion in structural biology*, 11(1) : 39–46.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, 409(6822) : 860–921.
- Lee, C.-f., Lai, H.-L., Lee, Y.-C., Chien, C.-L., et Chern, Y. (2014) The a2a adenosine receptor is a dual coding gene a novel mechanism of gene usage and signal transduction. *Journal of Biological Chemistry*, 289(3) : 1257–1270.
- Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., et Qian, S.-B. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proceedings of the National Academy of Sciences*, 109(37) : E2424–E2432.
- Lewis, S. C., Uchiyama, L. F., et Nunnari, J. (2016) Er-mitochondria contacts couple mtdna synthesis with mitochondrial division in human cells. *Science*, 353(6296) : aaf5549.
- Li, C., Goudy, K., Hirsch, M., Asokan, A., Fan, Y., Alexander, J., Sun, J., Monahan, P., Seiber, D., Sidney, J., et al. (2009) Cellular immune response to cryptic epitopes during

- therapeutic gene transfer. *Proceedings of the National Academy of Sciences*, 106(26) : 10770–10774.
- Liang, H. et Landweber, L. F. (2006) A genome-wide study of dual coding regions in human alternatively spliced genes. *Genome research*, 16(2) : 190–196.
- Liesa, M., Palacín, M., et Zorzano, A. (2009) Mitochondrial dynamics in mammalian health and disease. *Physiological reviews*, 89(3) : 799–845.
- Lim, S. C., Friemel, M., Marum, J. E., Tucker, E. J., Bruno, D. L., Riley, L. G., Christodoulou, J., Kirk, E. P., Boneh, A., DeGennaro, C., *et al.* (2013) Mutations in *lyrm4*, encoding iron-sulfur cluster biogenesis factor *isd11*, cause deficiency of multiple respiratory chain complexes. *Human molecular genetics*, page ddt295.
- Liu, B., Han, Y., et Qian, S.-B. (2013) Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. *Molecular cell*, 49(3) : 453–463.
- Liu, J., Yosten, G. L., Ji, H., Zhang, D., Zheng, W., Speth, R. C., Samson, W. K., et Sandberg, K. (2014) Selective inhibition of angiotensin receptor signaling through *erk1/2* pathway by a novel peptide. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 306(8) : R619–R626.
- Losón, O. C., Liu, R., Rome, M. E., Meng, S., Kaiser, J. T., Shan, S.-o., et Chan, D. C. (2014) The mitochondrial fission receptor *mid51* requires *adp* as a cofactor. *Structure*, 22(3) : 367–377.
- Losón, O. C., Song, Z., Chen, H., et Chan, D. C. (2013) *Fis1*, *mff*, *mid49*, and *mid51* mediate *drp1* recruitment in mitochondrial fission. *Molecular biology of the cell*, 24(5) : 659–667.
- MacVicar, T. et Langer, T. (2016) *Opa1* processing in cell death and disease—the long and short of it. *J Cell Sci*, pages jcs–159186.
- Mattick, J. S. et Makunin, I. V. (2006) Non-coding rna. *Human molecular genetics*, 15(suppl 1) : R17–R29.
- Michel, A. M., Ahern, A. M., Donohue, C. A., et Baranov, P. V. (2015) *Gwips-viz* as a tool for exploring ribosome profiling evidence supporting the synthesis of alternative proteoforms. *Proteomics*, 15(14) : 2410–2416.
- Michel, A. M., Choudhury, K. R., Firth, A. E., Ingolia, N. T., Atkins, J. F., et Baranov, P. V. (2012) Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome research*, 22(11) : 2219–2229.

- Michel, A. M., Fox, G., Kiran, A. M., De Bo, C., O'Connor, P. B., Heaphy, S. M., Mullan, J. P., Donohue, C. A., Higgins, D. G., et Baranov, P. V. (2014) Gwips-viz : development of a ribo-seq genome browser. *Nucleic acids research*, 42(D1) : D859–D864.
- Mishra, P. et Chan, D. C. (2014) Mitochondrial dynamics and inheritance during cell division, development and disease. *Nature reviews Molecular cell biology*, 15(10) : 634–646.
- Mouilleron, H., Delcourt, V., et Roucou, X. (2016) Death of a dogma : eukaryotic mnas can code for more than one protein. *Nucleic acids research*, 44(1) : 14–23.
- Na, D., Son, H., et Gsponer, J. (2014) Categorizer : a tool to categorize genes into user-defined biological groups based on semantic similarity. *BMC genomics*, 15(1) : 1.
- Nam, J.-W., Choi, S.-W., et You, B.-H. (2016) Incredible rna : Dual functions of coding and noncoding. *Molecules and cells*, 39(5) : 367.
- Nesvizhskii, A. I. (2010) A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of proteomics*, 73(11) : 2092–2123.
- Nickens, K. P., Wikstrom, J. D., Shirihai, O. S., Patierno, S. R., et Ceryak, S. (2013) A bioenergetic profile of non-transformed fibroblasts uncovers a link between death-resistance and enhanced spare respiratory capacity. *Mitochondrion*, 13(6) : 662–667.
- O'Carroll, A.-M., Lolait, S. J., Harris, L. E., et Pope, G. R. (2013) The apelin receptor apj : journey from an orphan to a multifaceted regulator of homeostasis. *Journal of Endocrinology*, 219(1) : R13–R35.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cdnas. *Nature*, 420(6915) : 563–573.
- Olexiouk, V. et Menschaert, G. (2016) Identification of small novel coding sequences, a proteogenomics endeavor. In *Proteogenomics*, pages 49–64. Springer.
- Otera, H., Ishihara, N., et Mihara, K. (2013) New insights into the function and regulation of mitochondrial fission. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1833(5) : 1256–1268.
- Otera, H., Miyata, N., Kuge, O., et Mihara, K. (2016) Drp1-dependent mitochondrial fission via mid49/51 is essential for apoptotic cristae remodeling. *The Journal of cell biology*, 212(5) : 531–544.

- Pain, V. (1986) Initiation of protein synthesis in mammalian cells. *Biochemical Journal*, 235(3) : 625.
- Palmer, C. S., Elgass, K. D., Parton, R. G., Osellame, L. D., Stojanovski, D., et Ryan, M. T. (2013) Adaptor proteins mid49 and mid51 can act independently of mff and fis1 in drp1 recruitment and are specific for mitochondrial fission. *Journal of Biological Chemistry*, 288(38) : 27584–27593.
- Palmer, C. S., Osellame, L. D., Laine, D., Koutsopoulos, O. S., Frazier, A. E., et Ryan, M. T. (2011a) Mid49 and mid51, new components of the mitochondrial fission machinery. *EMBO reports*, 12(6) : 565–573.
- Palmer, C. S., Osellame, L. D., Stojanovski, D., et Ryan, M. T. (2011b) The regulation of mitochondrial morphology : intricate mechanisms and dynamic machinery. *Cellular signalling*, 23(10) : 1534–1545.
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., et Blencowe, B. J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, 40(12) : 1413–1415.
- Peng, J.-Y., Lin, C.-C., Chen, Y.-J., Kao, L.-S., Liu, Y.-C., Chou, C.-C., Huang, Y.-H., Chang, F.-R., Wu, Y.-C., Tsai, Y.-S., *et al.* (2011) Automatic morphological subtyping reveals new roles of caspases in mitochondrial dynamics. *PLoS Comput Biol*, 7(10) : e1002212.
- Pinto, S. M., Manda, S. S., Kim, M.-S., Taylor, K., Selvan, L. D. N., Balakrishnan, L., Subbannayya, T., Yan, F., Prasad, T. K., Gowda, H., *et al.* (2014) Functional annotation of proteome encoded by human chromosome 22. *Journal of proteome research*, 13(6) : 2749–2760.
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., et Siepel, A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research*, 20(1) : 110–121.
- Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., et Koller, D. (2014) Causal signals between codon bias, mrna structure, and the efficiency of translation and elongation. *Molecular systems biology*, 10(12) : 770.
- Prabakaran, S., Hemberg, M., Chauhan, R., Winter, D., Tweedie-Cullen, R. Y., Dittrich, C., Hong, E., Gunawardena, J., Steen, H., Kreiman, G., *et al.* (2014) Quantitative profiling of peptides from rnas classified as noncoding. *Nature communications*, 5.

- Pueyo, J. I., Magny, E. G., et Couso, J. P. (2016) New peptides under the surface of the genome. *Trends in biochemical sciences*.
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et Lopez, R. (2005) Interproscan : protein domains identifier. *Nucleic acids research*, 33(suppl 2) : W116–W120.
- Remm, M., Storm, C. E., et Sonnhammer, E. L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of molecular biology*, 314(5) : 1041–1052.
- Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., et Ting, A. Y. (2013) Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, 339(6125) : 1328–1331.
- Ribrioux, S., Brünger, A., Baumgarten, B., Seuwen, K., et John, M. R. (2008) Bioinformatics prediction of overlapping frameshifted translation products in mammalian transcripts. *BMC genomics*, 9(1) : 1.
- Richter, V., Palmer, C. S., Osellame, L. D., Singh, A. P., Elgass, K., Stroud, D. A., Sesaki, H., Kvensakul, M., et Ryan, M. T. (2014) Structural and functional analysis of mid51, a dynamin receptor required for mitochondrial fission. *The Journal of cell biology*, 204(4) : 477–486.
- Rosenbaum, D. M., Rasmussen, S. G., et Kobilka, B. K. (2009) The structure and function of g-protein-coupled receptors. *Nature*, 459(7245) : 356–363.
- Rouault, T. A. (2015) Mammalian iron-sulphur proteins : novel insights into biogenesis and function. *Nature reviews Molecular cell biology*, 16(1) : 45–55.
- Runswick, M. J., Fearnley, I. M., Skehel, J. M., et Walker, J. E. (1991) Presence of an acyl carrier protein in nadh : ubiquinone oxidoreductase from bovine heart mitochondria. *FEBS letters*, 286(1-2) : 121–124.
- Savard, J., Marques-Souza, H., Aranda, M., et Tautz, D. (2006) A segmentation gene in *tribolium* produces a polycistronic mrna that codes for multiple conserved peptides. *Cell*, 126(3) : 559–569.
- Shteynberg, D., Nesvizhskii, A. I., Moritz, R. L., et Deutsch, E. W. (2013) Combining results of multiple search engines in proteomics. *Molecular & cellular proteomics*, 12(9) : 2383–2393.
- Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R., et Wiemann, S. (2000)

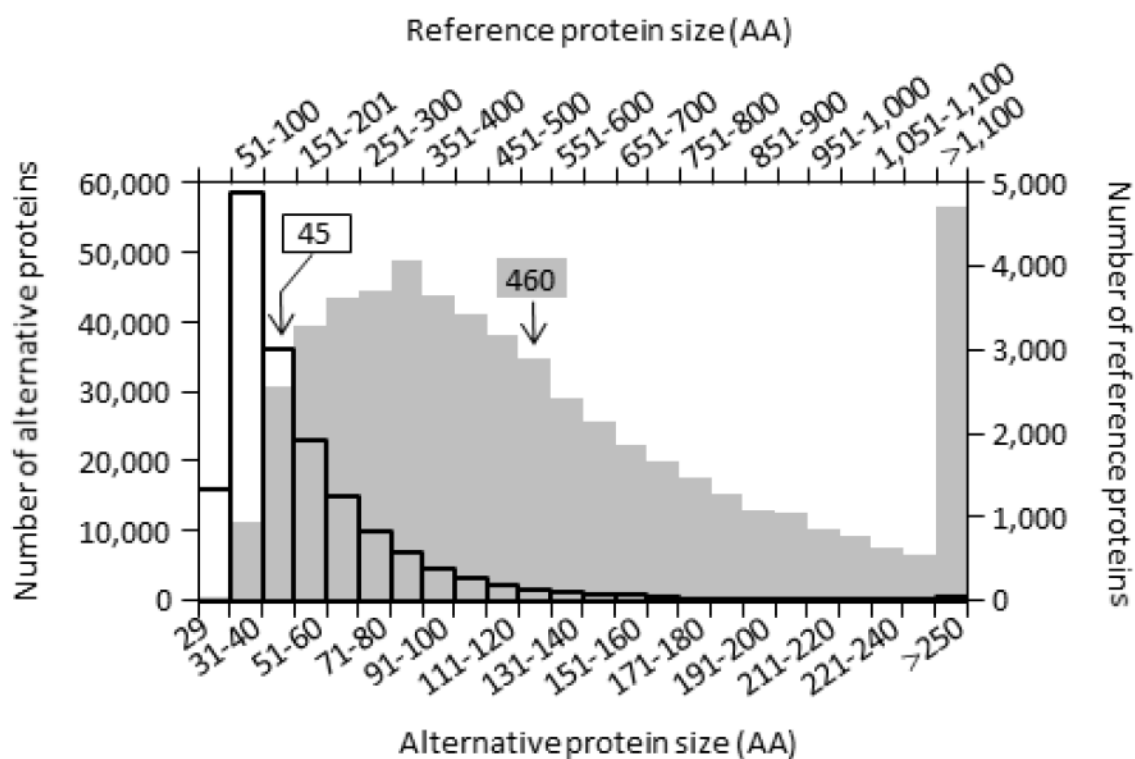
- Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO reports*, 1(3) : 287–292.
- Skarshewski, A., Stanton-Cook, M., Huber, T., Al Mansoori, S., Smith, R., Beatson, S. A., et Rothnagel, J. A. (2014) upepperoni : an online tool for upstream open reading frame location and analysis of transcript conservation. *BMC bioinformatics*, 15(1) : 1.
- Slager, R. E., Newton, T. L., Vlangos, C. N., Finucane, B., et Elsea, S. H. (2003) Mutations in *rai1* associated with smith–magenis syndrome. *Nature genetics*, 33(4) : 466–468.
- Slavoff, S. A., Heo, J., Budnik, B. A., Hanakahi, L. A., et Saghatelian, A. (2014) A human short open reading frame (sorf)-encoded polypeptide that stimulates DNA end joining. *Journal of Biological Chemistry*, 289(16) : 10950–10957.
- Smirnova, E., Griparic, L., Shurland, D.-L., et Van Der Blik, A. M. (2001) Dynamin-related protein drp1 is required for mitochondrial division in mammalian cells. *Molecular biology of the cell*, 12(8) : 2245–2256.
- Smith, A., McGavran, L., Robinson, J., Waldstein, G., Macfarlane, J., Zonona, J., Reiss, J., Lahr, M., Allen, L., Magenis, E., et al. (1986) Interstitial deletion of (17)(p11.2p11.2) in nine patients. *American journal of medical genetics*, 24(3) : 393–414.
- Smith, E., Meyerrose, T. E., Kohler, T., Namdar-Attar, M., Bab, N., Lahat, O., Noh, T., Li, J., Karaman, M. W., Hacia, J. G., et al. (2005) Leaky ribosomal scanning in mammalian genomes : significance of histone H4 alternative translation in vivo. *Nucleic acids research*, 33(4) : 1298–1308.
- Somers, J., Pöyry, T., et Willis, A. E. (2013) A perspective on mammalian upstream open reading frame function. *The international journal of biochemistry & cell biology*, 45(8) : 1690–1700.
- Sonnhammer, E. L. et Östlund, G. (2015) Inparanoid 8 : orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic acids research*, 43(D1) : D234–D239.
- Stadler, M. et Fire, A. (2011) Wobble base-pairing slows in vivo translation elongation in metazoans. *Rna*, 17(12) : 2063–2073.
- Starck, S. R., Tsai, J. C., Chen, K., Shodiya, M., Wang, L., Yahiro, K., Martins-Green, M., Shastri, N., et Walter, P. (2016) Translation from the 5' untranslated region shapes the integrated stress response. *Science*, 351(6272) : aad3867.
- Steen, H. et Mann, M. (2004) The abc's (and xyz's) of peptide sequencing. *Nature reviews Molecular cell biology*, 5(9) : 699–711.

- Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V. T. K., Hein, M. Y., Huang, S.-X., Ma, M., Shen, B., Qian, S.-B., Hengel, H., *et al.* (2012) Decoding human cytomegalovirus. *Science*, 338(6110) : 1088–1093.
- Taguchi, N., Ishihara, N., Jofuku, A., Oka, T., et Mihara, K. (2007) Mitotic phosphorylation of dynamin-related gtpase drp1 participates in mitochondrial fission. *Journal of Biological Chemistry*, 282(15) : 11521–11529.
- Tatusov, R. L., Koonin, E. V., et Lipman, D. J. (1997) A genomic perspective on protein families. *Science*, 278(5338) : 631–637.
- Tautz, D. (2009) Polycistronic peptide coding genes in eukaryotes—how widespread are they? *Briefings in functional genomics & proteomics*, 8(1) : 68–74.
- Twig, G., Elorza, A., Molina, A. J., Mohamed, H., Wikstrom, J. D., Walzer, G., Stiles, L., Haigh, S. E., Katz, S., Las, G., *et al.* (2008) Fission and selective fusion govern mitochondrial segregation and elimination by autophagy. *The EMBO journal*, 27(2) : 433–446.
- Vanderperre, B. (2013) *Les cadres ouverts de lecture alternatifs contribuent significativement au protéome des eucaryotes*. PhD thesis, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Québec, Canada.
- Vanderperre, B., Lucier, J.-F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., Wisztorski, M., Salzet, M., Boisvert, F.-M., et Roucou, X. (2013) Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PloS one*, 8(8) : e70698.
- Vanderperre, B., Lucier, J.-F., et Roucou, X. (2012) Haltorf : a database of predicted out-of-frame alternative open reading frames in human. *Database*, 2012 : bas025.
- Vanderperre, B., Staskevicius, A. B., Tremblay, G., McCoy, M., O'Neill, M. A., Cashman, N. R., et Roucou, X. (2011) An overlapping reading frame in the prnp gene encodes a novel polypeptide distinct from the prion protein. *The FASEB Journal*, 25(7) : 2373–2386.
- Vaudel, M., Burkhardt, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., Martens, L., et Barsnes, H. (2015) Peptideshaker enables reanalysis of ms-derived proteomics data sets. *Nature biotechnology*, 33(1) : 22–24.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith,

- H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001) The sequence of the human genome. *science*, 291(5507) : 1304–1351.
- Wei, J., Wu, C., et Sachs, M. S. (2012) The arginine attenuator peptide interferes with the ribosome peptidyl transferase center. *Molecular and cellular biology*, 32(13) : 2396–2406.
- Wells, J. N., Bergendahl, L. T., et Marsh, J. A. (2016) Operon gene order is optimized for ordered protein complex assembly. *Cell reports*, 14(4) : 679–685.
- Westermann, B. (2010) Mitochondrial fusion and fission in cell life and death. *Nature reviews Molecular cell biology*, 11(12) : 872–884.
- Wethmar, K., Barbosa-Silva, A., Andrade-Navarro, M. A., et Leutz, A. (2014) uorfdb—a comprehensive literature database on eukaryotic uorf biology. *Nucleic acids research*, 42(D1) : D60–D67.
- Willingham, A. T. et Gingeras, T. R. (2006) Tuf love for “junk” dna. *Cell*, 125(7) : 1215–1220.
- Xie, S.-Q., Nie, P., Wang, Y., Wang, H., Li, H., Yang, Z., Liu, Y., Ren, J., et Xie, Z. (2015) Rpfdb : a database for genome wide information of translated mrna generated from ribosome profiling. *Nucleic acids research*, page gkv972.
- Yamano, K. et Youle, R. J. (2011) Coupling mitochondrial and cell division. *Nature cell biology*, 13(9) : 1026.
- Yosten, G. L., Liu, J., Ji, H., Sandberg, K., Speth, R., et Samson, W. K. (2015) A 5'-upstream short open reading frame encoded peptide regulates angiotensin type 1a receptor production and signalling via the β -arrestin pathway. *The Journal of physiology*.
- Young, S. K. et Wek, R. C. (2016) Upstream open reading frames differentially regulate gene-specific translation in the integrated stress response. *Journal of Biological Chemistry*, pages jbc–R116.
- Zamore, P. D. et Haley, B. (2005) Ribo-gnome : the big world of small rnas. *Science*, 309(5740) : 1519–1524.
- Zhao, J., Liu, T., Jin, S., Wang, X., Qu, M., Uhlén, P., Tomilin, N., Shupliakov, O., Lendahl, U., et Nistér, M. (2011) Human mief1 recruits drp1 to mitochondrial outer membranes and promotes mitochondrial fusion rather than fission. *The EMBO journal*, 30(14) : 2762–2778.

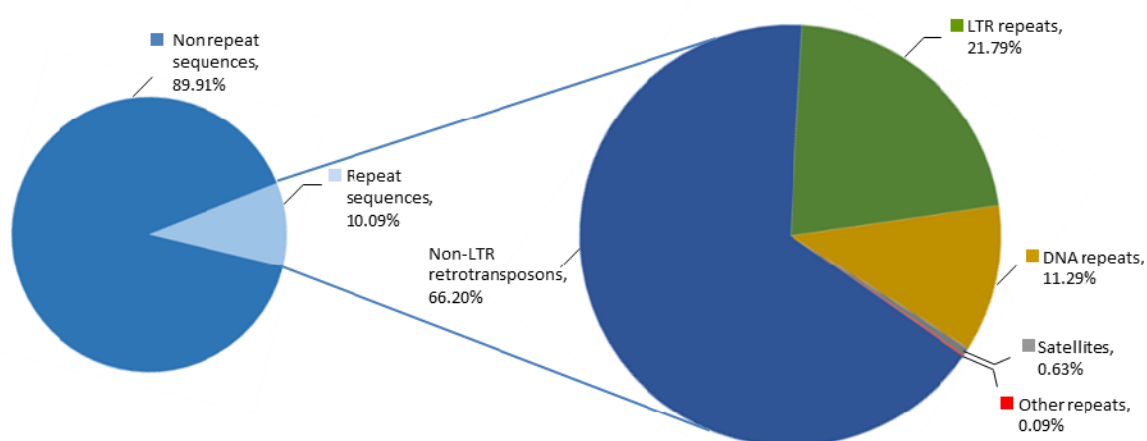
5 ANNEXES

5.1 Figures supplémentaires de l'article



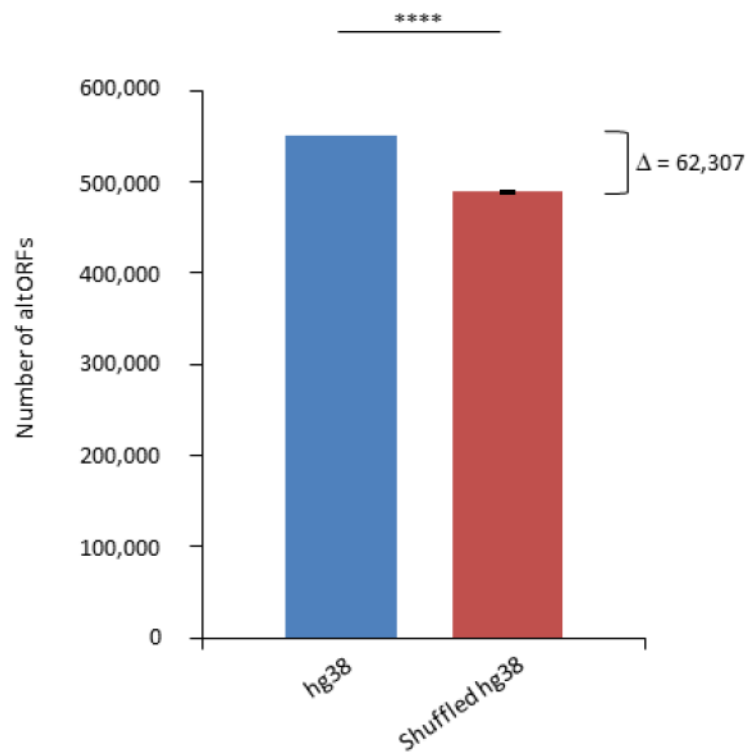
Supplementary Figure 1: Size distribution of alternative and reference human proteins.

Size distribution of alternative (empty bars, vertical and horizontal axes) and reference (grey bars, secondary horizontal and vertical axes) proteins. Arrows indicate the median size. The median alternative protein length is 45 amino acids (AA) compared to 460 for the reference proteins.



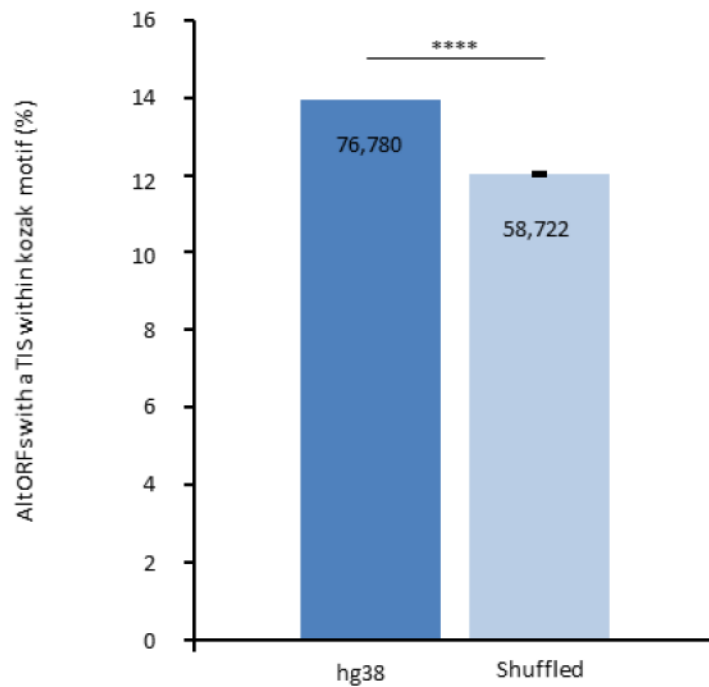
Supplementary Figure 2: 10% of altORFs are present in different classes of repeats.

More than half of the human genome is composed of repeated sequences, and only 10.09% of altORFs are located inside these repeats. These altORFs are detected in non-LTR retrotransposons, LTR repeats, DNA repeats, satellites and other repeats. Proportions were determined using RepeatMasker (version 3.3.0).



Supplementary Figure 3: There are more altORFs than would be expected from random occurrence in the human transcriptome.

Number of total altORFs in hg38 compared to shuffled hg38. Means and standard deviations for 100 replicates obtained by sequence shuffling are shown. Statistical significance was determined by using one sample t-test with two-tailed p -values. **** $P < 0,0001$.



Supplementary Figure 4: The proportion of altORFs with a translation initiation site with a Kozak motif in hg38 is significantly different from 100 shuffled hg38 transcriptomes.

Percentage of altORFs with an optimal Kozak sequence in hg38 (dark blue) compared to 100 shuffled hg38 (light blue). Mean and standard deviations for sequence shuffling are displayed, and significant difference was defined by using one sample t test. **** $P < 0,0001$. Note that shuffling all transcripts in the hg38 transcriptome generates a total of 489,073 altORFs in average, compared to 551,380 altORFs in hg38. Most transcripts result from alternative splicing and there are 183,191 unique altORFs in the hg38 transcriptome, while the 489,073 altORFs in shuffled transcriptomes are all unique. Fig. 1e shows the percentage of unique altORFs with a kozak motif (15%), while the current Fig. shows the percentage of altORFs with a kozak motif relative to the total number of altORFs (14%).

a **AltLINC01420^{nc}**
 MGDQPCASGRSTLPPGNAREAKPKKRCLLAPRWDYPEGTPNGGSTLLPSAPPPASAGLKSHPPPEK

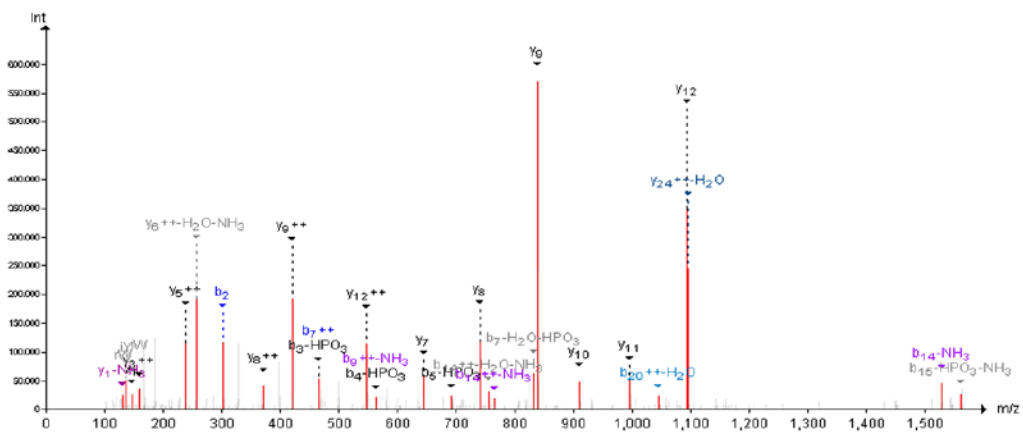
b

Spectrum & Fragment Ions (PR - NH2-WDY^qPEGTPNGGSTLLPSAPPPASAGLK-COOH - SH 3+ 917.09 m/z)

□ - + ?

NH2-W D **E** **F** **E** **G** T **P** N **G** G S T T **L** **P** **S** **A** **P** **P** **A** **S** **A** **G** **L** **K**-COOH

m/z = 917.09
 [M+3H]³⁺ = 2751.27 Da



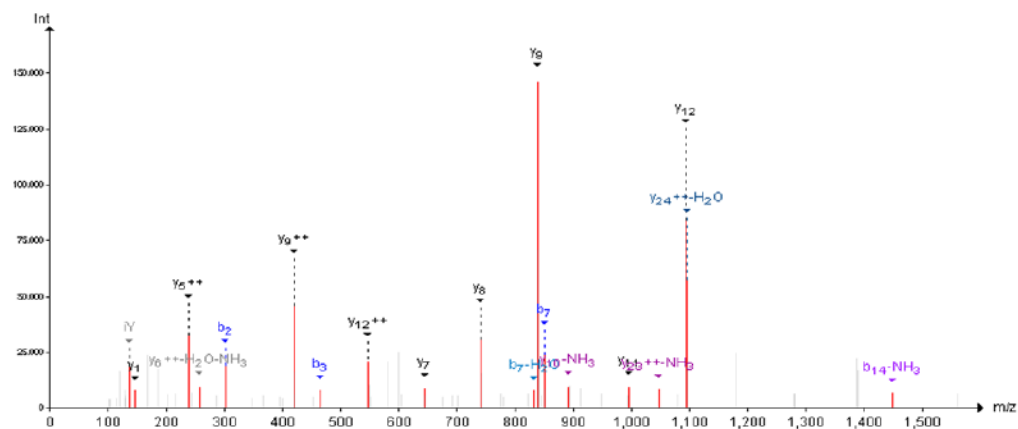
c

Spectrum & Fragment Ions (PR - NH2-WDY^qPEGTPNGGSTLLPSAPPPASAGLK-COOH - SH 3+ 890.43 m/z)

□ - + ?

NH2-W D **Y** **P** **E** **G** T **P** N **G** G S T T **L** **P** **S** **A** **P** **P** **A** **S** **A** **G** **L** **K**-COOH

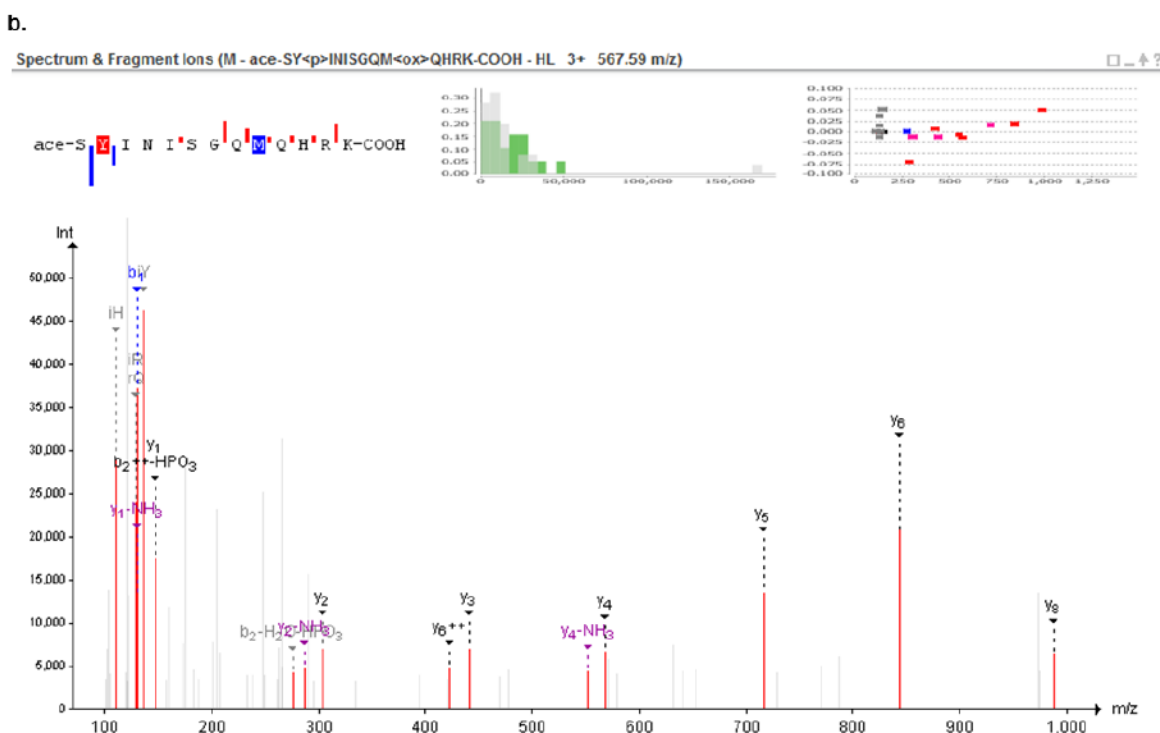
m/z = 890.4309
 [M+3H]³⁺ = 2671.29 Da



Supplementary Figure 5: Example of a phosphorylated peptide in mitosis, in the alternative protein AltLINC01420^{nc}.

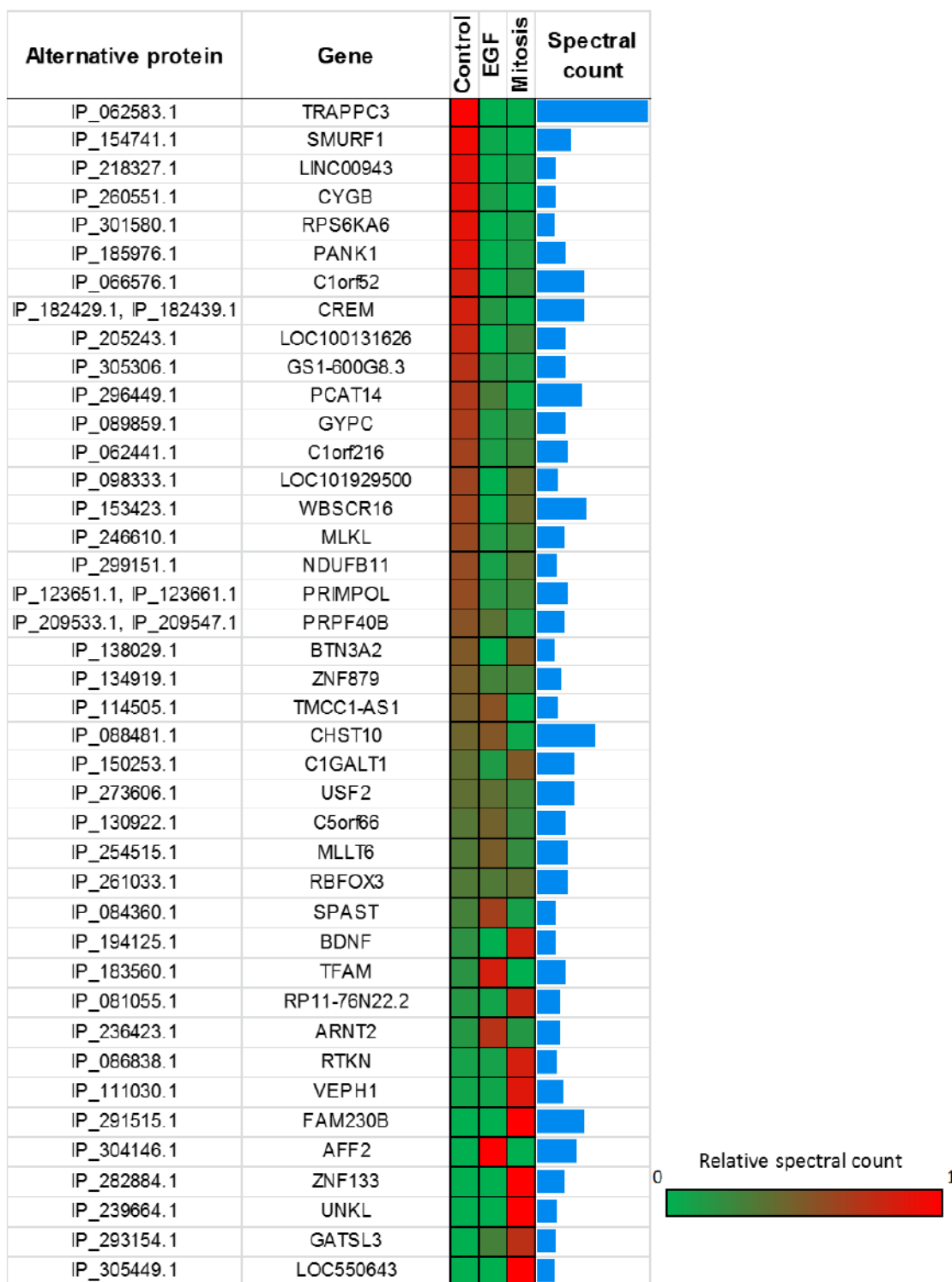
(a) AltLINC01420^{nc} amino acid sequence with detected peptides underlined and phosphorylated peptide in bold (73,9% sequence coverage). (b) MS/MS spectrum for the phosphorylated peptide (PeptideShaker graphic interface output). The phosphorylation site is the tyrosine residue, position 2. (c) MS/MS spectrum for the non-phosphorylated peptide. The mass difference between the precursor ions between both spectra corresponds to that of a phosphorylation, confirming the specific phosphorylation of this residue in mitosis.

a. **AltTFAM^{3'}**
 MSYINISGQMQRHKHLCYPAGKFPLSSFNINYPYFILNIHIIPSQIYWEVQC



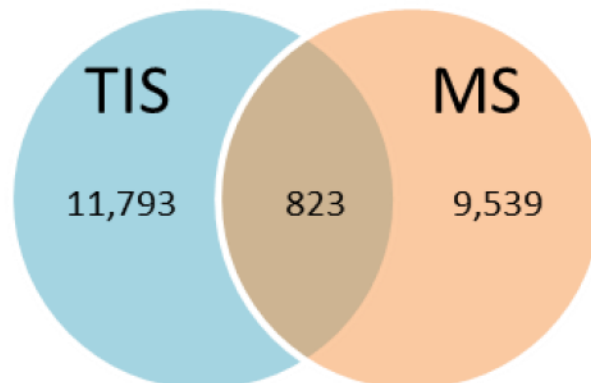
Supplementary Figure 6: Example of a phosphorylated peptide in EGF-treated cells, in the alternative protein AltTFAM^{3'}.

(a) AltTFAM^{3'} amino acid sequence with the detected phosphorylated peptide underlined (22,2% sequence coverage). (b) MS/MS spectrum for the phosphorylated peptide (PeptideShaker graphic interface output). The phosphorylation site is a tyrosine residue, position 2. The difference in spectral counting indicates an increase in phosphorylation in cells stimulated with EGF.



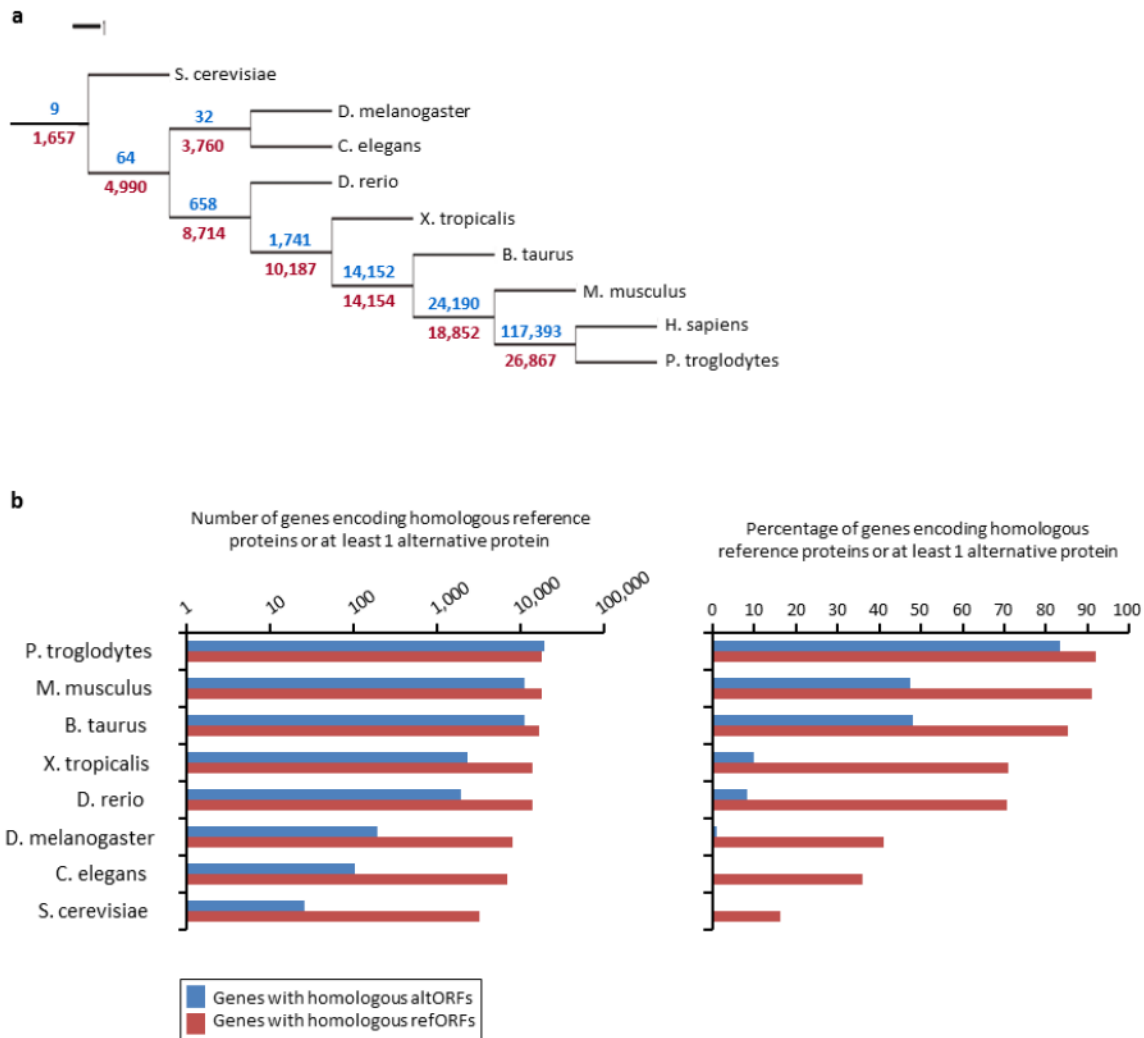
Supplementary Figure 7: The alternative phosphoproteome in mitosis and EGF-treated cells.

Heatmap showing relative levels of spectral counts for phosphorylated peptides following the indicated treatment²⁹. For each condition, heatmap colors show the percentage of spectral count on total MS/MS phosphopeptide spectra. Blue bars on the right represent the number of MS/MS spectra; only proteins with spectral counts covering a range between 70 and 10 are shown.



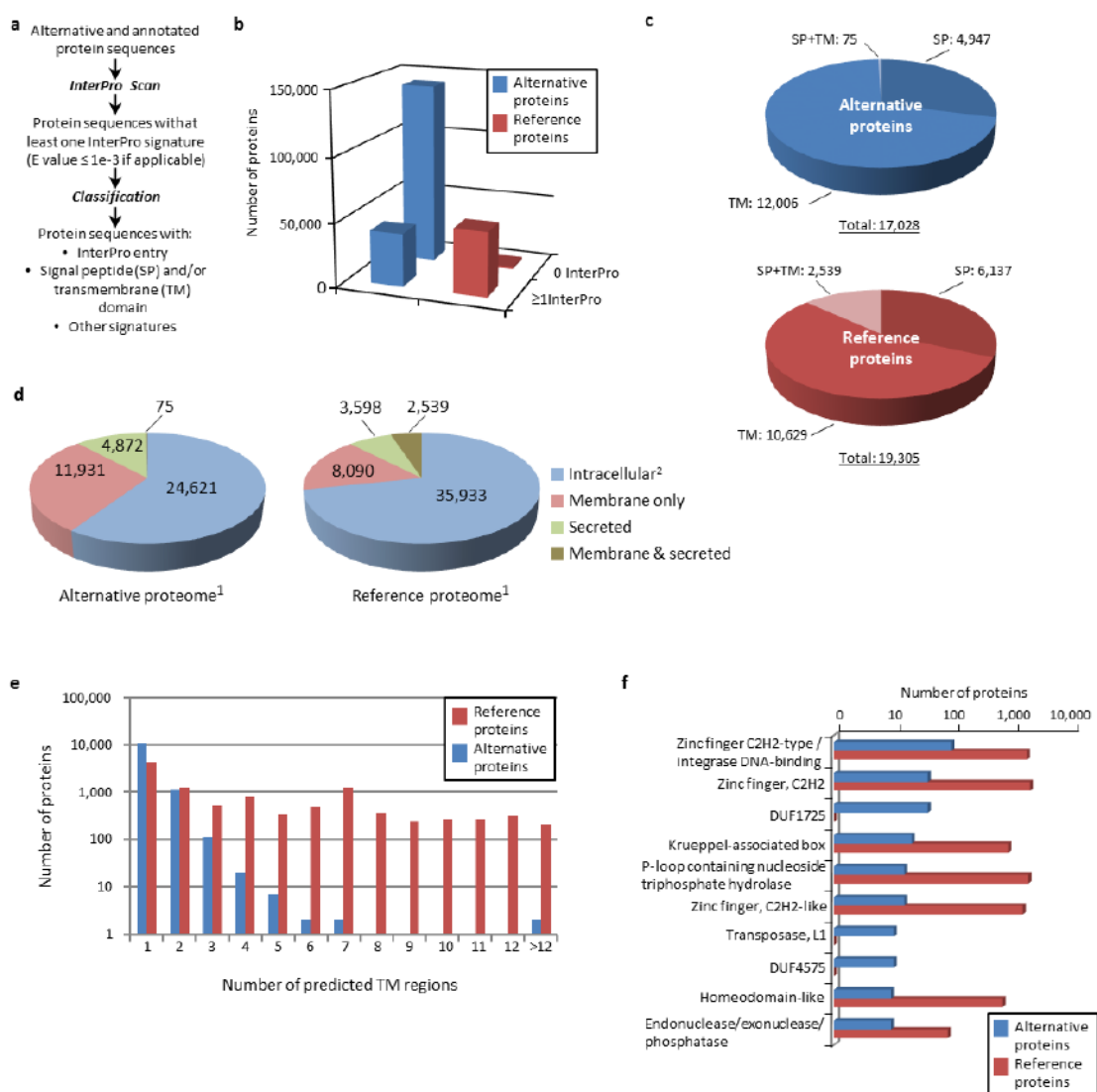
Supplementary Figure 8: Number of alternative proteins detected by ribosome profiling and mass spectrometry.

The expression of 823 alternative proteins was detected by both ribosome profiling (translation initiation sites, TIS) and mass spectrometry (MS).



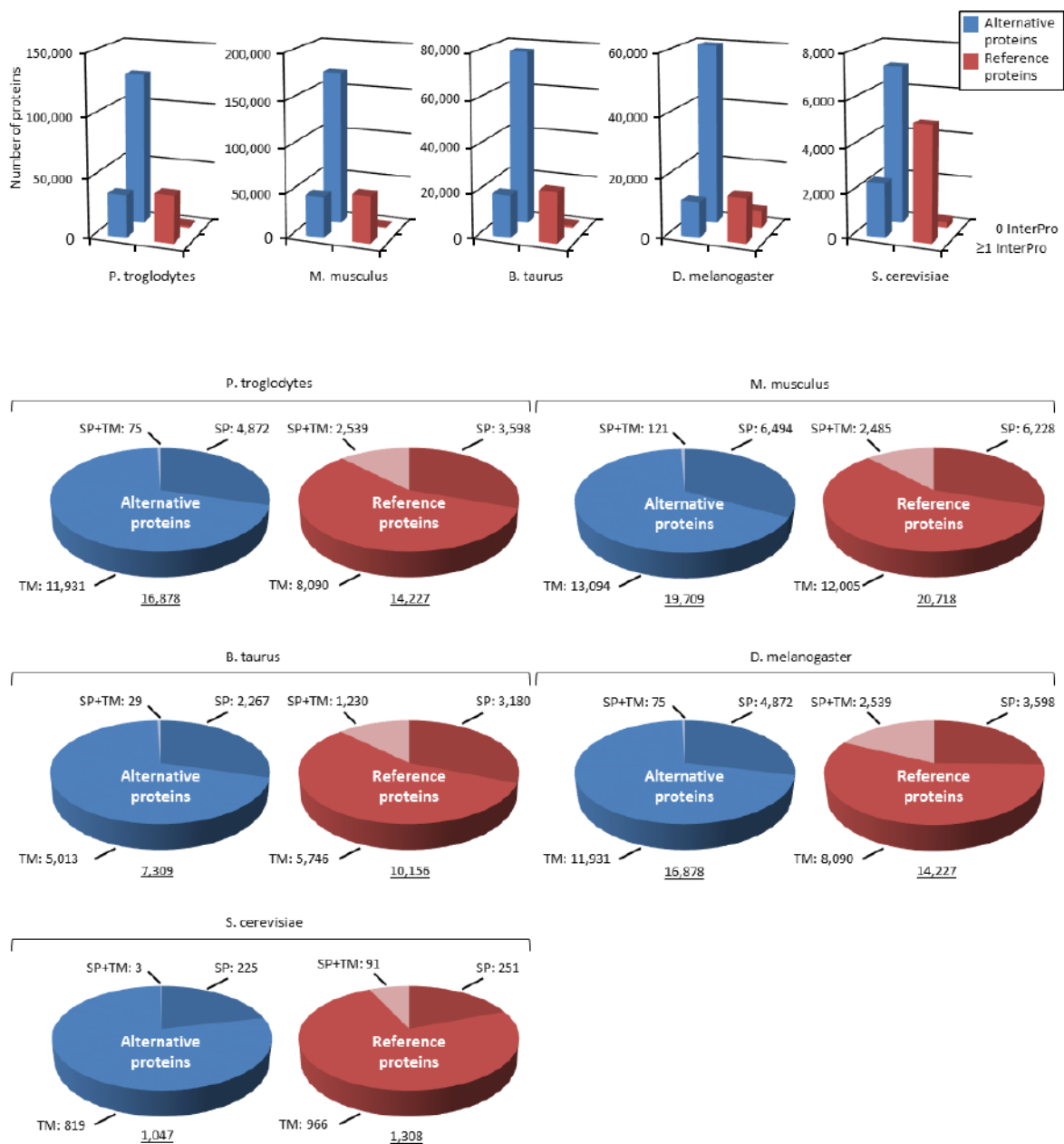
Supplementary Figure 9: Conservation of alternative and reference proteins across different species.

(a) Phylogenetic tree: conservation of alternative (blue) and reference (red) proteins across various eukaryotic species. (b) Number and fraction of genes encoding homologous reference proteins or at least 1 homologous alternative protein.

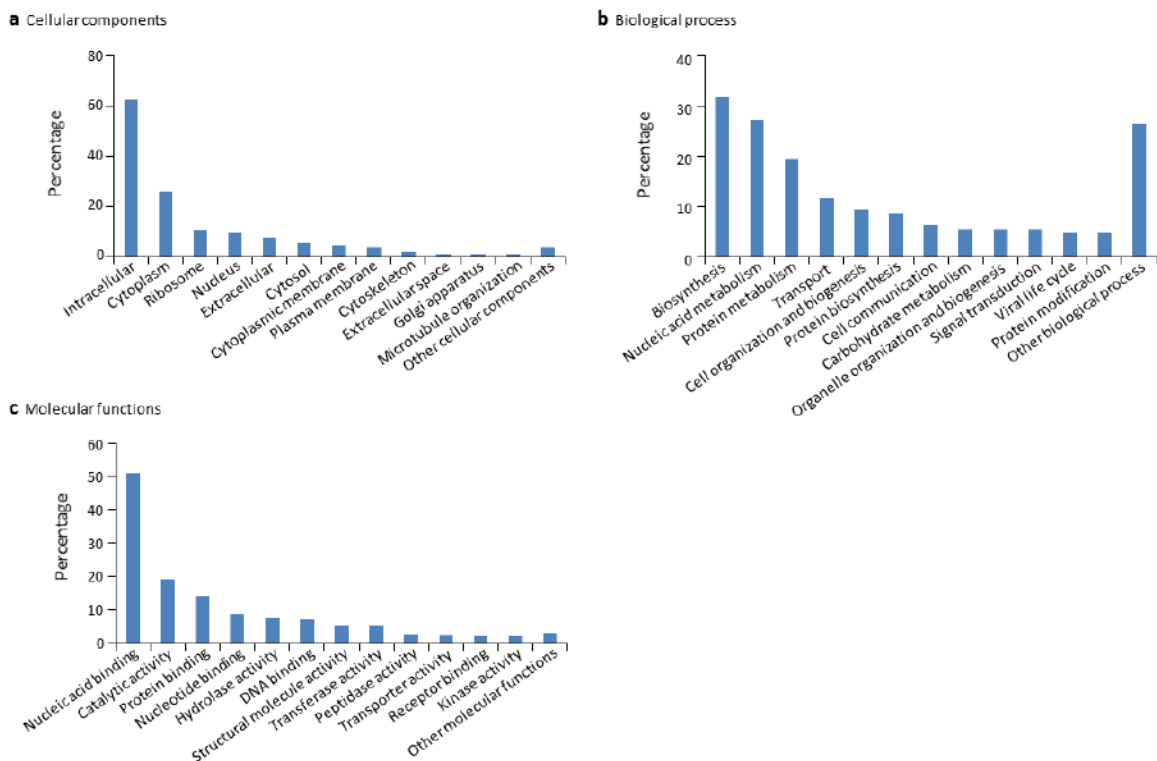


Supplementary Figure 10: Human alternative proteome sequence analysis and classification using InterProScan.

(a) InterPro annotation pipeline. (b) Alternative and reference proteins with InterPro signatures. (c) Number of alternative and reference proteins with transmembrane domains (TM), signal peptides (S) and both TM and SP. (d) Number of all alternative and reference proteins predicted to be intracellular, membrane, secreted and membrane-spanning and secreted. ¹Proteins with at least one InterPro signature; ²Proteins with no predicted signal peptide or transmembrane features. (e) Number of predicted TM regions for alternative and reference proteins. (f) The top 10 InterPro families in the alternative proteome.

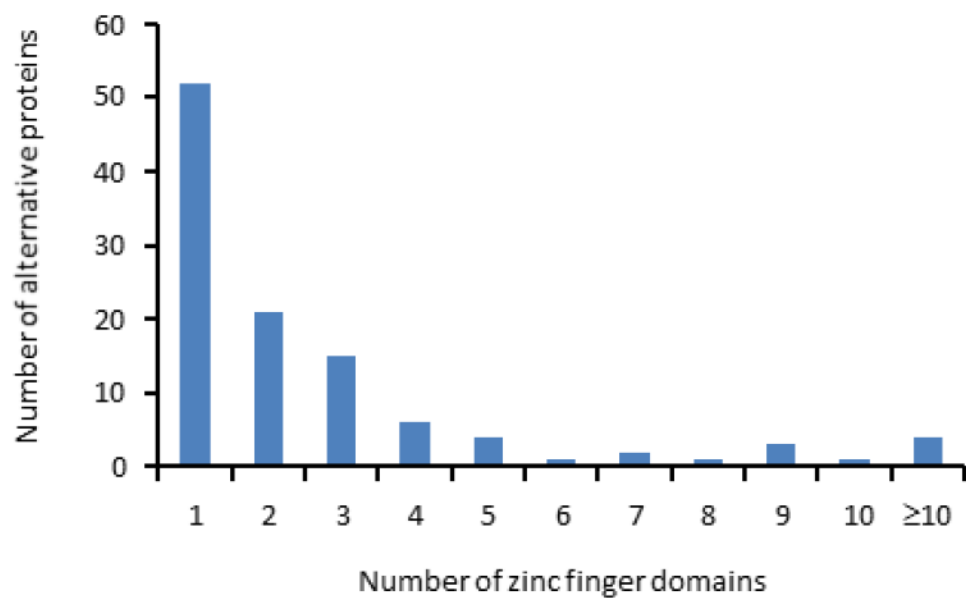


Supplementary Figure 11: Alternative proteome sequence analysis and classification in *P. troglodytes*, *M. musculus*, *B. Taurus*, *D. melanogaster* and *S. cerevisiae*. For each organism, the number of InterPro signatures (top graphs) and proteins with transmembrane (TM), signal peptide (SP), or TM+SP features (bottom pie charts) is indicated for alternative and reference proteins.



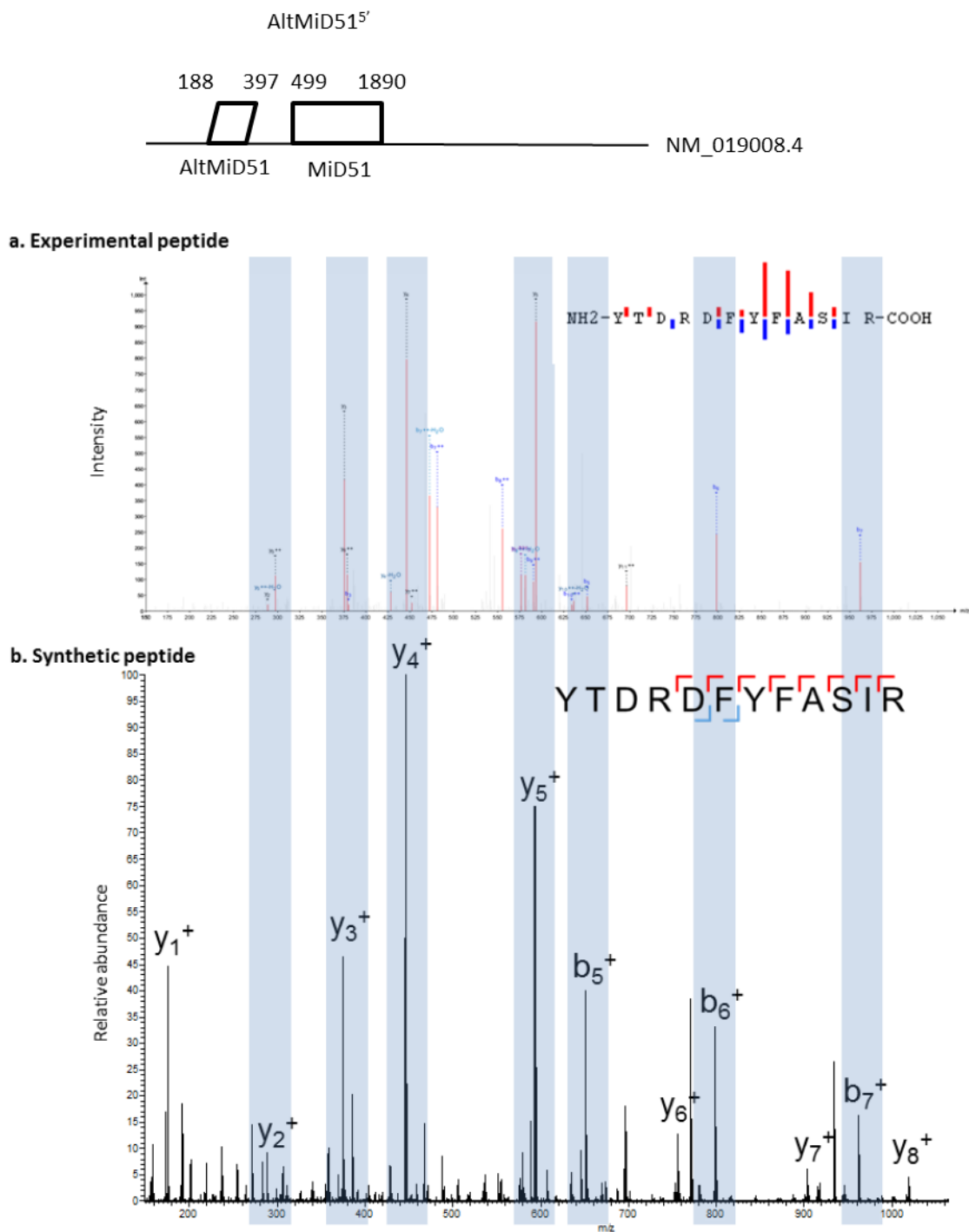
Supplementary Figure 12: Gene ontology (GO) annotations for human alternative proteins.

GO terms assigned to InterPro entries are grouped into 13 categories within the three ontologies. (a) 94 GO terms were categorized into molecular function for 302 alternative proteins. (b) 64 GO terms were categorized into biological process for 128 alternative proteins. (c) 34 GO terms were categorized into cellular component for 107 alternative proteins. The majority of alternative proteins with GO terms are predicted to be intracellular, to function in nucleic acid-binding, catalytic activity and protein binding and to be involved in biosynthesis and nucleic acid metabolism processes.



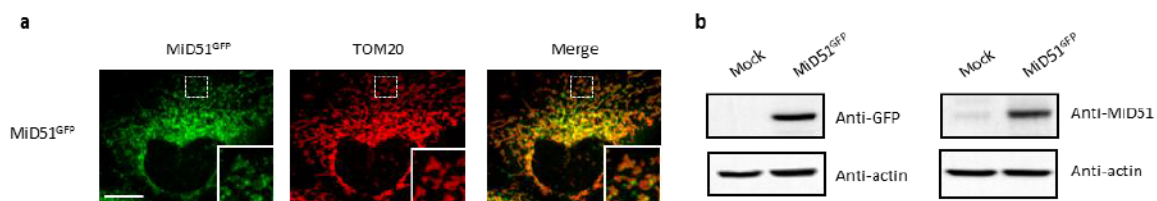
Supplementary Figure 13: Human alternative proteins with zinc finger domains.

A total of 110 alternative proteins have between 1 and 23 zinc finger domains.



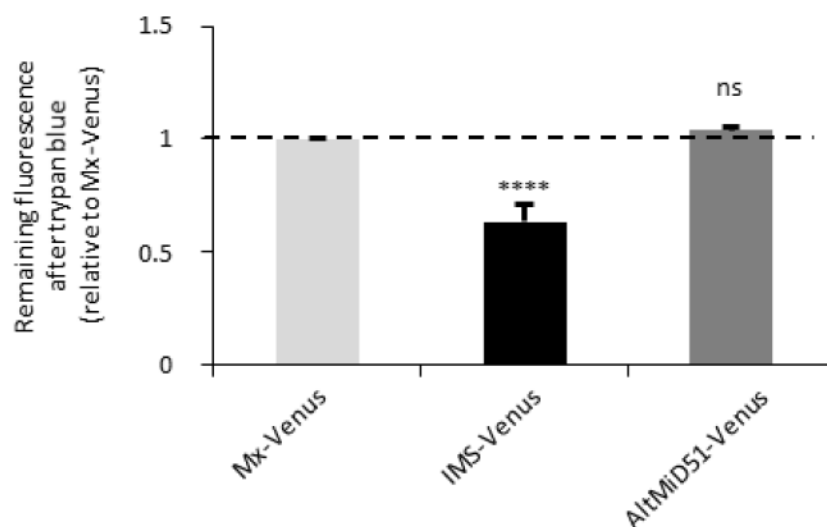
Supplementary Figure 14: Spectra validation for altMiD51

Example of validation for altMiD51 specific peptides YTD R D F Y F A S I R and GLVFLNGK. (a,c) Experimental MS/MS spectra (PeptideShaker graphic interface output). (b,d) MS/MS spectra of the synthetic peptides. Matching peaks are shown with blue masks. A diagram of the transcript with its accession number and the localization of the altORF and the CDS is shown at the top.



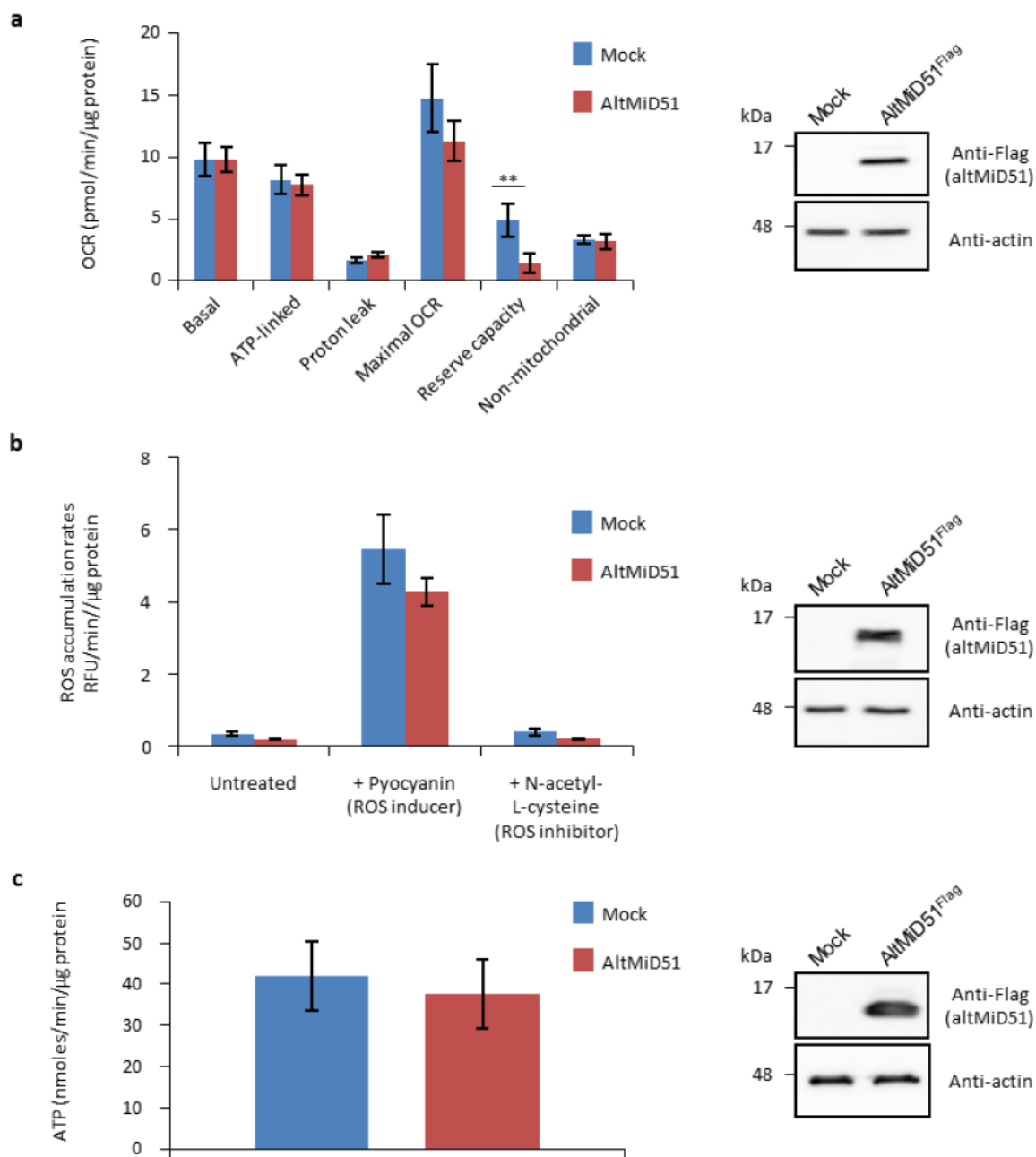
Supplementary Figure 15: MiD51 expression results in mitochondrial fission.

(a) Confocal microscopy of HeLa cells transfected with MiD51^{GFP} immunostained with anti-TOM20 (red channel) monoclonal antibodies. In each image, boxed areas are shown at higher magnification in the bottom right corner. The localization of MiD51 in fission sites is shown in merged higher magnification inset. Scale bar, 10 mm. (b) Human HeLa cells transfected with empty vector (mock) or MiD51^{GFP} were lysed and analyzed by western blot to confirm MiD51^{GFP} expression.



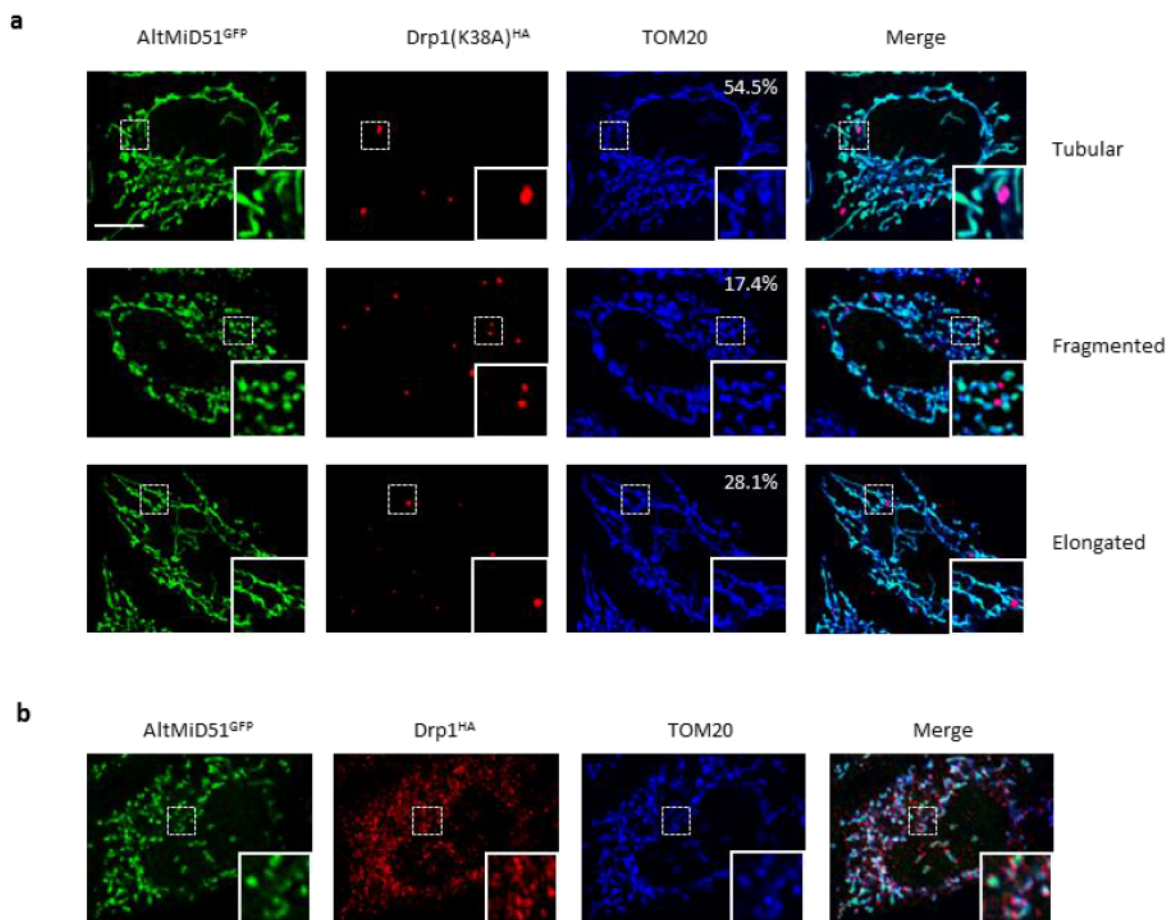
Supplementary Figure 16: AltMiD51 is localized in the mitochondrial matrix.

Trypan blue quenching experiment performed on HeLa cells stably expressing the indicated constructs. The fluorescence remaining after quenching by trypan blue is shown relative to Matrix-Venus (Mx-Venus) indicated by the dashed line. (**** $p < 0,0001$, one-way ANOVA). The absence of quenching of the fluorescence compared to IMS-Venus indicates the matricial localization of altMiD51. $n \geq 3$ cells were quantified per experiment, and results are from 6 independent experiments. Data are mean \pm SEM.



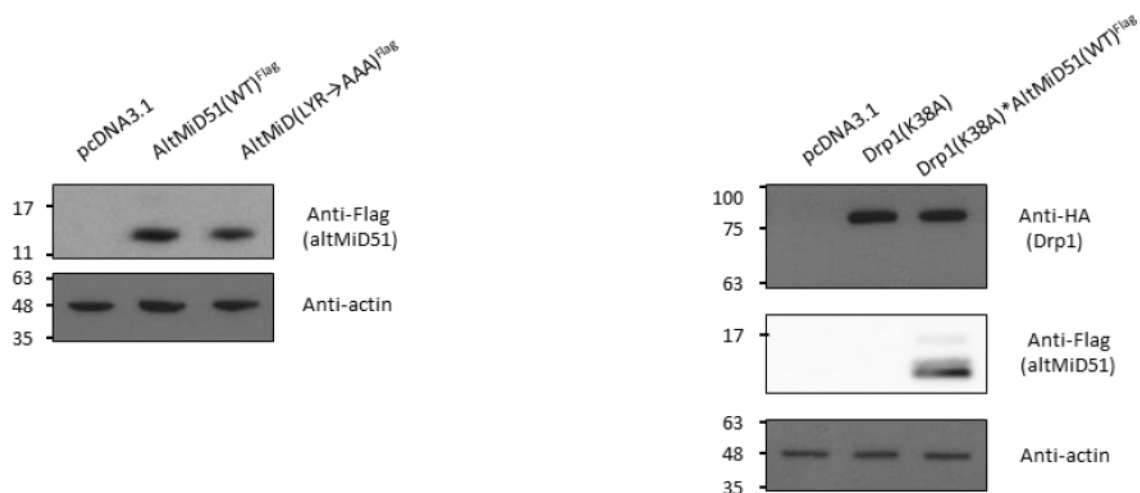
Supplementary Figure 17: Mitochondrial function parameters.

(a) Oxygen consumption rates (OCR) in HeLa cells transfected with empty vector (mock) or altMiD51^{Flag}. Mitochondrial function parameters were assessed in basal conditions (basal), in the presence of oligomycin to inhibit the ATP synthase (oxygen consumption that is ATP-linked), FCCP to uncouple the mitochondrial inner membrane and allow for maximum electron flux through the respiratory chain (maximal OCR), and antimycin A/rotenone to inhibit complex III (non-mitochondrial). The balance of the basal OCR comprises oxygen consumption due to proton leak and nonmitochondrial sources. The mitochondrial reserve capacity (maximal OCR- basal OCR) is an indicator of rapid adaptation to stress and metabolic changes. Mean values of replicates are plotted with error bars corresponding to the 95% confidence intervals. Statistical significance was estimated using a two-way ANOVA with Tukey's post-hoc test (** $p = 0,004$). (b) ROS production in mock and altMiD51-expressing cells. Cells were untreated, treated with a ROS inducer or a ROS inhibitor. Results represent the mean value out of three independent experiments, with error bars corresponding to the standard error of the mean (s.e.m.). Statistical significance was estimated using unpaired T-test. (c) ATP synthesis rate in mock and altMiD51-expressing cells. No significant differences in ATP production were observed between mock and altMiD51 transfected cells. Results represent the mean of mitochondrial ATP production out of three independent experiments. Error bars represent the standard error of the mean. At the end of the experiments, cells were collected and proteins analyzed by western blot with antibodies against the Flag tag (altMiD51) or actin, as indicated, to verify the expression of altMiD51. A representative western blot is shown on the right. Molecular weight markers are shown on the left (kDa).



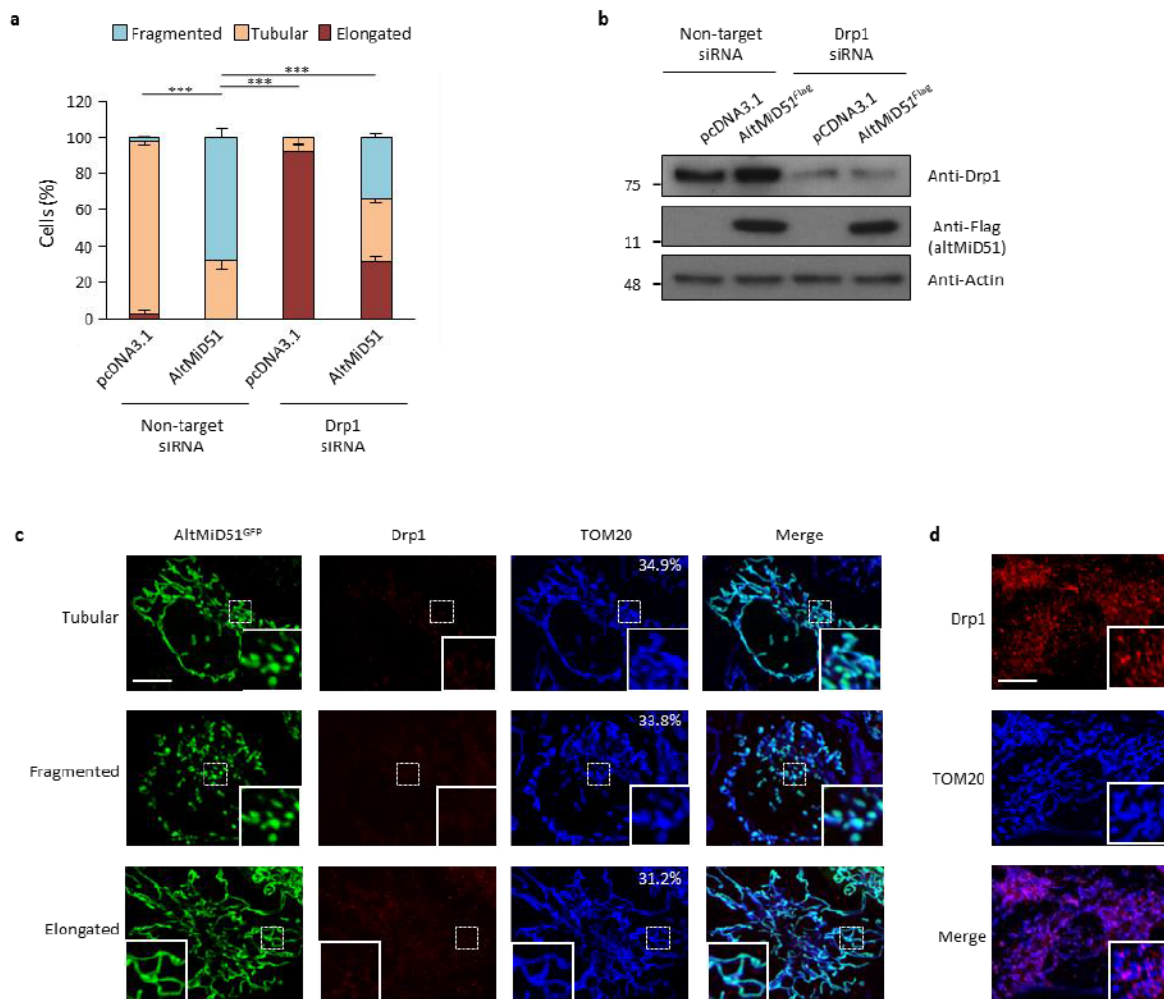
Supplementary Figure 18: Representative confocal images of cells co-expressing altMiD51^{GFP} and Drp1(K38A)^{HA}.

(a) Confocal microscopy of HeLa cells co-transfected with altMiD51^{GFP} and Drp1(K38A)^{HA} immunostained with anti-TOM20 (blue channel) and anti-HA (red channel) monoclonal antibodies. In each image, boxed areas are shown at higher magnification in the bottom right corner. % of cells with the indicated morphology is indicated on the TOM20 panels. (b) Confocal microscopy of HeLa cells co-transfected with altMiD51^{GFP} and Drp1(wt)^{HA} immunostained with anti-TOM20 (blue channel) and anti-HA (red channel) monoclonal antibodies. In each image, boxed areas are shown at higher magnification in the bottom right corner. Scale bar, 10 mm.



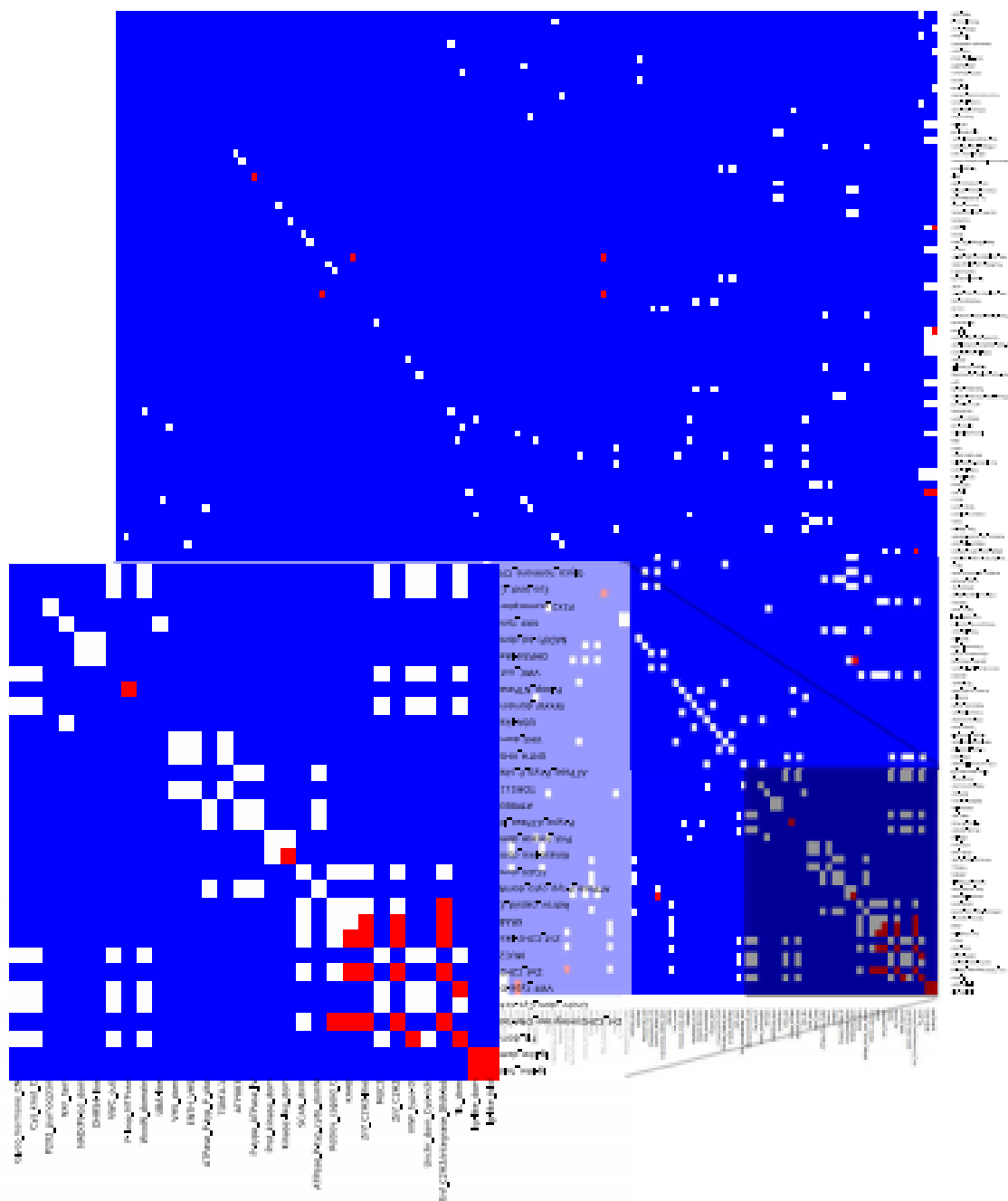
Supplementary Figure 19: Protein immunoblot showing the expression of different constructs in HeLa cells.

HeLa cells were transfected with empty vector (pcDNA3.1), altMiD51(WT)^{Flag}, altMiD51(LYR→AAA)^{Flag}, Drp1(K38A)^{HA}, or Drp1(K38A)^{HA} and altMiD51(WT)^{Flag}, as indicated. Proteins were extracted and analyzed by western blot with antibodies against the Flag tag (altMiD51), the HA tag (Drp1K38A) or actin, as indicated. Molecular weight markers are shown on the left (kDa).



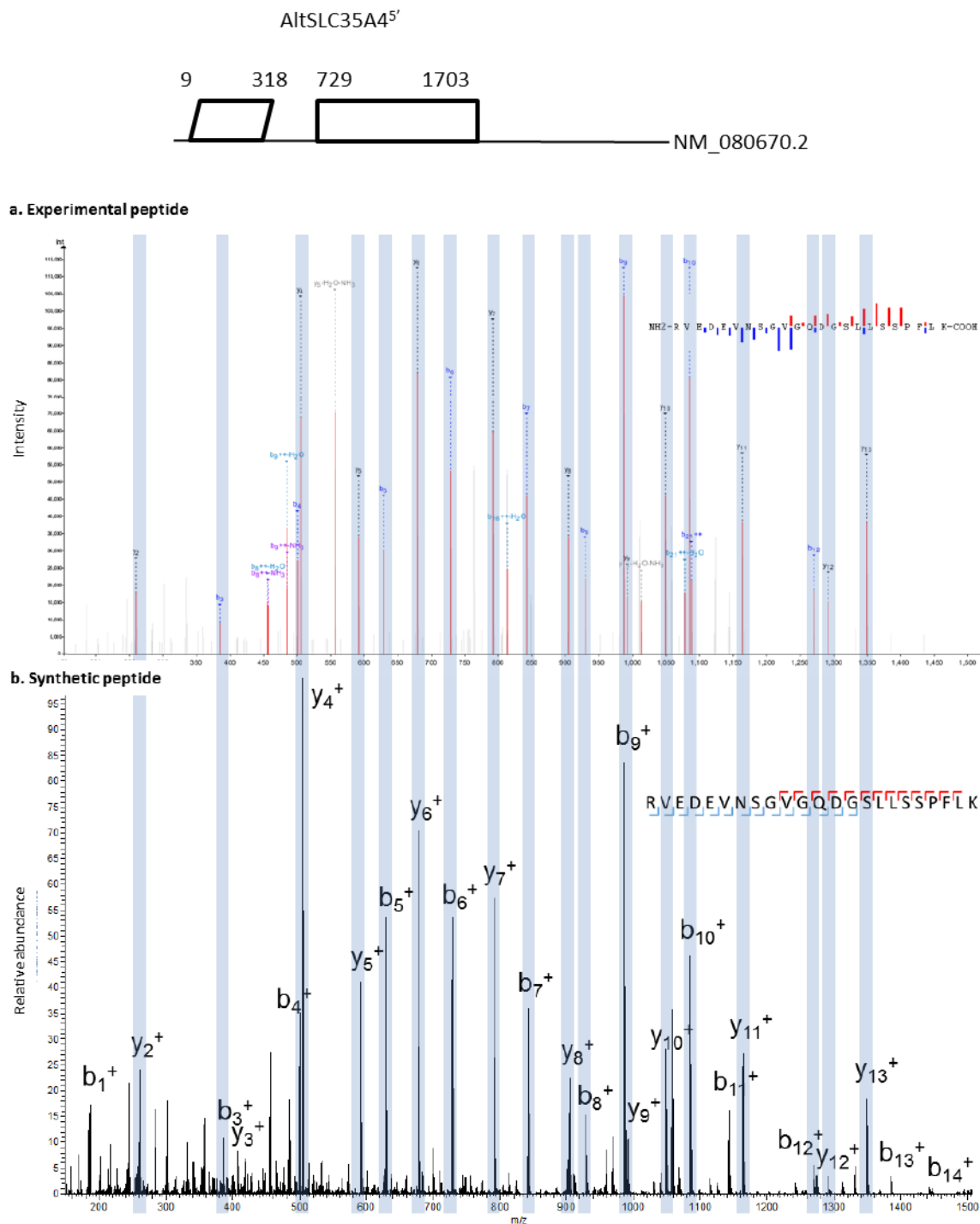
Supplementary Figure 20: AltMiD51-induced mitochondrial fragmentation is dependent on Drp1.

(a) Bar graphs show mitochondrial morphologies in HeLa cells treated with non-target or Drp1 siRNAs. Cells were mock-transfected (pcDNA3.1) or transfected with altMiD51^{Flag}. Means of three independent experiments per condition are shown. *** $p < 0.0005$ (Fisher's exact test) for the three morphologies between altMiD51 and the other experimental conditions. (b) HeLa cells treated with non-target or Drp1 siRNA were transfected with empty vector (pcDNA3.1) or altMiD51^{Flag}, as indicated. Proteins were extracted and analyzed by western blot with antibodies against the Flag tag (altMiD51), Drp1 or actin, as indicated. Molecular weight markers are shown on the left (kDa). (c) Confocal microscopy of Drp1 knockdown cells transfected with altMiD51^{GFP} immunostained with anti-TOM20 (blue channel) and anti-Drp1 (red channel) monoclonal antibodies. In each image, boxed areas are shown at higher magnification in the bottom right corner. % of cells with the indicated morphology is indicated on the TOM20 panels. Scale bar, 10 μ m. (d) Control Drp1 immunostaining in HeLa cells treated with a non-target siRNA. For (c) and (d), laser parameters for Drp1 and TOM20 immunostaining were identical.



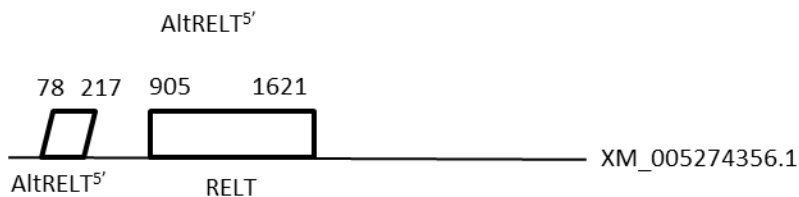
Supplementary Figure 21: Matrix of co-occurrence of InterPro entries between alternative/reference protein pairs coded by the same transcript.

Pixels show the number of times entries co-occur in reference and alternative proteins. Blue pixels indicate that these domains are not shared, white pixels indicate that they are shared once, and red that they are shared twice or more.

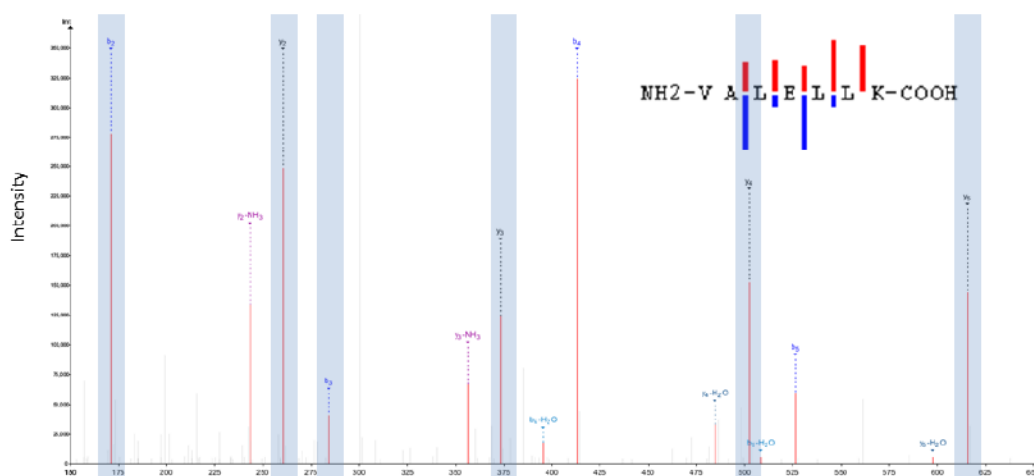


. Supplementary Figure 22: Spectra validation for altSLC35A4^{5'}

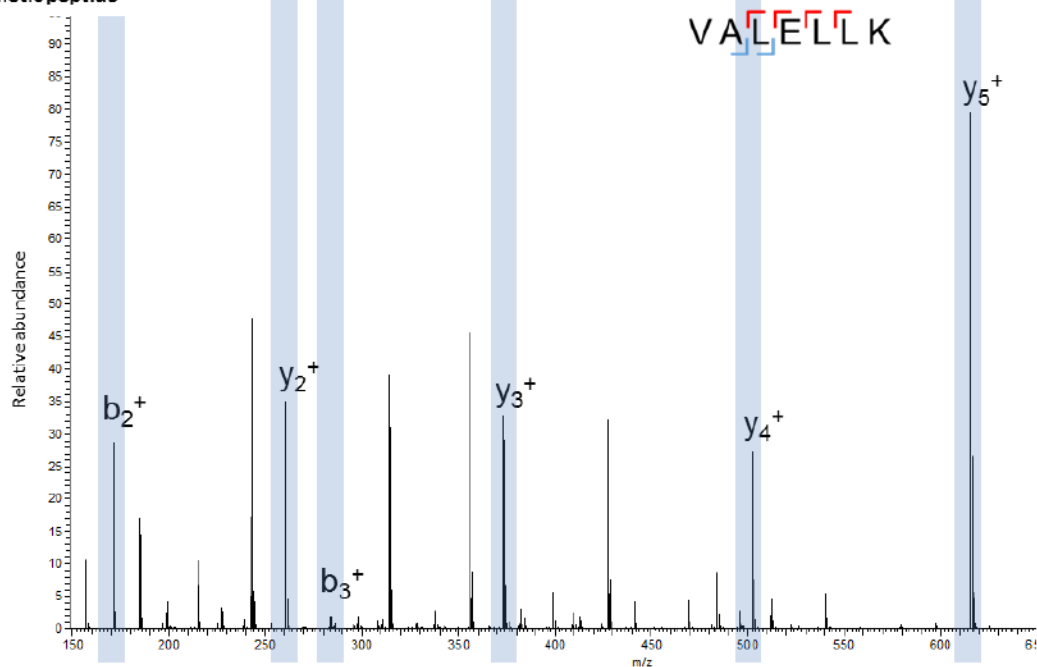
Example of validation for altSLC35A4^{5'} specific peptide RVEDEVNSGVGQDGSLLSSPFLK. (a) Experimental MS/MS spectra (PeptideShaker graphic interface output). (b) MS/MS spectra of the synthetic peptide. Matching peaks are shown with blue masks. A diagram of the transcript with its accession number and the localization of the altORF and the CDS is shown at the top.



a. Experimental peptide



b. Synthetic peptide



Supplementary Figure 23: Spectra validation for altRELT^{5'}

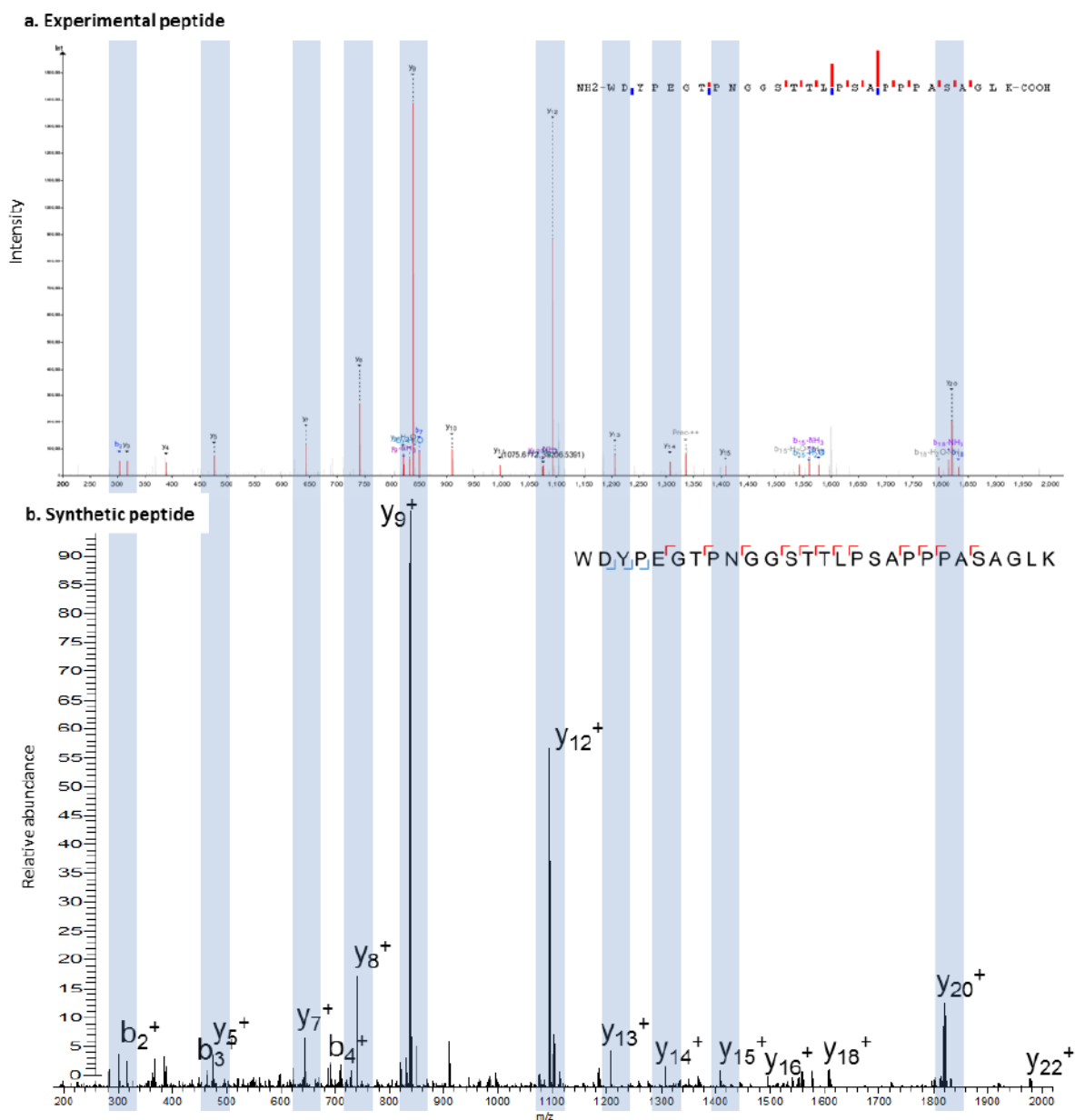
Example of validation for altRELT^{5'} specific peptide VALELLK. (a) Experimental MS/MS spectra (PeptideShaker graphic interface output). (b) MS/MS spectra of the synthetic peptide. Matching peaks are shown with blue masks. A diagram of the transcript with its accession number and the localization of the altORF and the CDS is shown at the top.

AltLINC01420^{nc}

70 273



NR_015367.2



Supplementary Figure 24: Spectra validation for altLINC01420^{nc}

Example of validation for altLINC01420^{nc} specific peptide WDYPEGTPNGGSTTLPSAPPASAGLK. (a) Experimental MS/MS spectra (PeptideShaker graphic interface output). (b) MS/MS spectra of the synthetic peptide. Matching peaks are shown with blue masks. A diagram of the transcript with its accession number and the localization of the altORF is shown at the top.

5.2 Tableaux supplémentaires de l'article

Supplementary Table 1: AltORFs and alternative protein annotations in different organisms

Genomes	Features						
	Transcripts		Current annotations		Annotations of alternative protein coding sequences		
	mRNAs	Others ¹	CDSs	Proteins	altORFs	Proteins	
<i>H. sapiens</i> RefSeq GCF_000001405.26	GRCh38	67,765	11,755	68,066	51,818	539,895	183,191
<i>P. troglodytes</i> RefSeq GCF_000001515.6	2.1.4	55,034	7,527	55,243	41,774	416,515	161,663
<i>M. musculus</i> RefSeq GCF_000001635.22	GRCm38p2,	73,450	18,886	73,551	53,573	642,203	215,472
<i>B. Taurus</i> RefSeq GCF_000003205.5	4.6.1 RefSeq	26,000	2151	25,971	23,619	137,844	96,015
<i>X. tropicalis</i> JGI_4.2.78	Ensembl	22,718	1,479	22,747	22,585	62,121	52,557
<i>D rerio</i> Ensembl ZV9.79		43,153 ²	6,429	43,221	39,758	129,328	71,705
<i>D. melanogaster</i> RefSeq GCA_000705575.1		30,255	3,474	30,715	20,995	174,771	71,705
<i>C. elegans</i> RefSeq GCF_000002985.6	WBcel235,	28,653	25,256	26,458	25,750	131,830	45,603
<i>S. cerevisiae</i> RefSeq GCA_000662435.1	YJM993_v1,	5,471	1,463	5,463	5,423	12,401	9,492

¹Other transcripts include miRNAs, rRNAs, ncRNAs, snRNAs, snoRNAs, tRNAs. ²Annotated retained-intron and processed transcripts were classified as mRNAs.

Supplementary Table 2: alternative zinc finger proteins detected by mass spectrometry (MS) and ribosome profiling (RP)

Alternative protein accession	Detection method ¹	Gene	Amino acid sequence
IP_238718.1	MS	RP11	MLVEVACSSCRSLHLKAGASEDGAALPAHTGGKENGATT
IP_278905.1	MS/RP	ZNF761	MSVARPLVGSILYAIIDFILERNLISVMSVARTLVRSHPLYATIDFILERNLTSVMS VARPLVRSQTLHAIVDFILEKNCNECGEVFNQQAHLAGHHRHTGKGP
IP_278681.1	MS	ZNF468	MNVARFLIKKQPLHITIDFILERNLTNGRNVTKVFSCKSNLKHKKIHIIEKPYRGK VCDKVFAYNAYLAKHTRHTGKLIISVMSVARPLVKIHTL
IP_106493.1	MS	ZNF717	MWKNLSSQVIPHHTPENSHGKPYGCNECGKTFQCQSYLIHQRTHTGKPYECN ECGKSFHQKANLQKHQGIHTGKPYECSKCGKTLSEVSPHCTS
IP_278745.1	MS/RP	ZNF816	MSVARPSVRNHPFNAIHYFTLERNLTVKNVTMFTFADHTLKDIGRFILERDHTNV RFVTRFSGVIHTLQNIREFILERNHTSVINVAGVSVGSHPFNTIHFILERNLTHVMN VARFLVEEKTLLHVIIDFMLERNLTVKNVTKFSVADHTLKDIGEFILGKNHTNVRF VTRLSGVIHALQTIREFILERNLTSVINRRFLIKKESLHNIREFILERNLTSVMNVA RFLIKKQALQNIREFILQRNLTSVMSVAKPLLDLQHLFTIKQSMGVGKLYKCNDC KVFSNATTIANHYRIHIEERSTSVINVANFSDVIHNL
IP_138289.1	MS	ZSCAN3 1	MNIGGATLERNPINVRSVKGKSPVAMASLDTEESTQGKNHMKNAKCVGRSSSAH ALFSIRGYTLERSAISVSVAKPSFRMQGFSSISESTLVRNPISAVSAVNSLVSGHFL RNIRKSTLERDHHKGFDEFGAFSHHCNLRHFRIHTVPAELD
IP_278564.1	MS	ZNF808	MIVTKSSVTLQQLQIIGESMMKRNLSSVINVACFSDIVHTLQFIGNLILERNLTVNM IEARSSVKLHPMQNRRHTGKPHKCDDCGKAFTSHSHLVGHQRIHTGQKCKCH QCGKVFSPRSLAEHEKIHIF
IP_275012.1	MS	ZNF780 A	MKPCECTECGKTFSCSSNIVQHVKIHTGKRYNVRNMGKHLWMSICLNIRKFR VRNFVTIRSVDKPDLCTKNLLNTRILMRNLVNIKECVKNFHHGLGFAQLLSIHTS EKSLSVRNVGRFIATLNTLEFGEDNSCEKVFE
IP_204754.1	RP	ZFP91- CNTF	MPGETEERPEQDQEGEAAKAAPEEPQORPPEAVAAAPAGTTSSRVLGRGRD RGRAAAAAAAAAVSRRRKAEPYRRRRSSPSARPPDPVPGQPPQAAKSPSPVQGGK SPRLLCIEKVTTDKDPKKEEKEEEDDSDALPQEVSIASRPSRGWRSSRTSVSRHRDTE NTRSSRSKTGSLQLICKSEPNTDQLDYDVGEEHQSPGGISSEEEEEEMLISEEEI PFKDDPRDETYKPHLERETPKPRRKSQKVKKEEKEKEIKVEVEVEVEKEENEIRED EPPRKRGRRRKDDKSPRLPKRRKPPPIQYVRCMEGCGTVLAHPRYLQHHIKYQ HLLKKYVCPHPSCGRLFRLLQKQLLRHAKHHTDQRDYICEYCARAFKSSHNLA VHRMIHTGKPLQCEICGFTCRQKASLNWHMKKHDADSFYQFSCNICGKKEFKKDS VVAHKAKSHPEVLIAEALANAGALITSDILGTNPESLTQPSDGOGLPLPLPEPLGN STSGECLLLEAEGMSKSYCSGTERSIIHR
IP_098649.1	RP	INO80B- WBP1	MSKLWRRGSTSGAMEAPEPEGALELSLAGAHGHGVHKKKHKHKKHKKHKKH QEEDAGPTQPSPAKQLKIKLGGQVLGKTSVPTFTVIPEGPRSPSPLMVVDNEE EPMEGVPLEQYRAWLDEDSNLSPLRDLGGLGGQEEEEQRWLDALKEGELD DNGDLKKEINERLLTARQRALLQKARSQSPMLPLPVAEGCPPALTEEMLLKREE RARKRRLQAARRAEHKNQTIERLTKTAATSGRGGRGARGERRGGRAAAPAPM VRYCSGAQGSTLSFPPGVPAPTAVSQRPSPGPPRCSVPGCPHPRRYACSRGTQA LCSLQCYRINLQMRLLGGPEGPGSPLLATFESCAQE
IP_115174.1	RP	ZNF721	MYIGEFILERNPHTVENAKPLDSLQIFMRIRKIFILERNPTRVETVAKPLDSLQIFM HIRKFILEIKPKCKEKGAFKSYYSILKHKRTHTRGMSYEGDECRGL
IP_275016.1	RP	ZNF780 A	MNVRSVGKALIVVHTLFSIRKFIIPMRNLLYVGNVVRWPLDIANLLNILEFILVTSHL NVKTVGRPSIVAQALFNIRVFTLVRSPMNVRVSGRLLDFTYFNPNIRKLTQVKNLH NVRNVGNSFVVVQILINIEVILERNPLNVRNVGKPFDFICTLFDIRNCILVRNPLNV RSVGKPFDFICNLFDIRNCILVRNPLNVRNVERFLVFPPLAIRFTQVRRHLECKE CGKSFNRVSNHVQHQSIRAGVKPCECKGCGKGFICGSNVIHQHQIHSSEKLFVCKE WRTTFRYHYHLFNITKFTLVKNPLNVKNVERPSVF
IP_278870.1	RP	ZNF845	MNVARFLIEKQNLHVIEFILERNIRNMKNVTKFTVVNQVLKDRRIHTGEKAYKCK SL
IP_278888.1	RP	ZNF765	MSVARPSAGRPHLHTIIDFILDRNLTVKIVMKLSVSNQTLKDIGEFILERNYTCNE CGKTFNQELTLCRRRLHSGEKPYKYEELDKAYNFKSNLEIHKIRTEENLTSVMS

VARP

IP_278918.1	RP	ZNF813	MNVARVLIQKHTLHVIIDFILERNLTSVMNVARFLIEKHTLHIIIDFILEINLTSVMN VARFLIKKHTLHVITIDFILERNLTSVMNVARFLIKKQTLHVIIDFILERNLTSLSVA KLLIEKQSLHIIIQFILERNKCNCEGKTFCHNSVLVIHKNSYWRETSVMNVAKFLIN KHTFHVIIDFIVERNLRNVKHVTKFTVANRASKDRRIHTGEKAYKGEEYHRVFSH KSNLERHKINHHTAEKP
IP_280349.1	RP	ZNF587	MNAVNVGNHFFPALRFMFIKEFILDKSLISAVNVENPFLNVPVSLNTGEFTLEKGL MNAPNVEKHFSEALPSFIIRVHTGERPYECSEYGKSAEASRLVKHRRVHTGERPY ECCQCGKHQNVCCPRS
IP_280385.1	RP	ZNF417	MNAMNVGNHFFPALRFMFIKEFILDKSLISAVNVENPLLNVPVSLNTGEFTLEKGL MNVPNVEKHFSEALPSFIIRVHTGERPYECSEYGKSAETSRLIKHRRVHTGERPYE CCQSGKHQNVCCSPWS

¹MS, mass spectrometry; RP, ribosome profiling.

Supplementary Table 3: Examples of proteins encoded in the same gene and functionally interacting

Gene	Polypeptides ¹	Reference
CDKN2A, INK4	Cyclin-dependent kinase inhibitor 2A or p16-INK4 (P42771), and p19ARF (Q8N726)	(45)
GNAS, XLalphas	Guanine nucleotide-binding protein G(s) subunit alpha isoforms XL α s (Q5JWF2) and Alex (P84996)	(46)
mlpt	Mlpt peptide 1 (Q0VU43), peptide 2 (Q0VU42), peptide 3 (Q0VU41), peptide 4 (Q0VU40)	(47)
pri, tal	Peptide tarsal-less 1A (C0HJX4), 2A (C0HJX5), AA (A3RLR1), 3A (A3RLR0), polished rice ORF5 (A4PJ05)	(48)
Scl	Sarcolamban A (C0HJH4) and B (C0HJH3)	(49)
ATXN1	Ataxin-1 (P54253) and altAtaxin-1	(50)
Adora2A	A2A adenosine receptor (P30543) and uORF5	(51)
AGTR1	Angiotensin type 1a receptor (P25095) and PEP7	(52)

¹The UniProtKB accession is indicated when available.

5.3 Matériels et méthodes de la figure 3.2

Des cellules HeLa ont été ensemencées sur des lamelles de verre, puis transfectées après 24h avec l'ADNc d'altMiD51-2xflag-Vénus(1-173) et l'ADNc d'ACPM-Vénus(155-239). Les deux constructions ont été transfectées simultanément dans un ratio 1 :1. Après 24h de transfection, les cellules ont été rincées au PBS 1X à température pièce puis fixées avec 4% de paraformaldéhyde pendant 20 minutes à température pièce. Ensuite, elles ont été lavées 3 fois au PBS 1X. Par la suite, elles ont été perméabilisées avec 0.15% triton pendant 5 minutes. Ensuite, elles ont été bloquées dans le NGS 10% (*Normal Goat Serum*) pendant 20 minutes. Les mitochondries ont été marquées avec un anticorps Tom20 (Abcam, ab186734) dilué 1/500 dans le NGS 10% pendant la nuit à 4°C pour marquer les mitochondries. Le lendemain, les lamelles ont été lavées 3 fois au NGS 10% pendant 5 minutes. L'anticorps secondaire Alexa fluor 647nm (Thermo Fisher Scientific) a été incubé pendant 1h à température pièce. Les lamelles ont été lavées 3 fois pendant 5 minutes au PBS 1X avant d'être rincées à l'eau nano et montées sur des lames avec du *SlowFade*, un produit pour réduire le photoblanchiment, et du vernis à ongles. L'interaction entre alt-MiD51 et ACPM se retrouvent dans des *foci* mitochondriaux verts et les mitochondries sont marquées en rouge.

5.4 Matériels et méthodes de la figure 3.3

(A, C et D) Des cellules HeLa ont été ensemencées dans une plaque de 6 puits de 35mm. Elles ont été transfectées après 24h avec un vecteur vide ou altMiD51-2xflag avec l'agent transfectant GeneCellin selon les procédures du fabricant. Ensuite, les cellules ont été grattées dans du RIPA (1% Triton X-100, 1% Sodium Deoxycholate, 0.1% SDS, 1mM EDTA, 50 mM Tris-HCl pH7.5), «soniquées» 2 fois 5 secondes à 12% puis centrifugées à 10000g pendant 5 minutes. Un dosage des protéines avec le réactif BCA (Pierce, Waltham, MA, USA) a été effectué. Un total de 50 µg de protéines par puits a été mélangé avec un volume 1 :1 de Laemmli (62,5 mM Tris-HCl 0.5 M pH 6.8, 16,25% Glycérol, 2% SDS, 5% 2-mercaptoéthanol, bleu de bromophénol) et migré dans des gels Tris-Glycine. Chacune des expériences a un gel de concentration de 4% acrylamide et un gel de résolution de (A, D) 10% et de (C) 7% acrylamide. La migration du SDS-PAGE s'est faite à 200 volts pendant 50-75 minutes. Le transfert a été réalisé avec une solution de transfert (25 mM Tris, 200 mM glycine et 20% méthanol) pendant 1h à 4°C avec des membranes de PVDF. Les membranes ont été bloquées avec une solution de lait 5% dilué dans du tampon TBS-T (150 mM NaCl, 20mM Tris et 0,1% Tween 20) pendant 1h avant de recevoir les anticorps 16h à 4°C dans une solution de lait 5% dilué dans du PBS 1X. Voici les différents anti-

corps primaires : (A) anti-Drp1 P-S616 (Cell Signaling, 3455) 1/500, (C) anti-Opa1 (Cell Signaling, 67589S) 1/1000 et (D) anti-Mfn2 (Abcam, ab56889) 1/1000. Après 3 lavages au TBS-T de 5 minutes, l'anticorps secondaire de souris (Cell Signaling, 7076S) ou de lapin (Cell Signaling, 7074S) couplé à l'enzyme peroxydase de raifort a été dilué 1/10000 dans une solution de lait 5% dilué dans du PBS 1X et incubé avec la membrane pendant 1h à température pièce. Ensuite, 3 lavages au TBS-T pendant 5 minutes ont été effectués avant la révélation avec du réactif ECL (PerkinElmer, USA) et un ImageQuant LAS4000.

Le *stripping* des membranes fut effectué après la révélation. Un blocage de 1h fut effectué. Des anticorps primaires ont été incubés avec la membrane pour 16h comme précédemment : (A) anti-Drp1 (BD Transduction Laboratories, 611112) 1/500, (C et D) anti-actine (Sigma, A5441) 1/10000 et l'anti-Flag 1/2000 (Sigma, F1804). (A) L'anti-actine et l'anti-Flag ont pu être réalisés après un deuxième *stripping*. (A, C, D) Ensuite, 3 lavages au TBS-T pendant 5 minutes ont été effectués. L'anticorps secondaire a été incubé comme décrit précédemment avant la révélation. (B) Quantification avec le logiciel ImageJ de l'immunobuvardage fait en (A). Le test statistique utilisé fut le test t de Student non-apparié.