

Université de Montréal

Identification et modélisation cellulaire d'une mutation homozygote non-sens dans le gène  
*MLIP* causant une myopathie distale à apparition tardive

*Par*

Jean Mezreani

Département de Neurosciences, Université de Montréal, Faculté de Médecine

Mémoire présenté en vue de l'obtention du grade de Maîtrise ès Sciences (M.Sc.) en  
Neurosciences

Mars 2023

© Jean Mezreani, 2023

Université de Montréal

Unité académique : Département Neurosciences, Université de Montréal, Faculté de Médecine

---

*Ce mémoire intitulé*

**Identification et modélisation cellulaire d'une mutation homozygote non-sens dans le gène  
*MLIP* causant une myopathie distale à apparition tardive**

*Présenté par*

**Jean Mezreani**

*A été évalué par un jury composé des personnes suivantes*

**Inge Meijer**

Président-rapporteur

**Martine Tétreault**

Directeur de recherche

**Nicolas Chrestian**

Membre du jury

## Résumé

Les myopathies héréditaires représentent un large groupe de pathologies neuromusculaires progressives affectant l'intégrité générale, structurelle et fonctionnelle du muscle squelettique. Elles engendrent une myriade de symptômes, ternissant qualité de vie et autonomie, et pouvant même s'avérer mortelles. La pose d'un diagnostic juste peut être difficile, entravée notamment par une faible prévalence de certaines myopathies, l'importante hétérogénéité clinique existante, et le chevauchement symptomatique des diverses formes. Malgré les avancées récentes faites dans le domaine des techniques de séquençage qui contribuent grandement au dépistage, au moins 25% des individus atteints de myopathies demeurent sans diagnostic génétique.

Suivant l'investigation clinique d'un patient (Z46) atteint d'une myopathie distale à apparition tardive, l'analyse par séquençage ARN (RNA-seq) a révélé un variant non-sens homozygote de signification inconnue (VUS) à la fin de l'exon 5 du gène *MLIP*. Les niveaux d'expression génique altérés de « Protéine musculaire interagissant avec *LMNA* » (*MLIP*) et son partenaire « Lamine de type A » (*LMNA*) ont poussé à approfondir l'investigation. Davantage d'altérations -omiques furent identifiées par les techniques de RT-PCR, qPCR et WB, renforçant l'effet pathogénique du variant. Consolidant tous les résultats, le séquençage de longues lectures (LRS) a révélé un mécanisme d'épissage alternatif compensatoire de *MLIP*, qui tend à contourner et minimiser la production de transcrits arborant l'exon 5 muté.

La présente étude vise à : 1) apporter un diagnostic génétique définitif au patient Z46, posant le variant *MLIP* comme causatif de la myopathie distale; 2) démontrer le pouvoir diagnostique du RNA-seq dans la résolution de ce cas complexe par l'identification et l'élucidation du VUS; 3) témoigner de l'étendue de la caractérisation transcriptomique offerte par les longues lectures du LRS. Couplé à cela, la modélisation du variant pathogénique par CRISPR/Cas9 dans une lignée cellulaire de myoblastes humains immortalisés permettra l'évaluation des impacts morpho-fonctionnels; conférant un supplément d'informations relatives aux fonctions musculaires normales et pathologiques de *MLIP*, faiblement caractérisées jusqu'à présent.

**Mots-clés** : Myopathie, Transcriptomique, RNA-seq, *MLIP*, *LMNA*, LRS, Myoblaste, CRISPR/Cas9.

## Abstract

Hereditary myopathies represent a large group of progressive neuromuscular disorders affecting the general, structural and functional integrity of skeletal muscle. They cause a myriad of symptoms, impairing quality of life and autonomy, and can even be fatal. Making an accurate diagnosis can be difficult, hampered in particular by the low prevalence of certain myopathies, the significant clinical heterogeneity that exists, and the symptomatic overlap of the various forms. Despite recent advances in sequencing techniques that greatly assist in screening, at least 25% of individuals with myopathies remain without a genetic diagnosis.

Following the clinical investigation of a patient (Z46) with a late-onset distal myopathy, RNA-sequencing (RNA-seq) analysis revealed a homozygous nonsense variant of unknown significance (VUS) at the end of exon 5 of the *MLIP* gene. Altered gene expression levels of “Muscular *LMNA*-Interacting Protein” (*MLIP*) and its partner “A-type Lamin” (*LMNA*) prompted further investigation. More -omic alterations were identified by RT-PCR, qPCR and WB technics, reinforcing the pathogenic effect of the variant. Consolidating all results, Long-Read Sequencing (LRS) revealed a compensatory alternative splicing mechanism of *MLIP*, which tends to bypass and minimize the production of transcripts carrying the mutated exon 5.

The present study aims to: 1) provide a formal genetic diagnosis for patient Z46, positing the *MLIP* variant as causative of the distal myopathy; 2) demonstrate the diagnostic power of RNA-seq in resolving this complex case through identification and elucidation of the VUS; 3) testify to the breadth of transcriptomic characterization afforded by the long reads of LRS. Coupled with this, CRISPR/Cas9 modeling of the pathogenic variant in an immortalized human myoblast cell line will allow assessment of morpho-functional impacts; conferring additional information related to the normal and pathological muscles functions of *MLIP*, poorly characterized thus far.

**Keywords** : Myopathy, Transcriptomic, RNA-seq, *MLIP*, *LMNA*, LRS, Myoblast, CRISPR/Cas9.

# Table des matières

Résumé.....	iii
Abstract.....	iv
Table des matières.....	v
Liste des tableaux.....	ix
Liste des figures.....	x
Liste des sigles et abréviations.....	xi
Remerciements.....	xiv
1 – Introduction.....	1
1.1 Myopathie – Définition.....	1
1.2 Approches diagnostiques cliniques actuelles.....	2
1.2.1 L'importance de la recherche clinique.....	3
1.3 Dépistage génétique.....	4
1.3.1 Séquençage de première génération – Méthode Sanger.....	5
1.3.2 Techniques de séquençage de deuxième génération.....	6
1.3.2.1 Avantages du séquençage Illumina.....	10
1.4 Applications NGS.....	11
1.4.1 Analyse par panel de gènes.....	12
1.4.2 Séquençage de l'exome complet.....	12
1.4.3 Séquençage du génome complet.....	14
1.4.4 Séquençage transcriptomique.....	15
1.4.5 Comparatifs des forces et limitations.....	17
1.5 Séquençage de troisième génération – Séquençage de longues lectures.....	19

1.5.1 Séquençage nanopore.....	20
1.5.2 Parallèle avec ses prédécesseurs .....	23
1.5.3 Aboutissements et innovations.....	23
1.6 Présentation clinique du patient Z46.....	25
1.6.1 Aperçu de <i>MLIP</i> .....	26
1.6.2 Aperçu de la lamina nucléaire.....	28
1.7 Myoblastes humains immortalisés – Modèle d’étude d’intérêt.....	29
1.8 Mécanisme de défense bactérien CRISPR/Cas9 .....	29
1.8.1 Adaptation en tant qu’outil d’édition génomique.....	32
1.8.2 Réparation cassure double brin : Réparation NHEJ vs HDR.....	33
1.8.3 Principe de digestion d’hétéroduplexe .....	35
1.8.4 Culture unicellulaire – Génération de clones.....	36
1.9 Hypothèse et objectifs .....	36
2 – Méthodologie .....	38
2.1 Évaluation clinique et obtention des échantillons.....	38
2.1.1 Extraction des macromolécules de la biopsie musculaire .....	38
2.2 Techniques de séquençage employées.....	39
2.2.1 Séquençage transcriptomique (RNA-seq).....	39
2.2.2 Séquençage de longues lectures (LRS).....	41
2.2.3 Séquençage Sanger .....	43
2.3 Immunobuvardage (WB).....	43
2.4 Réaction en chaîne par polymérase (PCR) .....	43
2.4.1 Électrophorèse sur gel d’agarose.....	44
2.5 PCR quantitative (qPCR) .....	44

2.6 PCR précédée d'une transcription inverse (RT-PCR).....	45
2.7 Culture et entretien de la lignée cellulaire.....	45
2.7.1 Électroporation des myoblastes .....	46
2.7.1.1 Capture d'imagerie en microscopie fluorescente.....	46
2.7.2 Génération de clones par culture unicellulaire.....	46
2.7.3 Extraction de l'ADN des clones .....	48
2.8 Conception de l'édition génomique avec CRISPR/Cas9 .....	48
2.8.1 Synthèse des ARN guides .....	50
2.8.2 Test de digestion d'hétéroduplexe .....	51
2.8.3 Test de digestion avec enzyme de restriction.....	51
3 – Résultats .....	52
3.1 Article publié – Démonstration pathogénicité de <i>MLIP</i> .....	52
3.1.1 Abstract .....	53
3.1.2 Introduction.....	54
3.1.3 Results .....	55
3.1.4 Discussion.....	61
3.2 Modélisation de la mutation candidate <i>MLIP</i> par CRISPR/Cas9.....	66
3.2.1 Optimisation des paramètres expérimentaux CRISPR/Cas9.....	66
3.2.2 Conception de l'édition génomique KI de <i>MLIP</i> .....	68
3.2.3 Vérification du génotype des clones.....	70
4 – Discussion .....	72
4.1 Pertinence de l'utilisation du RNA-seq comme outil diagnostique .....	72
4.1.1 Pertinence de l'utilisation du LRS.....	73
4.2 Cohorte pédiatrique avec variants <i>MLIP</i> .....	74

4.2.1 Absence de phénotype cardiaque.....	75
4.3 Modélisation CRISPR/Cas9 .....	76
4.4 Optimisation du taux de rendement de KI.....	77
4.4.1 Promotion de la voie HDR .....	77
4.4.2 Promotion de l'utilisation de la matrice ADNss donneuse .....	78
4.4.3 Analyse par courbe de fusion à haute de résolution .....	78
4.5 Expansion du modèle cellulaire .....	79
4.5.1 Morphologie des myoblastes.....	79
4.5.2 Prolifération cellulaire .....	80
4.5.3 Différenciation cellulaire .....	80
4.5.4 Localisation intracellulaire de MLIP .....	81
4.5.4.1 Interactome de MLIP.....	82
4.5.5 Expérience de sauvetage.....	82
5 – Conclusion et perspectives.....	83
6 – Références bibliographiques.....	85
7 – Annexe.....	96
Methods .....	96
References.....	100
Supplementary Material .....	103



# Liste des tableaux

## Corps du mémoire

Tableau 1. – Comparatif de l'efficacité d'appel de variant entre WES, WGS et RNA-seq .....	19
Tableau 2. – Séquences des amorces utilisées pour LRS, PCR et RT-PCR .....	39
Tableau 3. – Séquences du matériel nucléique utilisé pour la conception des expériences d'édition génomique CRISPR/Cas9 .....	49

## Corps de l'article

Supplementary Table1. – Clinical Genetic Testing.....	103
---	-----

# Liste des figures

## Corps du mémoire

Figure 1. –	Vue d’ensemble et fonctionnement du pyroséquençage.....	8
Figure 2. –	Vue d’ensemble et fonctionnement du séquençage Illumina .....	9
Figure 3. –	Vue d’ensemble et fonctionnement du séquençage Nanopore.....	22
Figure 4. –	Carte génomique des exons et domaines protéiques de <i>MLIP</i> .....	27
Figure 5. –	Vue d’ensemble et fonctionnement du système CRISPR/Cas9 .....	31
Figure 6. –	Compétition des voies de réparation : NHEJ vs HDR .....	35
Figure 7. –	Schéma du pipeline bio-informatique RNA-seq .....	40
Figure 8. –	Préparation de la librairie et pipeline bio-informatique LRS.....	42
Figure 9. –	Préparation de la population cellulaire par dilutions séquentielles .....	47
Figure 10. –	Schéma d’assemblage des matrices ADN encodant les ARNg .....	50
Figure 11. –	Optimisation des paramètres expérimentaux CRISPR/Cas9.....	67
Figure 12. –	Conception du design KI CRISPR #3 de <i>MLIP</i> .....	69
Figure 13. –	Vérification du génotype des clones .....	71

## Corps de l’article

Figure 1. –	<i>MLIP</i> and <i>LMNA</i> respective expression in muscular tissue.....	57
Figure 2. –	Evaluation of nonsense variant effect on transcript balance of <i>MLIP</i> .....	60
Supplementary Figure 2. –	Exon usage of <i>MLIP</i> .....	104
Supplementary Figure 3. –	<i>MLIP</i> and <i>LMNA</i> respective expression in muscular tissue.....	105

## Liste des sigles et abréviations

ADN : Acide désoxyribonucléique

ADNc : ADN complémentaire

ADNss : ADN simple brin

ARN : Acide ribonucléique

ARNcr : ARN CRISPR

ARNg : ARN guide

ARNm : ARN messenger

ARNtracr : ARN CRISPR trans-activant

ATP : Adénosine triphosphate

Cas9 : Protéine 9 associée à CRISPR

CESGQ : Centre d'expertise et de services Génome Québec

CHUM : Centre hospitalier de l'Université de Montréal

CNV : Variabilité du nombre de copies

CRISPR : Courtes répétitions palindromiques groupées et régulièrement espacées

ddNTP : Didésoxyribonucléotide

DMD : Dystrophie musculaire de Duchenne

D-PBS : Tampon phosphate salin de Dulbecco

DSB : Cassure double brin

ECHO : Échocardiogramme

ECG : Électrocardiogramme

EMG : Électromyogramme

FBS : Sérum fœtal bovin

GFP : Protéine fluorescente verte

HDR : Réparation par recombinaison homologue

HRM : Courbe de fusion à haute résolution

Indel : Insertions et délétions de petite taille

IRM : Imagerie par résonance magnétique

kb : Kilobase, 1 000 paires de bases  
KI : *Knock-In*  
KO : *Knockout*  
LMNA : Lamine de type A/C (*Lamin A/C*)  
LRS : Séquençage de longues lectures  
Mb : Mégabase, 1 000 000 paires de bases  
MLIP : Protéine musculaire interagissant avec LMNA (*Muscular LMNA-Interacting Protein*)  
NGS : Techniques de séquençage de deuxième génération  
NHEJ : Jonction d'extrémités non-homologues  
NLS : Signal de localisation nucléaire  
NMD : Dégradation des ARNm non-sens  
ONT : Oxford Nanopore Technologies  
PAM : Motif de reconnaissance du proto-espaceur  
pb : Paire(s) de bases  
PCR : Réaction de polymérisation en chaîne  
PML : Promyelocytic Leukemia  
Poly(A) : Polyadénylée  
PPi : Pyrophosphate inorganique  
qPCR : PCR quantitative  
RNA-seq : Séquence transcriptomique / Séquençage de l'ARN  
RT-PCR : PCR précédée d'une transcription inverse  
SV : Variant de structure  
VUS : Variant de signification inconnue  
WB : Immunobuvardage  
WES : Séquençage de l'exome complet  
WGS : Séquençage du génome complet

*Malgré une vraisemblable confusion vis-à-vis de mon travail et certains de ses termes techniques, tel qu'adénosine triphosphate, séquençage transcriptomique ou le fameux -courtes répétitions palindromiques groupées et régulièrement espacées-, je dédie cet ouvrage à ma famille, pour votre écoute, support et amour.*

## Remerciements

Tout d’abord, je tiens à profondément remercier ma directrice de recherche, Martine Tétreault. Merci d’avoir cru en moi. Merci de m’avoir pris sous ton aile et offert cette incroyable opportunité scientifique qu’a été de joindre ton laboratoire novateur. Ta gentillesse, ta sagesse, et ton support sont sans limite, t’assurant que tes étudiants ne manquent de rien avec le plus stimulant des environnements de travail et ton appui irremplaçable. Tu as tout mis en place pour que l’on ait le meilleur des parcours possibles : entre des projets uniques, de l’équipement inédit, des conférences locales et internationales, des opportunités d’enseignement et de vulgarisation, des activités extracurriculaires, et j’en passe. J’ai hérité d’une chance inouïe en rejoignant tes rangs. Ayant su en profiter pleinement, c’est en tant que personne grandie et habilement outillée que j’en ressors; prêt à me dévouer aux projets futurs qui me seront alloués. Du fond du cœur, mille mercis pour cette formidable aventure.

Une attention particulière à mon ami et collègue Sébastien. Merci pour ton amitié incommensurable depuis le tout début de nos parcours universitaires respectifs. Merci pour les centaines de séances passées dans des cafés afin de rédiger notre article scientifique, nos mémoires, travailler sur nos cours, compléter des applications ou bien étudier. Et ça n’est pas sans oublier les soirées de jeux, les séances de gym, de golf et plus encore. Je ne peux mettre en mots à quel point ton support et ta camaraderie m’ont été précieux. Merci d’avoir été présent dans toutes ces multiples facettes de ma vie.

Je tiens également à remercier ardemment Jennifer et Lovatiana. Merci Jennifer pour ta jovialité, m’avoir aidé à perfectionner mon anglais, et avoir agrémente l’ambiance au labo avec tes multiples desserts préparés avec passion! Lovatiana, Lou, tu es une personne dotée d’une gentillesse sans pareille, tu es bienveillante et toujours prête à écouter et aider autrui. Je suis absolument ravi que nos chemins se soient croisés. Merci à toutes les deux pour votre camaraderie chaleureuse.

Mentions spéciales aux membres des laboratoires voisins. Merci Julie de m’avoir fait part de ton savoir infailible dans moult aspects des techniques expérimentales, te rendant disponible pour

une myriade de discussions stimulantes. De plus, merci pour toutes ces soirées d'astronomie passées en compagnie de Stéphanie et Quentin dans les différents coins de la province. Merci également à Kasia pour ton incroyable énergie propagatrice, pour ces nombreux fous rires que l'on a eus jusqu'à en avoir les larmes aux yeux avec nos humours farceuses, et pour m'avoir fait découvrir ces succulentes spécialités sucrées polonaises bien de chez toi.

Je n'oublie pas l'ensemble des membres du laboratoire et autres membres du CRCHUM que j'ai côtoyé de près ou de loin. Vous avez tous su rendre mon parcours unique à votre façon. Merci Annie, Camille M., Camberly, Marjorie, Valérie, Adrien, Nab, Saiyet, Alexie, Gaël, Stéphanie, Camille P., Renaud, Moustafa, Floriane, Ahmed, Éric.

Finalement, un immense merci à ma famille, mes amis et tous mes proches, vous n'avez jamais cessé de croire en moi, constamment en train de me pousser afin de donner le meilleur que j'ai à offrir. Votre amour, camaraderie et encouragement m'ont été d'une motivation implacable. Merci Samir, Fadia, Cynthia, Yanik, Sarah-Jade, Justine, Kencie, Éric, Leïla, Antoine, Rachel, Laurie, Joseph, Sami, Abanoub, Charles, Maëla.

# 1 – Introduction

## 1.1 Myopathie – Définition

Le terme myopathie, venant du grec myo - muscle et patheia/pathie - affection, constitue un terme générique désignant littéralement une affection musculaire. Il englobe un large éventail de pathologies musculaires pouvant être héréditaires ou acquises. Les myopathies héréditaires forment un vaste groupe hétérogène de maladies neuromusculaires génétiques portant atteinte principalement au tissu musculaire squelettique des individus affectés. Elles peuvent être classées suivant quatre grands axes, différenciable par l'origine et la nature du désordre : congénitales, métaboliques, mitochondriales et dystrophies musculaires (1-4). À cela se rajoute une panoplie de sous-catégories passant par les myopathies centrales, centronucléaires, à nemalines, à corps d'inclusion, myofibrillaire; jusqu'au canalopathies, myotonie congénitale; des myopathies avec un impact proximal (tronc, buste) ou distal (bras, jambes), et bien d'autres (4). Suivant l'importante hétérogénéité clinique surplombant ces pathologies, elles peuvent être transmises héréditairement par tous les modes de transmissions génétiques connus (4, 5).

Leurs répercussions peuvent être multiples tout dépendamment du défaut génétique causal, altérant tant l'intégrité générale, que la fonctionnalité ou encore la structure du muscle. Ces affections se manifestent au travers d'une myriade de symptômes variés; allant de la faiblesse, à la douleur, rigidité, crampes ou spasmes musculaires (4). Une démarche dandinante, de la maladresse ou des déformations, tel que de la pseudohypertrophie musculaire ou encore une déviation de la colonne vertébrale (scoliose/cyphose/lordose), sont un autre groupe de signaux cliniques rencontrés (4). En somme, la qualité de vie des personnes touchées est profondément ternie et ce avec des complications dans le temps. En effet, une composante progressive est observable dans ces pathologies, attribuable à la dégénérescence graduelle des fibres musculaires et causant une détérioration des symptômes éprouvés (4, 6). Avec ces complications, le faiblissement des fonctions motrices peut se transformer en une perte de celles-ci, handicapant plus sévèrement les individus atteints. Une importante perte d'autonomie peut en découler, certains individus se retrouvant dans l'impossibilité de travailler, s'occuper d'eux-mêmes ou



restreint à un fauteuil roulant. Dans les cas les plus sévères, lorsque les voies respiratoires et/ou cardiaques sont profondément impactées, des décès prématurés peuvent s'ensuivre (7-9). De façon générale, les phénotypes de myopathie se traduisent davantage dans un contexte pédiatrique, mais rencontrer des manifestations plus tardives n'est pas à exclure (6, 10). Ceci peut en partie être expliqué par les impacts variables que peuvent avoir différents acteurs génétiques défectueux sur l'organisme; certains ne devenant drastiques qu'après plusieurs années de dégénérescence musculaire (4, 6). Finalement, si une myopathie n'est pas d'origine génétique, alors elle peut être acquise suivant différentes provenances : immunitaire, environnementale, endocrinienne, infectieuse (11-16).

## **1.2 Approches diagnostiques cliniques actuelles**

Dans un cadre clinique, les méthodes diagnostiques couramment utilisées reposent principalement sur l'évaluation clinique des symptômes et une analyse morphologique du patient par un clinicien spécialisé, habituellement un neurologue. Très souvent, ces examens sont accompagnés d'une multitude de tests non invasifs, conférant un meilleur pouvoir diagnostique aux professionnels de la santé. On peut citer des tests électrophysiologiques, tel que l'électromyogramme (EMG), permettant d'attester le bon fonctionnement électrique des muscles testés, ou encore un électrocardiogramme (ECG), renseignant sur le rythme et l'efficacité du muscle cardiaque (17, 18). On compte également des tests à ultrasons, permettant entre autres la détection d'atrophie musculaire, de fibrose intramusculaire ou encore d'infiltration graisseuse du muscle (19). Un échocardiogramme (ECHO) est un type de test à ultrasons permettant spécifiquement une visualisation cardiaque, et conférant des informations sur la forme, texture, volume, capacité de pompage du cœur (20). L'imagerie par résonance magnétique (IRM) est un autre outil communément employé dans le milieu, et permettant l'observation et l'évaluation des tissus mous (organes et muscles) pour la précision du patron myopathique (21).

Sur une autre lancée, les récoltes de données biométriques, telles que la mesure du niveau de créatine kinase sérique ou encore le prélèvement et analyse d'une biopsie musculaire, sont des standards bien établis qui confèrent leur lot d'informations cruciales. Le muscle squelettique

étant riche en enzyme créatine kinase, une hausse de ce niveau d'enzyme sérique est un indicatif d'un désordre ou d'un dommage musculaire, la hausse étant dû au relâchement du contenu cytoplasmique des cellules musculaires dans le plasma (22). Il n'y a pas de normes absolues établies pour le niveau de créatine kinase sérique normal, puisqu'il varie énormément en fonction de plusieurs facteurs (âge, sexe, ethnie, musculature, activité physique). Selon une étude norvégienne, on parle d'hyperCKémie lorsque les mesures biométriques excèdent 210U/L chez la femme, ou >400U/L pour un homme de moins de 50 ans et >280U/L pour un homme de plus de 50 ans (23). D'autre part, dans l'analyse histologique d'une biopsie musculaire, un pathologiste va s'attarder à la présence ou non de nécrose, l'inflammation du tissu (myosite/polymyosite), la morphologie et la variation en taille de différentes fibres musculaires, une internalisation accrue des noyaux au niveau des fibres, des anomalies mitochondriales (microscopie électronique) et bien d'autres (24-27). L'ensemble de ces indicatifs va entre autres permettre au clinicien de rapidement distinguer la pathologie entre une myopathie nécrosante, métabolique, inflammatoire, mitochondriale, ou autre (24).

### **1.2.1 L'importance de la recherche clinique**

Toutes ces méthodes diagnostiques composent un formidable arsenal pour les cliniciens dans leur quête d'apporter les diagnostics les plus justes qu'ils soient à leurs patients. Mais malgré tout cela, il n'en demeure pas moins que certaines limitations s'imposent suivant ces approches. Établir efficacement des diagnostics pour les patients atteints de myopathie est une tâche complexe en raison de l'importante hétérogénéité clinique qui surplombe ces pathologies, et de manière concertée, de leurs multiples traits et symptômes qui peuvent se chevaucher. Couplé à cela, le fait que certaines formes de ces maladies soient à faible incidence, les cas d'étude peuvent se faire rares, compliquant le tout davantage.

En ce sens, le séquençage clinique constitue une technique additionnelle de choix pour l'établissement de diagnostic robuste. En ne s'attardant qu'au génome de l'individu pour repérer le ou les défauts géniques causatifs de la maladie, un généticien est en mesure d'apporter un supplément d'information sur un aspect non exploré jusqu'à là par les autres méthodes d'évaluation clinique. Ce complément d'information joue un rôle clé dans la pose du diagnostic

et sert considérablement à la classification efficace des myopathies. Et en n'étant pas impacté par l'obstacle du chevauchement symptomatique, le séquençage clinique possède ainsi un pouvoir de distinction se démarquant des autres techniques cliniques abordées. De plus, une forme rarissime de myopathie, avec manifestation atypique ou inusuelle, pourra être identifiée plus aisément au travers d'un dépistage génétique, qu'elle compose une découverte fortuite ou qu'elle ait déjà été répertoriée dans la littérature au préalable.

Apporter un diagnostic clinique juste aux patients est crucial. En plus de proposer une prise en charge appropriée suivant les cas, il peut également contribuer à une certaine source de réconfort psychologique et une baisse de l'anxiété chez ces individus. Même si un traitement n'est pas proposé dans l'immédiat, le simple fait de mettre le doigt sur la cause peut être rassurant pour eux, témoignant de la prise en charge de leur cas et l'avancée qui a été faite à leur endroit. Pour toutes ces raisons, il est primordial d'avoir les moyens de dépistages génétiques des plus efficaces qu'ils soient. Et pour que tout cela soit vrai, il est impératif que la littérature sur laquelle s'appuie ce dépistage génétique soit la plus développée, robuste et précise possible.

### **1.3 Dépistage génétique**

Le dépistage génétique est un test médical cherchant à établir un diagnostic moléculaire en identifiant le ou les défaut(s) génétique(s) causatif(s) de la maladie du patient. Cette méthode d'investigation a été rendue possible grâce aux nombreuses techniques de séquençage qui ont successivement vu le jour et qui se sont perfectionnées au fil des dernières décennies. Les techniques de séquençage ont pour but de donner à son utilisateur la séquence, ou l'enchaînement en nucléotides du matériel génétique à l'étude. Dans un contexte clinique, elles permettent aux généticiens d'évaluer les données -omiques (exomiques, génomiques, transcriptomiques) de patients afin de cibler et identifier le ou les variants pathogéniques en cause. Un variant pathogénique est une mutation, une altération génétique causant l'apparition d'une maladie ou augmentant la susceptibilité ou la prédisposition d'un individu et/ou sa progéniture de développer ladite maladie. Il existe plusieurs types de variant parmi lesquels on peut répertorier : les mutations faux-sens modifiant un nucléotide dans un codon et intervertissant subséquemment son acide aminé correspondant; les mutations non-sens

généralant un codon d'arrêt de traduction prématuré dans la séquence du transcrit (codons UGA, UAA, UAG); ou encore une insertion ou délétion (indel) de nucléotide(s), pouvant mener soit à l'ajout ou au retrait de codons. Si l'indel n'est pas divisible par trois, alors il cause un décalage du cadre de lecture des codons, altérant toute la chaîne en acides aminés résultante, et menant dans la majorité des cas à un codon d'arrêt prématuré. Les variants pathogéniques les plus impactant sont de loin les mutations non-sens et les décalages de cadre de lecture, causant des altérations structurelles et fonctionnelles majeures; soit au travers de portions de protéines manquantes, mal repliées ou avec des domaines protéiques non opérationnels (28). Outre les impacts fonctionnels et structurels, la biodisponibilité d'une protéine dans un tissu peut également être affectée. L'expression de la protéine peut être réprimée entièrement ou partiellement, et est en partie due à la voie de dégradation des ARNm non-sens (NMD) ou la dégradation par le protéasome; deux voies métaboliques de la cellule responsables d'assurer la qualité des protéines produites (29, 30). Il existe également des mutations dites silencieuses qui modifient le nucléotide d'un codon, généralement le troisième, sans pour autant altérer l'acide aminé qui en découle. Cela est dû au fait que le code génétique est un code dégénéré avec plusieurs codons codants pour un même acide aminé (31, 32). Bien que l'acide aminé ne soit pas modifié, certains variants silencieux peuvent être pathogéniques en abolissant ou créant de nouveaux sites d'épissage (33).

L'émergence, le développement et toute l'effervescence entourant cette branche scientifique sont en partie dus à une approche pionnière dans le domaine, la technique de séquençage établie par Frederick Sanger en 1977 : la méthode Sanger.

### **1.3.1 Séquençage de première génération – Méthode Sanger**

La méthode Sanger repose sur le principe de la terminaison d'élongation de chaînes, en employant des inhibiteurs de polymérisation de l'ADN, des didésoxyribonucléotides (ddNTP). L'élongation du brin d'ADN est freinée en raison de l'absence d'un groupement 3'-OH sur le didésoxyribose du nucléotide inhibiteur, groupement chimique normalement essentiel pour la ligature du prochain acide nucléique (34). Ainsi, en préparant quatre mélanges réactionnels séparés permettant la synthèse du brin sœur d'un ADN simple brin (ADNss) d'intérêt, et en dosant adéquatement et distinctement les inhibiteurs (ddATP, ddTTP, ddCTP, ddGTP) dans chacune des

préparations; alors on est statistiquement en mesure de générer toutes les longueurs de fragments complémentaires possibles (34). Donc, le premier mélange contiendra toutes les longueurs de fragments complémentaires se terminant par la base azotée adénine, le deuxième tous les fragments se terminant par une thymine, etc. Après réaction, chargement et migration des produits sur gel d'acrylamide, on peut reconstituer la séquence nucléique complémentaire en partant du plus petit fragment et en remontant (34).

Même si les séquences obtenues par cette méthode étaient incontestables avec une précision quasi parfaite, cette dernière s'avère être incroyablement laborieuse (35). Et malgré sa modernisation avec l'utilisation d'inhibiteurs couplés à des fluorochromes, permettant le multiplexage des ddNTP en un seul mélange réactionnel, et l'automatisation informatique de la lecture de la séquence au travers d'une électrophorèse capillaire, son débit de séquençage est encore très loin d'approcher celui proposé par les techniques de séquençage de deuxième génération (NGS) (36-38). À des fins de comparaisons, il a nécessité plus d'une décennie pour séquencer le génome humain entier à l'aide de la technique Sanger, tandis qu'aujourd'hui la même tâche est réalisable en moins de 24 heures avec les techniques plus récemment implantées (39, 40). La méthode Sanger n'a toutefois pas été délaissée de nos jours, et est encore couramment utilisée pour valider la séquence d'un ou plusieurs gènes (environ 20 gènes cibles). Néanmoins, elle ne peut certainement pas être employée pour évaluer un exome ou un génome dans leur entièreté, ce sont ses homologues de seconde génération qui rempliront ces rôles.

### **1.3.2 Techniques de séquençage de deuxième génération**

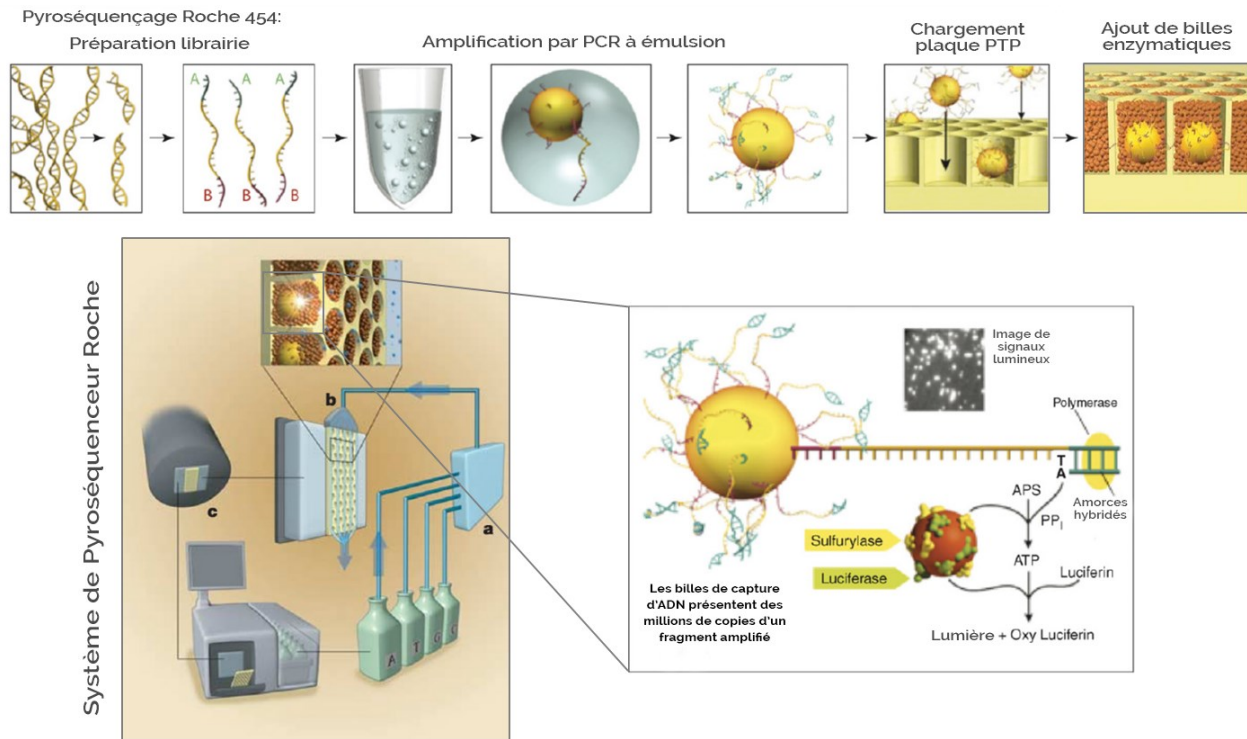
La deuxième génération de techniques de séquençage (NGS) inclut entre autres le pyroséquençage (Roche) et l'utilisation de nucléotides terminateurs fluorescents réversibles (Illumina), deux des techniques les mieux implantées en NGS et suivant le principe de séquençage par synthèse. Ces techniques sont différenciables de leurs prédécesseurs suivant trois grands points : la préparation préalable d'une librairie de séquençage, le principe de détermination du nucléotide, et le fait que ces processus soient massivement parallélisés (41-43).

Les deux approches démarrent donc avec une préparation de librairie de séquençage. Le matériel génétique est fragmenté en morceaux de quelques centaines de paires de bases (pb). Ces

longueurs (~100-400 pb) varient en fonction de la technique de séquençage employée et sont déterminantes dans la continuité de la procédure de préparation de la librairie. Les fragments sont ensuite ligués à des oligonucléotides adaptateurs à leurs extrémités, dénaturés à l'état simple brin, immobilisés sur billes (pyroséquençage) ou sur puces à ADN (Illumina), avant de finalement être amplifiés localement par PCR à émulsion (pyroséquençage) ou par amplification en pont sur matrice (Illumina) (41-43). Le but de l'amplification clonale est de générer des billes ou des agrégats locaux de plusieurs milliers de fragments identiques, permettant ainsi une bonne capture et quantification du signal à chaque ajout de nucléotide durant le processus de séquençage.

La seconde différence majeure repose sur la façon dont la séquence d'intérêt est déterminée, soit le concept de séquençage par synthèse. À l'inverse des méthodes de première génération, où la séquence est reconstituée suivant une électrophorèse de multiples fragments, ici les techniques de deuxième génération possèdent le pouvoir de caractériser le nucléotide ajouté, et ce durant le processus de synthèse du brin sœur. Par exemple dans le pyroséquençage, l'incorporation d'un nucléotide est détectée au travers d'un signal lumineux. La chambre de séquençage (*flowcell*) n'est exposée qu'à un seul type de nucléotide à la fois (41, 43). S'il y a incorporation de l'acide nucléique en question dans la séquence cible, alors sa ligature causera un relâchement de groupements pyrophosphates inorganiques (PPi) dans le puit réactionnel (41, 43). Des billes couplées aux enzymes sulfate adényltransférase et luciférase sont présentes à proximité des brins en processus de séquençage (41, 43). La molécule de PPi subira ainsi une cascade catalytique où, en la couplant à une molécule de 3'-phosphoadénosine 5'-phosphosulfate, elle sera convertie en molécule d'adénosine triphosphate (ATP) grâce à l'enzyme sulfate adényltransférase (41, 43). L'ATP nouvellement synthétisée servira ensuite de source énergétique à la luciférase dans sa réaction d'oxydation de son substrat, convertissant la luciférine en oxyluciférine (41, 43). Cette réaction, à la base de la capacité de bioluminescence de la luciole, va émettre des photons pouvant être détectés par l'appareil (41, 43). Ceux-ci seront émis à l'unisson et en quantité suffisante pour détection, grâce aux billes arborant des milliers de fragments identiques à leur surface. En résumé, la détection d'un signal lumineux confirme l'incorporation du nucléotide

auquel a été exposée la *flowcell*, au travers de la cascade catalytique subie par le PPI relâché. La Figure 1 survole le fonctionnement du pyroséquençage.

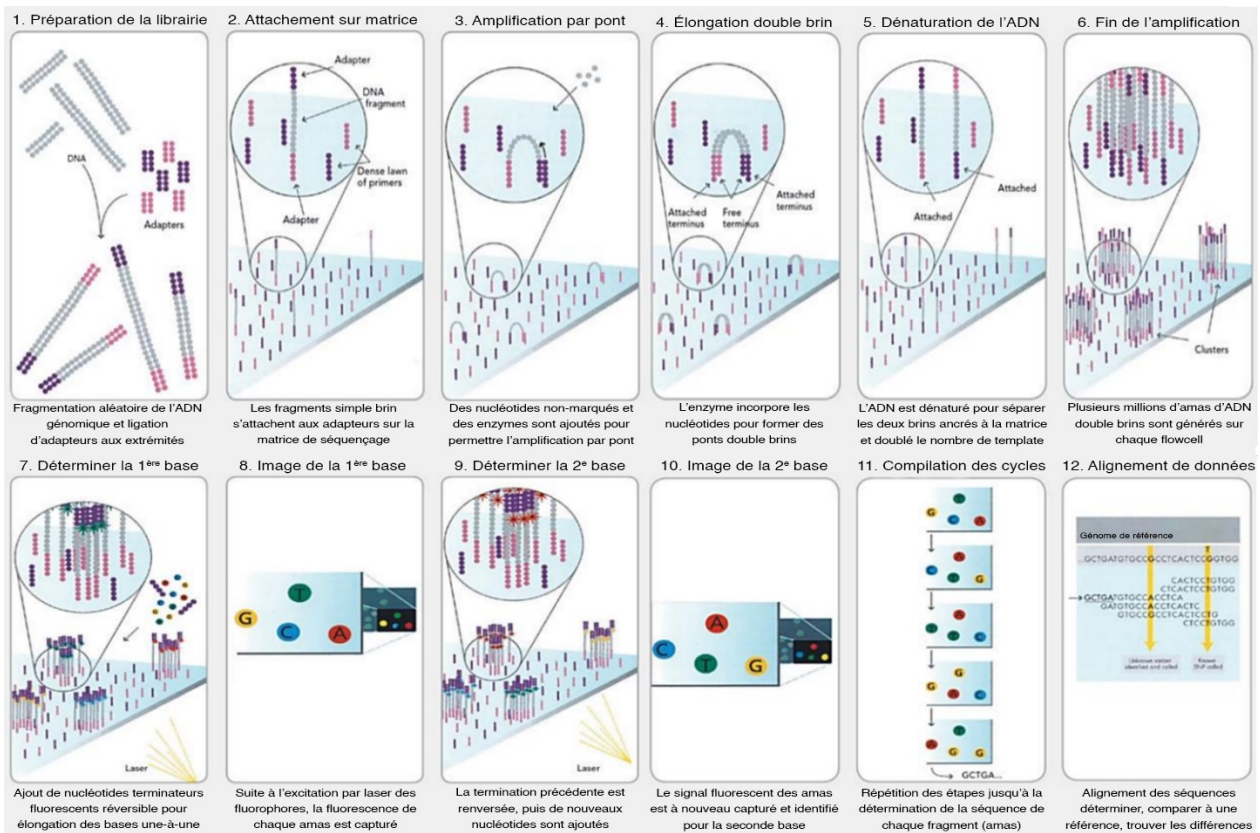


**Figure 1. – Vue d’ensemble et fonctionnement du pyroséquençage**

Préparation de la librairie d’ADN par fragmentation, ajout d’adapteurs, dénaturation, immobilisation sur bille et amplification locale par PCR à émulsion. Chargement d’une bille par puit sur plaque PTP accompagnée des billes enzymatiques. Exposition de la chambre de séquençage à un nucléotide à la fois et détection des puits qui émettent de la lumière. Figure adaptée à partir de Margulies *et al.* et Mardis (41, 44).

D’autre part, l’utilisation de nucléotides terminateurs fluorescents réversibles dans la méthode Illumina permet de freiner la polymérisation du brin sœur à chaque cycle d’addition, et de caractériser le nucléotide ajouté suivant le fluorophore y étant rattaché (45). La génération de la séquence se fait au travers d’une succession d’images, présentant des signaux fluorescents variant selon l’agrégat et le nucléotide s’y étant incorporé (42, 43). Suite à l’excitation laser du fluorophore et la capture du signal, le groupement fluorescent (rattaché à la base azotée) et le groupement 3’-O-allyle terminateur (présent sur le désoxyribose) sont clivés en employant le Tris-

(2-carboxyéthyl)phosphine comme agent réducteur (42, 43). En libérant le nucléotide de ses groupements supplémentaires et en régénérant la fonction 3'-OH sur le désoxyribose, le traitement chimique va permettre l'addition du prochain nucléotide (42, 43). Le processus est répété ainsi de suite de façon cyclique jusqu'à complétion du séquençage. Ce pouvoir de caractérisation de l'incorporation de nucléotide contribue à diminuer grandement la complexité de la tâche pour l'utilisateur. La Figure 2 résume le principe du séquençage Illumina.



**Figure 2. – Vue d'ensemble et fonctionnement du séquençage Illumina**

Préparation de la librairie d'ADN par fragmentation, ajout d'adaptateurs, dénaturation, immobilisation sur puce à ADN et amplification locale en pont sur matrice. Ajout des nucléotides terminateurs fluorescents réversibles à la chambre de séquençage, et capture et caractérisation cyclique de chaque nucléotide incorporé. Figure adaptée à partir de documentation Illumina (46).



Finalement, ces techniques se démarquent par leur processus massivement parallélisé qui permettent le séquençage en simultané de plusieurs millions de fragments, point phare accélérant considérablement le rythme de séquençage. Dans le pyroséquençage, le processus se déroule habituellement dans une plaque *PicoTiterLitre*, une plaque contenant des puits de l'ordre du picolitre ( $10^{-12}$ ) et entassés par millions (41). Avec la taille du puit et celle des billes portant la librairie à séquencer, il ne peut y avoir qu'une bille par puit (41). Cela assure qu'une seule séquence possible ne peut être générée par puit. L'ensemble de la plaque est ensuite exposé à un seul nucléotide à la fois durant le séquençage, et la technologie fait la distinction entre les puits qui émettent ou non des photons, en plus de quantifier cette intensité lumineuse lorsqu'un nucléotide est incorporé plus d'une fois dans la séquence (41). Tout ceci simultanément, pour les millions de puits en parallèle. Il en va de même du côté de l'approche Illumina, le séquençage se déroule sur une puce à ADN, soit une lame de verre avec de multiples canaux/couloirs possédant à sa surface plus de 10 millions d'agrégats locaux d'ADN par  $\text{cm}^2$  (47). À chaque ajout de nucléotide, l'appareil fait une capture en imagerie fluorescente pour l'entièreté des agrégats, générant des centaines de millions de séquences en simultané.

#### 1.3.2.1 Avantages du séquençage Illumina

Autre point révolutionnaire de ces plateformes NGS est l'accessibilité de ces techniques en termes de coût, radicalement moins chères que leur prédécesseur. À titre comparatif, le coût moyen pour le séquençage d'un million de paires de bases (megabase, Mb) s'élève jusqu'à 500\$/Mb pour la méthode Sanger, soit plus de 1 500 000\$ pour le séquençage d'un génome humain entier (48). Les plateformes NGS arrivent à une fraction des coûts de la méthode Sanger. Évalué à quelques dizaines de dollars par Mb pour le pyroséquençage, c'est cependant Illumina qui l'emporte haut la main, avec son meilleur tarif allant aussi bas que 0,10\$/Mb (48). Cette importante différence de prix s'explique par l'exigence davantage de réactifs pour le pyroséquençage, notamment avec ces enzymes et substrats pour réaliser la cascade enzymatique du PPI (43). Somme toute, ces percées ont éventuellement permis de franchir la barre symbolique du 1000\$ par génome.

Outre le coût étant un clair avantage en faveur de la plateforme Illumina, la chimie derrière le fonctionnement du pyroséquençage s'accompagne de certains désavantages. La cascade

enzymatique du PPI et les multiples étapes de lavage requises sont les principales limitations. L'accumulation de sous-produits dans le puit réactionnel peut affecter la cinétique enzymatique, impactant en retour l'intensité du signal lumineux produit (49). De l'autre côté, les étapes de lavage répétées nuisent à la qualité de la librairie à partir d'un certain nombre de cycles, pouvant mener à la perte de fragments d'ADN (49). À l'inverse, la caractérisation nucléotidique est plus juste dans la méthode Illumina pour deux raisons : il y a un frein de polymérisation à chaque cycle d'addition et un signal fluorescent distinct pour chaque base azotée (42, 45). Il n'y a ni fluctuations du signal perçu, ni perte de matériel.

Finalement, pour une même période de temps allouée, les quantités de données générées via l'approche Illumina sont considérablement plus élevées que son concurrent (43). Cela s'explique par l'ajout systématique de nucléotides à chaque cycle d'addition, à l'inverse du pyroséquençage qui procède par distribution rotationnelle des nucléotides, jonchée de temps morts où il peut y avoir absence d'extension. De plus, les différentes puces à ADN employées par Illumina ont des capacités de chargement bien plus grandes que les plaques *PicoTiterLitre*, résultant en des différences de l'ordre du milliard, en termes de lectures produites par expérience (41, 47).

## 1.4 Applications NGS

Avec l'important débit de séquençage pouvant être atteint grâce à ces percées technologiques, cela a permis l'instauration de plusieurs applications dans les domaines de la science fondamentale et clinique. Elles permettent des analyses de plus grande envergure, avec des temps et coûts moindres. De toutes celles qui ont pu émerger, on peut souligner l'analyse par panel de gènes, le séquençage de l'exome complet (WES), le séquençage du génome complet (WGS) et le séquençage transcriptomique (RNA-seq), tous des outils décisifs dans l'établissement de diagnostics moléculaires. Chacune possède ses avantages et limitations, avec des coûts, évaluations, couvertures, renseignements, interprétations et demandes en infrastructures informatiques variables. En fonction des besoins et l'hypothèse scientifique, il faut choisir celle qui convient le mieux au cas d'étude, ou opter pour la combinaison d'approches, tout dépendamment des ressources pouvant être déployées dans le processus d'investigation.

### **1.4.1 Analyse par panel de gènes**

L'analyse par panel de gènes est une méthode routinière bien implantée dans le domaine clinique. En fonction de la pathologie suspectée par le clinicien évaluateur, ce dernier peut recommander un panel de gènes approprié à faire passer au patient pour tenter de cerner le défaut génique causatif. Un panel de gènes est un test génétique rapide investiguant une liste précise de candidats : cette liste étant basée sur des gènes établis et associés dans la littérature à une pathologie donnée ou connexe à celle suspectée par le clinicien. Cette approche peut aussi bien servir à cerner le gène défectueux causatif, qu'à préciser un diagnostic lorsqu'un cas de figure présente une hétérogénéité génétique importante. On retrouve une multitude de panels de gènes adaptés et déjà en place pour tester toutes sortes de maladies, conditions, syndromes (panels neuromusculaires, ataxies, épilepsies, cancers), mais ils peuvent également être personnalisés en fonction des besoins (50-52). C'est une application qui constitue un juste milieu entre l'analyse d'un seul gène, comme par exemple *BRCA1* ou *BRCA2*, causes connues de développement du cancer du sein chez la femme, et le séquençage de l'exome complet (WES), qui va regarder toute la portion codante du génome (53). La rapidité et l'interprétation relativement simple des résultats de ce test en font un outil grandement pertinent pour un cadre clinique. Cependant, en n'évaluant qu'une portion très restreinte du génome de l'individu, il peut être facile pour cette méthode de passer à côté du défaut génétique, surtout lorsque le gène défectueux en question n'a pas encore été associé dans la littérature à une quelconque pathologie. En réponse à cela, il peut être pertinent d'employer des approches avec des champs d'analyse plus élargis pouvant procurer davantage d'informations, susceptible de mieux répondre aux besoins du cas à l'étude.

### **1.4.2 Séquençage de l'exome complet**

Le séquençage de l'exome complet (WES) est une approche qui va se concentrer sur l'investigation de toute la portion codante du génome, soit l'exome d'un individu. L'exome englobe l'entièreté des exons des gènes et représente approximativement 1% du génome, soit environ 30 000 000pb (54). C'est une piste d'investigation de choix, puisque cette portion du matériel génétique est la plus susceptible de contenir les variants pathogéniques, en raison des instructions essentielles à la synthèse protéique qui y sont encodées. Les mutations en son sein

seront les plus enclines de causer des défauts structuraux et fonctionnels protéiques, et d'entraîner conséquemment des phénotypes pathologiques divers.

L'investigation exclusive de l'exome repose sur le principe de préparation de la librairie de séquençage suite à l'extraction de l'ADN génomique. Selon la plateforme de séquençage choisie, il existe quelques variations durant le processus de préparation de la librairie, mais dans l'ensemble le traitement est similaire. Il y a une première étape de fragmentation de l'ADN par voie mécanique (ultrasonication) ou biologique (enzymatique) (55). Les extrémités des fragments sont ensuite émoussées et ligaturées avec des adaptateurs, ces derniers ayant des affinités diverses pour les régions codantes du génome. Par exemple, les trousseaux de préparation d'Illumina (*Illumina Nextera Rapid Capture Expanded Exomes* ou *Illumina TruSeq Exome Enrichment*) sollicitent des centaines de milliers de sondes avec des affinités accrues pour les régions non traduites (UTR) flanquant en amont et en aval les portions codantes des gènes (55). S'ensuit une capture spécifique des fragments adaptés, pour finalement laisser place à un enrichissement ciblé des régions exoniques (55). Le séquençage subséquent de cette librairie fournira une importante couverture de l'exome de l'individu.

Certes, les demandes en ressources informatiques et interprétation des résultats sont plus grandes, mais ça n'est pas sans apporter son lot d'informations cruciales. Un des avantages qui est ainsi permis avec cette approche, comparativement au séquençage Sanger ou l'analyse par panel de gènes, c'est le fait qu'il ne soit pas nécessaire de connaître le ou les gènes cibles à séquencer. En investiguant tout l'exome d'un individu, l'application favorise les découvertes fortuites au travers d'identification d'acteurs génétiques qui étaient jusqu'alors non suspectés de causer la pathologie à l'étude. Les chances d'établir un diagnostic moléculaire sont ainsi plus importantes. À cet effet, le WES a fait ses preuves en tant qu'outil diagnostique robuste à multiples reprises, identifiant et associant une myriade de variants pathogéniques à leurs pathologies connexes (56). Selon plusieurs études, un taux d'établissement de diagnostics moléculaires approchant les 25-50% peut lui être associé en ce qui a trait l'évaluation de maladies diverses, témoignant de son efficacité d'utilisation (57).

### 1.4.3 Séquençage du génome complet

Sur un plan encore plus large, le séquençage du génome complet (WGS) va prendre en compte l'entièreté du génome humain : tous ses chromosomes et leurs régions codantes/non codantes, régions régulatrices, régions satellites répétées en tandem, télomères, tout. L'ADN total récolté est séquençé en son entier, et représente une investigation d'environ 3 milliards de pb (54). Pourquoi explorer un champ aussi élargi, alors qu'il a été établi que les variants dans les régions codantes étaient les plus susceptibles d'altérer la fonction protéique? Et bien même si quelque peu dilué par la quantité colossale de données générées, cette application est en mesure de procurer les mêmes renseignements que le WES, avec un supplément d'information cruciale concernant les régions non codantes et régulatrices du génome. En effet, de plus en plus d'études démontrent les nombreux impacts que peuvent avoir des altérations dans ces régions sur la bonne fonction et expression d'une protéine donnée. Par exemple, une multitude de pathologies ont pu être reliées à des mutations identifiées au sein d'amplificateurs (*enhancer*) (58, 59). Un amplificateur est un élément génétique régulateur avec pour fonction d'augmenter la transcription de son gène cible situé en aval. Un défaut dans sa séquence causerait un mauvais amarrage de la machinerie protéique responsable de l'initiation et l'amplification de la transcription, menant à une baisse ou une répression d'expression du gène (59). L'amplificateur est distant de plusieurs milliers de pb du gène qu'il régule, et peut même parfois agir à partir d'un chromosome différent (60). Ceci explique la nécessité d'avoir une importante couverture des régions non codante pour mener à bien l'identification de ces variants.

Les variants d'épissage sont une autre catégorie de mutations difficilement identifiable par WES. Il s'agit de mutations situées à la frontière exon-intron et provoquant des événements d'épissage aberrants. Tout dépendamment de l'altération, on peut soit observer une rétention intronique anormale ou alors une exclusion exonique, tous deux menant à la génération de transcrits fautifs. Cela peut en retour impacter la traduction, la bonne fonctionnalité et la biodisponibilité de la protéine correspondante. Certains variants localisés profondément dans l'intron peuvent aussi affecter l'épissage et résulter dans la formation d'un pseudo-exon qui ne devrait normalement pas apparaître dans la séquence du transcrit (61).

D'autre part, les variants de structure (SV) forment de larges restructurations chromosomiques de plusieurs centaines à plusieurs milliers de pb (62). Les SV sont typiquement générés lors des processus de réplication, de réparation et de recombinaison de l'ADN et englobent entre autres des événements d'insertions, inversions et translocations de larges fragments génomiques, fructifiant la biodiversité génétique d'une espèce (63). Cependant, des complications ou des défauts rencontrés lors de ces remaniements structurels peuvent mener à des duplications et délétions de segments chromosomiques. Cela affecte en retour le nombre de copies d'un ou plusieurs gènes présents sur un chromosome ou une chromatide. Cette catégorie de SV porte le nom de variabilité du nombre de copies (CNV), et outre la distorsion de l'architecture chromosomique normale, ces CNV peuvent également avoir un impact fonctionnel important sur une protéine précise, notamment au travers d'une altération de son bon profil d'expression dans un tissu donné (62, 64). Cela peut être expliqué directement par le gain ou la perte de gènes entiers, provoquant un effet de dosage, ou suivant la modification de la relation spatiale entre un élément régulateur et son gène correspondant, n'adoptant plus leur conformation usuelle (62).

L'époque où l'on octroyait la connotation péjorative d'ADN poubelle à la portion non codante du génome est révolue depuis bien longtemps. Ses rôles de régulation génétique, de préservation de l'architecture chromosomique ou de maintenance de signatures épigénétiques ne sont que quelques exemples de son importance capitale au sein du génome. On peut citer le projet *ENCODE* lancé par la National Human Genome Research Institute, un projet cherchant à établir une base de données robuste avec l'identification de tous les éléments fonctionnels du génome, à la fois chez l'humain et la souris. En définitive, l'utilisation de WGS peut être déterminante dans la pose d'un bon diagnostic moléculaire, les variants énumérés ci-haut témoignant des cas de figure qui peuvent être rencontrés.

#### **1.4.4 Séquençage transcriptomique**

Bien que des piliers dans l'établissement de diagnostics moléculaires, les taux de succès diagnostique pour le WES et le WGS combinés pour une variété de pathologies données se situent approximativement entre 25 et 50% (57). Leur atout majeur reposant principalement sur l'identification de variants, ils fournissent en contrepartie très peu à aucun renseignement sur les

impacts fonctionnels et cliniques de ces derniers, résultant en l'identification de variants de signification inconnue (VUS). Ainsi, une approche permettant d'améliorer l'interprétation de ces VUS consiste à incorporer de l'information génétique fonctionnelle. C'est là qu'entre en jeu la plus grande force du séquençage transcriptomique (RNA-seq), apportant un aperçu fonctionnel additionnel et permettant une caractérisation approfondie des variants identifiés.

En séquençant de l'ADN complémentaire (ADNc) issu de la rétrotranscription d'ARN, le nombre de lectures perçu pour un même gène varie proportionnellement à la quantité initiale de transcrits ARN présente dans l'échantillon lors de sa récolte. Cela confère une représentation fidèle du niveau d'expression du gène dans le tissu analysé. La préparation de la librairie de séquençage débute avec une récolte de l'ARN total, suivant des traitements à la nucléase ADN (DNase) et/ou une purification sur colonne. À noter, que l'ARN total d'une cellule forme un tout très diversifié (ARNm, ARNr, ARNt, snRNA, miRNA, siRNA, lncRNA), dont 80 à 90% de sa composition est de l'ARN ribosomique (ARNr), non pertinent dans le cadre de l'investigation (65). En tirant profit des queues polyadénylées (poly(A)) qu'arbovent les ARN messagers (ARNm) matures à leur extrémité 3', il est possible d'utiliser des billes d'appâts arborant des oligos poly(T) pour leur capture spécifique par complémentarité des bases azotées (65). S'ensuivront finalement les étapes classiques de préparation de librairie : rétrotranscription en ADNc, fragmentation, ajout d'adaptateur, immobilisation, amplification et séquençage. Le séquençage de cette librairie fournira une importante couverture du transcriptome du tissu analysé, soit tout ce qui est activement transcrit dans les cellules composant l'échantillon.

Au même titre que le WES et le WGS, le RNA-seq est également capable d'appeler de variants dans les régions codantes des gènes, vis-à-vis du transcriptome tissulaire. En plus, les dérèglements fonctionnels engendrés par un variant peuvent être étudiés au travers d'analyses de données d'expression différentielles. Ainsi, afin de tirer pleins bénéfices des nombreux avantages proposés par le RNA-seq, il est primordial que l'investigation soit effectuée à partir d'un tissu affecté par la pathologie. Cette double facette informationnelle identification-interprétation fonctionnelle a pour effet d'augmenter les chances d'établir un diagnostic moléculaire. Dans cette perspective, de nombreuses études ont identifié moult variants pathogéniques reliés à une variété de désordres musculaires, grâce au pouvoir du RNA-seq. Dans la majorité des cas, il s'agissait de

variants produisant des événements d'épissage aberrants, soulignant le RNA-seq comme application NGS de choix pour l'identification efficace de cette catégorie de variant. Nicolau *et al.* ont élucidé un cas mettant en lumière un variant faux-sens et un variant d'épissage en trans, causant le saut de l'exon 8 au sein des transcrits de *CAPN3* (Calpaïne 3) dans la dystrophie musculaire des ceintures récessive 1 (LGMDR1) (66). Hamanaka *et al.* ont identifié dans le gène *NEB* (Nébuline), un nouveau variant intronique profond et une nouvelle mutation silencieuse pathogénique chez trois patients présentant une myopathie à némaline (67). Un cas un peu plus complexe et impressionnant, Cummings *et al.* ont été en mesure d'identifier un variant intronique *de novo* dans le gène *COL6A1* (Collagène VI A1) générant un événement de gain d'épissure et causant la rétention anormale d'un pseudo-exon récurrent et observé chez 27 patients (57). L'incorporation de ce pseudo-exon est maintenant la cause la plus fréquente pour les myopathies liées au collagène VI.

Une autre des grandes forces du RNA-seq, c'est sa capacité à déceler des fusions de gènes. Ces événements anormaux constituent des signatures moléculaires pour plusieurs formes de cancer, et peuvent se produire dans différentes circonstances (68). Elles peuvent être due à des réarrangements structuraux des chromosomes, menant à la fusion de gènes, et subséquemment à la transcription et traduction de transcrits et protéines chimériques (69). La transcription simultanée de gène voisin ou des événements irréguliers d'épissage d'ARNm sont de plus en plus reconnus comme les causes conduisant à la formation d'une grande partie des fusions de transcrits, justifiant la performance du RNA-seq pour la détection de ce type de variant (69).

Entre l'appel de variants dans les régions codantes du transcriptome tissulaire, l'apport d'une facette fonctionnelle supplémentaire pour une meilleure caractérisation des VUS, la détection de tous les types de variants d'épissage possibles, et l'identification de fusions de transcrits; l'importante diversité informationnelle qu'apporte le RNA-seq en fait un des outils NGS des plus puissants qui soient.

#### **1.4.5 Comparatifs des forces et limitations**

Toutes ces applications NGS sont irréfutablement de puissants outils diagnostiques. En fonction du cas d'étude, il faut savoir les distinguer afin de les utiliser à bon escient pour leurs forces



respectives. Ici, les panels de gènes sont écartés de la discussion car ils sont le fruit de décennies de recherche active, suivant le développement de la littérature et l'identification de gènes causals : développement et identification permis grâce aux autres applications NGS.

Toutes ces approches performant brillamment dans l'identification de variants dans les régions codantes, mais à des profondeurs variables. Étant la spécialité du WES avec une vitesse d'analyse supérieure, le WGS s'avère performer de mieux en mieux dans cet aspect, pouvant détecter davantage de variants codants de haute qualité (70). Et par rapport aux différents types de variants des régions non codantes, la large couverture fournie par le WGS permet leur entière identification. Si le généticien évaluateur a accès au tissu affecté par la pathologie, la question n'a pas lieu d'être, il faut opter pour le RNA-seq pour tirer profit des précieuses informations fonctionnelles du tissu. Le RNA-seq est capable de déceler la quasi-majorité des variants codants et non codants, vis-à-vis des gènes activement transcrits dans la biopsie, et possède le meilleur pouvoir pour la résolution des VUS.

En définitive, le WES est la méthode d'investigation la plus rapide et la moins chère, se concentrant exclusivement sur les régions codantes; le WGS est le seul avec le potentiel de détecter de manière exhaustive tous les types de variants génomiques (71); le RNA-seq fournit la plus grande diversité d'information par sa double facette identification-interprétation de variants, focalisant son investigation sur le transcriptome tissulaire. Le Tableau 1 récapitule les forces respectives des différentes applications NGS dans l'identification de variants.

**Tableau 1. – Comparatif de l'efficacité d'appel de variant entre WES, WGS et RNA-seq**

Comparaison de l'efficacité d'appel de variant pour les applications de séquençage NGS			
Types de variants	WES	WGS	RNA-seq
SNV codant	✓✓	✓✓✓	✓✓
Indels codants	✓✓	✓✓	✓✓
Variants d'épissage	✓✓	✓✓✓	✓✓✓
Variants d'épissage non codants profonds	✓	✓✓	✓✓✓
Variants régulateurs	✓	✓✓	✓
SNV intronique	✓	✓✓✓	✓
CNV	✓✓	✓✓✓	—
Fusion de gènes	—	✓	✓✓
— : Aucune détection.			
✓ : Quelques variants de ce type sont détectés.			
✓✓ : La plupart des variants de ce type sont détectés.			
✓✓✓ : Tous les variants de ce type sont détectés.			
SNV: Variant affectant un seul nucléotide; CNV: Variabilité du nombre de copies			

\* Table récapitulative traduite à partir de Tétreault *et al.* (71).

## 1.5 Séquençage de troisième génération – Séquençage de longues lectures

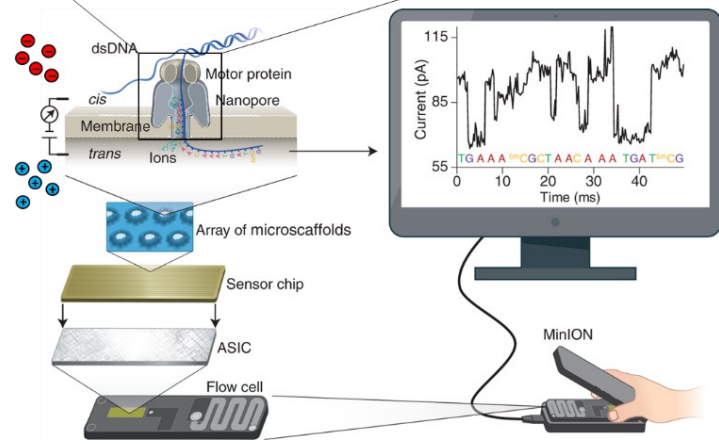
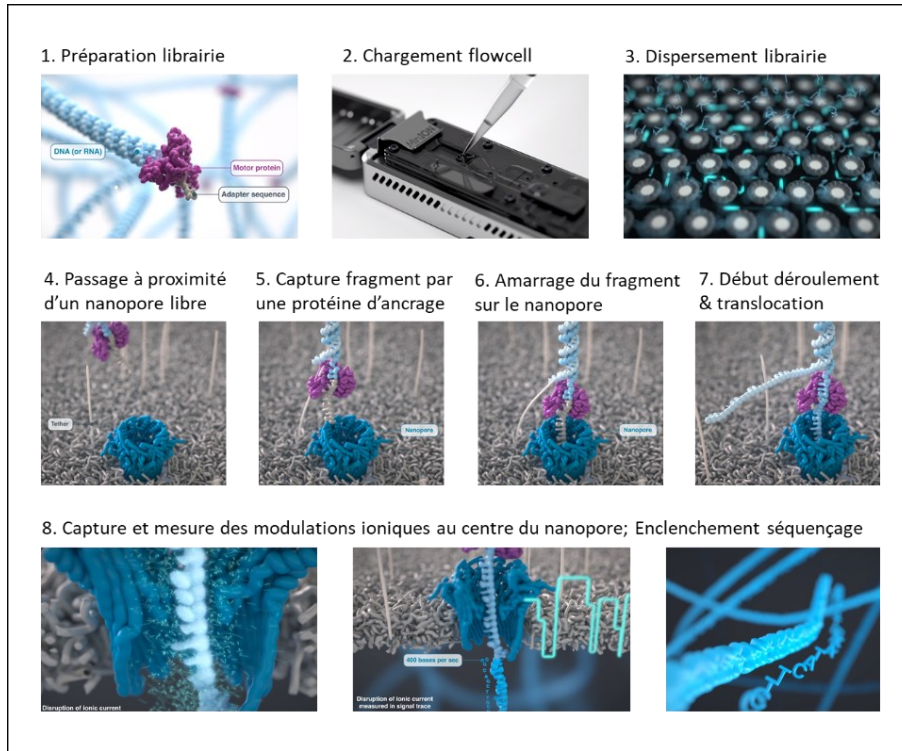
Dans cette troisième grande révolution des techniques de séquençage de longues lectures (LRS), les technologies émergentes proposent des lectures de fragments considérablement plus longues que leurs prédécesseurs NGS; étant question ici de plusieurs milliers de paires de bases (kilobases, kb) à plusieurs dizaines de kb, une étude ayant même développé un protocole pour l'assemblage de fragments ultra-longes s'étalant de 100kb à 882kb (72). De plus, la sensibilité de ces nouvelles technologies est telle, qu'il est possible de générer des lectures à partir d'une seule molécule d'ADN, délaissant les étapes d'amplification clonale localisées (73). Ces innovations sont rendues possibles grâce à de nouvelles percées faites dans le domaine du séquençage. S'éloignant des approches classiques de terminaison d'élongation et de séquençage par synthèse, la librairie préparée par l'utilisateur est directement séquencée par l'appareil, tirant profit de signature moléculaire propre à chacune des bases azotées de l'ADN et de l'ARN (73).

### 1.5.1 Séquençage nanopore

Le principe de séquençage proposé par Oxford Nanopore Technologies (ONT) repose sur la prémisse que : les interactions d'une molécule analyte avec un pore encastré dans une membrane isolante, dans un champ électrique constant, auront pour effet de modifier les niveaux de conductance ionique dans le micro-environnement de ce pore (74). Ainsi, la technologie ONT évalue l'altération d'un flux ionique au sein d'un pore de quelques nanomètres de diamètre (nanopore), suivant le passage d'un simple brin d'ADN ou d'ARN en son centre. Les bases azotées ayant des tailles, effets stériques et fonctions chimiques variables; elles vont moduler le flux ionique de manière caractéristique, leur conférant une « empreinte électrique », une signature discernable et reconnaissable par la technologie (75-77). Les chambres de séquençage (*flowcell*) sont composées d'une membrane résistante à l'électricité ornée de 2048 nanopores, et séparant le côté *cis* (chargé négativement durant le séquençage) du côté *trans* (chargé positivement) (78). Le séquençage de la librairie est régi en parallèle dans la *flowcell* par la disponibilité des nanopores de séquençage environnant. Outre la qualité de la librairie, il y a deux constituants absolument essentiels pour le bon déroulement du séquençage : la protéine motrice greffée aux extrémités des fragments d'ADN lors de la préparation de la librairie, et le cœur même du nanopore, spécialement sa portion profonde et étroite, servant de région sensitive pour la détection des fluctuations ioniques (78).

La protéine motrice utilisée par ONT n'a pas été divulguée par ce dernier, mais fort à parier qu'il s'agirait de la polymérase à ADN phi29, enzyme grandement employée dans divers champs de la biologie moléculaire (78, 79). Suite à la capture et l'amarrage d'un fragment d'ADN sur un nanopore libre, l'activité hélicase de la polymérase va initier le déroulement du double brin d'ADN et va contrôler sa translocation simple-brin au travers du nanopore (78, 80). Sous l'effet du voltage imposé dans la *flowcell* et avec la nature négative des groupements phosphates de l'ADN, les fragments de la librairie vont chercher à traverser la membrane afin de quitter le côté négatif *cis*, et se rendre du côté positif *trans* (78). La protéine motrice va contrôler le rythme de translocation à hauteur de 400-450 pb/seconde, rythme optimal pour le bon fonctionnement de la technologie (78, 80).

Du côté des nanopores, ceux successivement employés par ONT sont l' $\alpha$ -hémolysine de *Staphylococcus Aureus* (~1,2 à ~2,4nm), la porine A de *Mycobacterium smegmatis* (MspA, ~1,2nm) et la porine *Curli specific genes* dans *Escherichia coli* (CsgG, ~1nm) (74, 78, 81, 82). Tous ces canaux ont subi de nombreuses éditions précises par bio-ingénierie pour optimiser : la conformation pour l'amarrage de la protéine motrice; l'ouverture du nanopore pour régir la vitesse de translocation; et la région sensitive pour une évaluation plus précise des modulations ioniques (81). La portion sensitive du nanopore est reliée à une nanoélectrode sur une puce captatrice qui relaye ensuite le signal électrique au circuit ASIC (*Application-Specific Integration Circuit*), faisant les captures et mesures finales des fluctuations électriques (78). Le signal électrique brut est capté en k-mer de longueurs variables (dépendamment de la région sensitive du nanopore) avant d'être converti en séquences nucléiques correspondantes, soit par le logiciel opérateur MinKNOW, soit par traitement bio-informatique à l'aide de l'outil Guppy (78, 83). La Figure 3 récapitule des éléments clés du séquençage nanopore proposé par ONT.



**Figure 3. – Vue d'ensemble et fonctionnement du séquençage Nanopore**

La librairie de séquençage est préparée suivant les directives du manufacturier et chargée dans la *flowcell*. Les fragments vont se répandre jusqu'à trouver un nanopore libre, où ils seront capturés, amarrés, déroulés, puis transloqués au travers du pore. Les modulations ioniques provoquées par le passage de la molécule sont ressenties dans la portion étroite du nanopore et le signal est relayé à une puce sensitive, puis vers le circuit ASIC. Figure adaptée à partir de Wang *et al.* et des animations produites par ONT (78, 84).

### 1.5.2 Parallèle avec ses prédécesseurs

Avec un taux théorique de 2048 nanopores séquençant en parallèle à hauteur de 400-450 pb/sec (70pb/sec pour de l'ARN), la technologie de séquençage proposée par ONT permet l'atteinte d'un débit de séquençage considérable. Ainsi, les multiples applications ayant d'écoulées des progrès des NGS sont également transposables avec cette nouvelle méthode de séquençage, la seule différence reposant sur la procédure de préparation de la librairie. Il faut toutefois noter qu'en générant des lectures à partir de molécules uniques et sur de très longs tronçons, le taux d'erreur de la méthode nanopore est plus élevé que les autres techniques de séquençage. S'élevant initialement autour de 7%, il est respectivement de 0,001%, 0,49% et 0,7% pour ses homologues Sanger, pyroséquençage et Illumina (35, 42, 80, 85). Cependant, il est facilement atténué en compilant une multitude de lectures couvrant la même séquence, rehaussant la technologie à une précision quasi parfaite. Et grâce aux récentes améliorations apportées aux *flowcells*, aux librairies de séquençage et aux algorithmes basés sur les réseaux neuronaux de ONT, le taux d'erreur a chuté à environ 5% (72). Des raffinements technologiques ne cessent de voir le jour, ONT ayant mis à disposition de nouveaux kits de séquençage Q20, évitant la nécessité d'étapes de corrections renforcées.

### 1.5.3 Aboutissements et innovations

Cette nouvelle approche de séquençage apporte son lot d'avantages avec elle. Dans un cadre plus pratique, la puissance computationnelle de cette nouvelle technologie relaye en temps réel des informations relatives au bon déroulement du séquençage. Elle peut renseigner sur le rythme de séquençage, la qualité de la librairie, la qualité de la *flowcell*, les longueurs des fragments et les gènes séquencés tout dépendamment de l'application en cours. Selon les renseignements fournis, l'utilisateur peut apporter des ajustements à son expérience ou la mettre à l'arrêt. Cela contribue grandement à une meilleure préservation des réactifs, notamment des *flowcells*.

Sur un plan plus informationnel, la flexibilité de la méthode de séquençage ONT permet une caractérisation nucléotidique plus diversifiée, couvrant à la fois ADN et ARN sans passer par d'étape de rétrotranscription (73, 86). Des données relatives au taux d'expression génique peuvent ainsi être recueillies sans détour, en séquençant directement de l'ARN (86). De manière

concertée, l'analyse de transcrits complets sur une seule lecture, qu'ils soient de l'ARN ou de l'ADNc, permet une quantification plus précise des niveaux d'isoformes ainsi que leur composition exonique, offrant le potentiel de révolutionner le domaine de l'épissage alternatif (87). Sur une lancée similaire, il est possible pour la première fois d'évaluer les marques épigénétiques du méthylome directement sur le matériel nucléaire extrait; la technologie faisant la distinction entre les bases méthylées de celles qui ne le sont pas (88). Cela permet de sauter les étapes classiques de préparation de librairie sollicitant le bisulfite; traitement qui convertit des cytosines non-méthylées en uracile, mais préserve les cytosines méthylées telles quelles (89). Lors de ce type d'investigation, il est important de garder à l'esprit que les séquences bisulfites ne sont pas complémentaires au génome référence, et qu'il faut faire la distinction entre une thymine génomique authentique et une thymine convertie à partir d'une cytosine non-méthylée (C non-méthylée, Traitement bisulfite → U, Amplification PCR → T) (89). Dans ces circonstances, les outils d'alignement conventionnels ne peuvent être employés, nécessitant des outils bio-informatiques adaptés spécifiquement pour cette tâche. En somme, le séquençage nanopore aura pour effet d'énormément décomplexifier la tâche de l'utilisateur lors de ce type d'investigation, œuvrant par séquençage direct des marques épigénétiques plutôt qu'indirecte.

Outre son puissant pouvoir de caractérisation nucléotidique, le plus grand atout du séquençage nanopore demeure indéniablement la longueur des séquences qu'il procure sur une même lecture. Une des principales limitations en NGS est justement les lectures relativement courtes, nécessitant des efforts d'assemblage et d'alignement accrus pour être interprétées adéquatement. Le séquençage de longues lectures possède le potentiel de combler ces limitations. La longueur maximale pouvant être séquencée est théoriquement limitée par les manipulations de l'utilisateur lors de la préparation de la librairie de séquençage (90). Ainsi, cette méthode de séquençage est l'outil de choix pour l'instauration de génome de référence encore plus fiable avec des lectures ultra-longues de plusieurs centaines de kb, voire quelques Mb. Jusqu'à très récemment, il y avait toujours eu des lacunes dans les séquences chromosomiques, spécifiquement au niveau des régions centromériques et télomériques. Grâce à cette nouvelle technologie, Miga *et al.* ont été en mesure d'assembler *de novo* la séquence du chromosome X

humain dans son entièreté, d'un télomère à l'autre (91). Il n'est que question de temps avant que ces quelques trous dans le génome soient tous comblés.

Dans un autre ordre d'idée, la technologie va exceller dans la quantification d'expansions répétées ou autres régions avec motifs répétés en tandem; séquences qui représentent un véritable défi en NGS lors de l'assemblage des lectures courtes (92). Dans un contexte d'évaluation de variants pathogéniques, en plus d'être plus qu'apte à identifier les variants et indels dans les régions codantes, le séquençage nanopore va notamment exceller dans l'identification des SV, des CNV et des fusions de gènes en raison de l'importante taille que représentent ces restructurations (93). Également, la couverture de plusieurs jonctions d'exons sur une même lecture devrait grandement faciliter l'identification des variants d'épissage, conférant à la technologie un redoutable pouvoir de caractérisation transcriptomique. En définitive, le séquençage nanopore est d'une modularité sans précédent, conférant d'énormes potentiels lors d'investigations multiples à travers sa flexibilité analytique et ses longues lectures.

## **1.6 Présentation clinique du patient Z46**

Le patient (Z46) s'était présenté à un rendez-vous médical pour consultation à l'âge de 5 ans par rapport à ses performances physiques, inapte à rivaliser avec ses pairs (incapacité à patiner ou courir aussi vite que les autres enfants). Il n'avait cependant pas été diagnostiqué pour quoique ce soit, les symptômes jugés non débilitants à l'époque. Vers la cinquantaine, il avait développé davantage de faiblesse musculaire affectant les membres distaux du corps (bras et jambes), avec une faiblesse accrue au niveau des mains, ainsi que les muscles responsables de la dorsiflexion des pieds. Dans la soixantaine, il avait développé une lordose lombaire accentuée et une démarche de Trendelenburg, démarche caractérisée par un balancement prononcé des hanches pour compenser une atteinte ou parésie des muscles abducteurs des hanches (94). De la douleur et des crampes musculaires avaient également été enregistrées chez ce dernier. Aucun phénotype cardiaque n'a été noté. Des tests ECG, ECHO et Holter (24 heures) furent réalisés, mais les résultats n'étaient pas remarquables. Un test EMG avait démontré quelques altérations dans divers enregistrements électriques : petites unités motrices polyphasiques avec fibrillations et ondes aiguës positives.



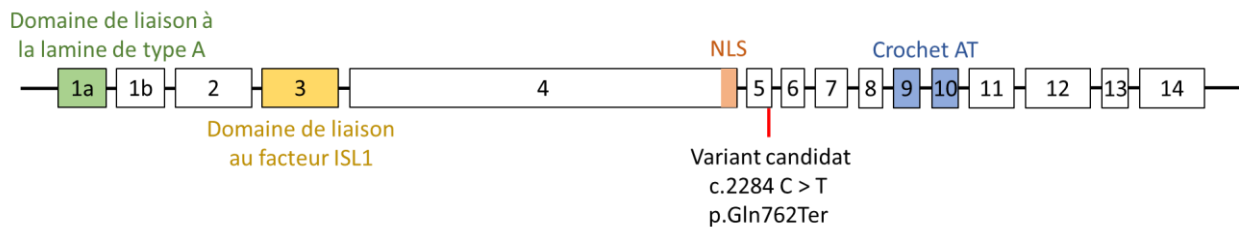
La mesure du taux sérique de créatine kinase s'élevait à 888U/L, ce qui est nettement supérieur à la norme, plaçant l'individu dans la fourchette d'hyperCKémie (23). L'analyse histologique de biopsie musculaire a pour sa part révélé plusieurs points : une fibre nécrotique, plusieurs fibres avec des noyaux internes, des fibres hypertrophiées, une atrophie des fibres de type 1 et des fibres de type 2 contenant de petites vacuoles cerclées. Toutes ces mesures biométriques consolident l'atteinte musculaire dont le patient souffre, mais demeurent non spécifiques pour caractériser une catégorie spécifique de myopathie.

Finalement, une analyse par panel de gènes avait ressorti deux variants hétérozygotes dans *RAPSN* (rapsyn) et *SCN4A* (Sous-unité  $\alpha$  4 du canal sodium voltage-dépendant), mais pour lesquelles les symptômes du patient ne concordaient pas avec la littérature rattachée aux variants identifiés. Une analyse RNA-seq a mené à l'identification d'une mutation non-sens homozygote à la fin de l'exon 5 du gène *MLIP*. Malgré une faible littérature entourant la fonction moléculaire de cette protéine, *MLIP* a été retenu comme variant candidat, jugé pertinent pour investigation approfondie. Les processus d'identification et caractérisation du VUS seront plus détaillés dans les sections des résultats et discussion.

### **1.6.1 Aperçu de *MLIP***

*MLIP* (Protéine musculaire interagissant avec *LMNA* - *Muscular LMNA-Interacting Protein*) est une protéine à expression ubiquitaire, avec une forte association au niveau du tissu musculaire squelettique et cardiaque, et identifiée relativement récemment dans la dernière décennie par Ahmady *et al.* en 2011 (95, 96). Elle est encodée par un gène du même nom, présent sur le locus 6p12.1 du chromosome 6 (95). Son interactant protéique principal, tel que désigné dans son nom, est la lamine de type A (*LMNA*), une des composantes de la lamina nucléaire qui tapisse l'intérieur du noyau cellulaire. À ce jour, le rôle moléculaire de *MLIP* n'a pas encore été complètement élucidé. Elle est suspectée d'avoir de multiple implication au niveau du maintien de l'homéostasie cardiaque à titre de répresseur transcriptionnel, notamment au travers de son interaction avec le facteur de transcription *ISL1*, un important progéniteur cardiaque (97, 98). *MLIP* aurait ainsi un rôle de régulateur sur la voie de signalisation *Akt/mTOR/FOXO1*, en partie responsable à l'adaptation au stress cardiaque (99, 100). D'autre part, des évidences ont démontré une

colocalisation entre MLIP et les corps PML au niveau du noyau (95). Cette superstructure nucléaire d'appariement aléatoire de protéines non apparentées peut tant être reliée à des processus nucléaires (réplication ADN, transcription, silence épigénétique), qu'à des troubles cellulaires causant la séquestration, modification ou dégradation de protéines nucléaires (101). Hormis son observation, aucune fonction n'a pu être attribuée à cette interaction (95). Plus récemment, des articles scientifiques ont mis en évidence un lien de cause à effet entre de multiples mutations de *MLIP* (majoritairement des troncations prématurées) et divers désordres musculaires (myopathie, rhabdomyolyse, myalgie), supportant un rôle important dans le muscle (102-105). Le mécanisme pathologique n'est toutefois pas encore connu.



**Figure 4. – Carte génomique des exons et domaines protéiques de *MLIP***

Représentation schématique de la carte génomique de *MLIP* illustrant le domaine de liaison à LMNA, le domaine de liaison à ISL1, la séquence NLS et le crochet AT respectivement sur les exons 1a, 3, 4 et 9-10. Le variant pathogénique identifié par RNA-seq apparaît à la fin de l'exon 5. Figure adaptée à partir de Lopes Abath Neto *et al.* et Cattin *et al.* (102, 106).

*MLIP* est un gène qui possède une architecture transcriptomique fortement complexe, avec entre autres 14 exons, permettant la génération de plus d'une douzaine d'isoformes différents issus de l'épissage alternatif (106). Certains transcrits sont nucléaires et d'autres cytoplasmiques le tout variant en fonction du tissu d'expression (106). On peut compter quatre domaines protéiques au sein de *MLIP*, tel qu'illustré dans sa représentation schématique à la Figure 4. Les exons 1a et 1b sont mutuellement exclusifs et déterminent la capacité des isoformes à pouvoir interagir avec LMNA ou non, puisque le domaine de liaison se situe dans l'exon 1a (106). L'exon 3 renferme le domaine de liaison au facteur de transcription ISL1 (106). À la fin du large exon 4 se trouve le

signal de localisation nucléaire (NLS), signal dictant le destin nucléaire ou cytoplasmique de la protéine une fois traduite (106). Finalement, un motif de liaison à l'ADN, un crochet AT, est présent conjointement sur les exons 9 et 10 (102). Ce domaine confère à MLIP un pouvoir d'interaction avec la chromatine, pour un potentiel rôle de remodelage ou répression transcriptionnelle devant encore être confirmé (107). Finalement, dans certaines branches de la littérature, la protéine peut prendre l'appellation CIP pour *Cardiac ISL1-Interacting Protein*, mais sa dénomination principale demeure MLIP.

### **1.6.2 Aperçu de la lamina nucléaire**

La présence d'organelles arborant des fonctions spécialisées et compartimentant la cellule est une signature des eucaryotes. Le noyau a pour sa part le rôle de protéger et séquestrer le matériel génétique du reste du contenu cytoplasmique. Sa structure est composée d'une double paroi de bicouche lipidique, une paroi interne et une externe, prenant l'appellation d'enveloppe nucléaire (108). Outre ses rôles de protection et contrôle des macromolécules entrantes et sortantes, la présence d'une entité sphérique autour de l'ADN permet également sa régulation et son remodelage (109). Ce contrôle sur l'ADN est notamment possible grâce à la lamina nucléaire, un réseau complexe de protéines filamenteuses intermédiaires tapissant la paroi nucléaire interne. Ce réseau est constitué de quatre protéines principales : les lamines A, B1, B2, C et leur expression est exclusive aux cellules somatiques (110). Les lamines de type A et C sont issues de l'épissage alternatif, toutes deux encodées dans le gène *LMNA*, tandis que les lamines B1 et B2 sont respectivement encodées par les gènes *LMNB1* et *LMNB2* (111-113).

La lamina nucléaire possède une myriade de protéines interactrices, dont MLIP est un partenaire spécifique pour la lamine de type A. Les rôles transcriptionnels, d'interactions et de remodelage de la chromatine par la lamina seraient ainsi dictés par ses différents partenaires protéiques, ayant chacun des affinités et des modes d'interaction variables avec différentes régions de la chromatine ou avec d'autres sous-partenaires protéiques (114). Intrinsèquement, la lamina nucléaire participe également à la maintenance et la stabilité de l'intégrité nucléaire (115). Lorsque dérégulée et/ou altérée, un groupe de maladies du nom de laminopathies peuvent en découler. Elles englobent une multitude de pathologies dont des troubles neuromusculaires, tels

que des cardiomyopathies dilatées, la dystrophie musculaire d'Emery-Dreifuss, et d'autres formes de dystrophies musculaires (116). Parmi les acteurs génétiques en cause, un nombre considérable de variants pathogéniques ont été identifiés à l'intérieur du gène *LMNA*, le reliant à une multitude de ces désordres neuromusculaires et mettant en évidence le rôle important qu'il occupe au sein du tissu musculaire (116-121).

## 1.7 Myoblastes humains immortalisés – Modèle d'étude d'intérêt

L'utilisation de myoblastes humains immortalisés constitue un modèle d'étude pertinent dans ce projet pour plusieurs raisons. Tout d'abord, les myoblastes se sont établis au fil du temps comme un modèle standard dans l'étude, la caractérisation et le développement morpho-fonctionnel du muscle squelettique (122, 123). D'autre part, les myoblastes du patient sont suspectés d'être affectés par la mutation, d'où l'intérêt de vouloir le transposer *in vitro*. Également, avec le haut niveau de potence des myoblastes, ceux-ci peuvent être différenciés en d'autres sous-types cellulaires, permettant l'évaluation des impacts fonctionnels de la mutation candidate dans plusieurs types de cellules à partir d'un modèle initial (124). Finalement, Ahmady *et al.* ont démontré l'implication de *MLIP* dans le processus de différenciation des myoblastes C2C12 en myotubes, montrant la pertinence de travailler avec un tel modèle (97).

L'utilisation de modèles plus classiques, tels que le ver *C elegans* et le poisson-zèbre *Danio Rerio*, est impossible en raison de l'absence de *MLIP* de leur génome, *MLIP* étant une innovation des amniotes (97). L'utilisation d'un modèle murin est trop avancée pour ce stade-ci du projet, en plus que Cattin *et al.* avait produit une souris « *Knockout* » (KO) *MLIP*<sup>-/-</sup> qui n'exhibait pas de phénotype musculaire remarquable, fort probablement dû à la complexité transcriptomique de *MLIP*, résultant en une répression d'expression incomplète de ce dernier (99, 106).

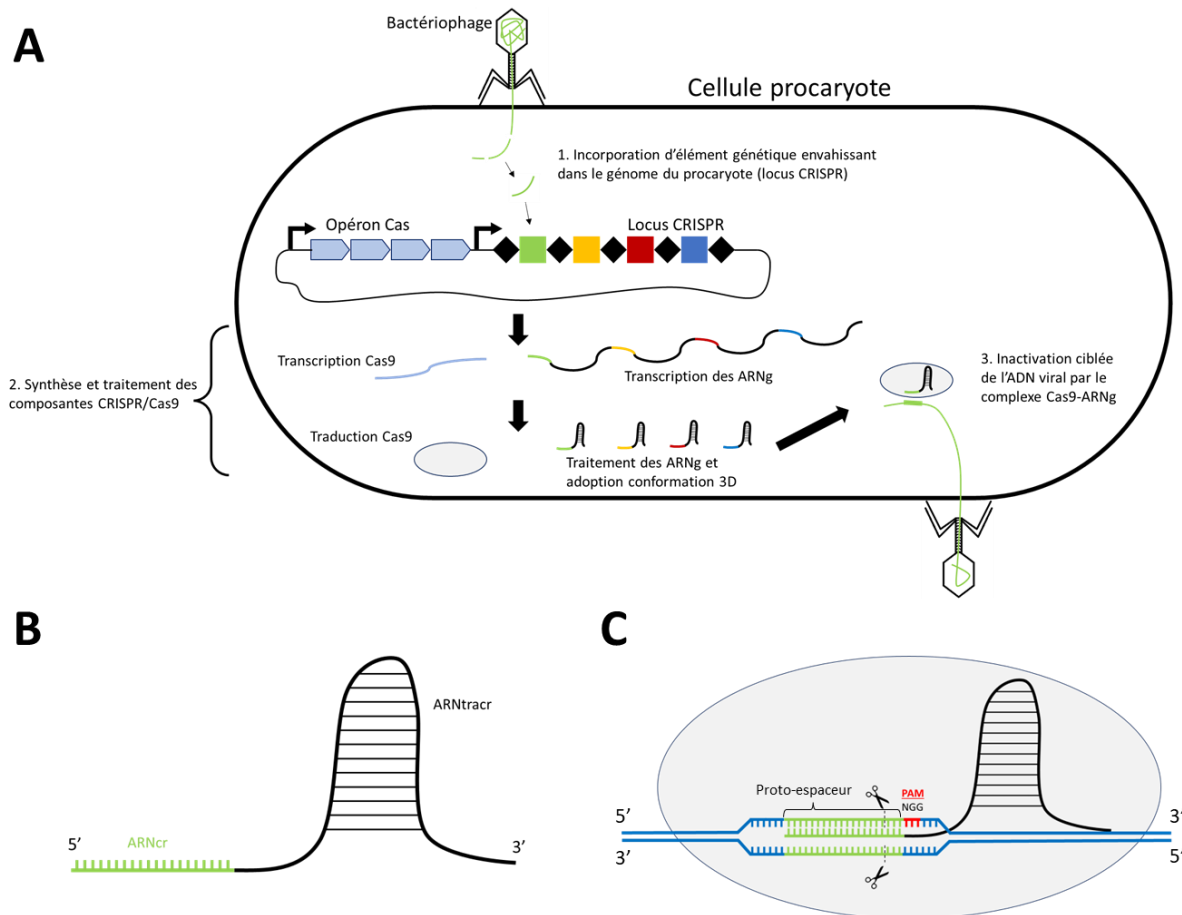
## 1.8 Mécanisme de défense bactérien CRISPR/Cas9

Avant d'être le puissant outil d'édition génomique que l'on connaît aujourd'hui, CRISPR/Cas9 est avant tout un système de défense à immunité adaptative chez les procaryotes, permettant aux bactéries et aux archées de se protéger contre les infections aux bactériophages (125). Le but de ce système est d'inactiver de l'ADN viral exogène en induisant une cassure double brin (DSB) dans

le matériel envahissant (126). Deux composantes essentielles régissent le bon fonctionnement de ce mécanisme : la protéine 9 associée à CRISPR (Cas9) et un ARN guide (ARNg) se complexant à la Cas9 et lui servant de guide pour un ciblage spécifique de l'ADN exogène. La protéine Cas9 est une endonucléase, classe d'enzyme avec le pouvoir catalytique de cliver un fragment d'ADN en son centre (126). Elle est encodée dans l'opéron *Cas* du chromosome bactérien, juste à côté du locus CRISPR (Courtes répétitions palindromiques groupées et régulièrement espacées), qui lui renferme les instructions essentielles à la synthèse de différents ARNg (126). Un ARNg possède deux segments distincts : l'ARN CRISPR (ARNcr) du côté 5' et l'ARN CRISPR trans-activant (ARNtracr) sur le reste de la séquence. L'ARNcr représente un tronçon d'une vingtaine de nucléotides servant au ciblage direct de l'ADN viral par appariement des paires de bases façon Watson-Crick (126). L'ARNtracr pour sa part constitue une séquence palindromique conservée qui adopte une structure tridimensionnelle précise, servant d'échafaudage moléculaire pour l'ancrage de l'ARNg dans le corps de la Cas9 (126). Ces deux composantes doivent impérativement se complexer ensemble afin d'être pleinement fonctionnelles.

Afin de ne cibler qu'exclusivement de l'ADN exogène et ne pas couper de l'ADN bactérien du « soi » par erreur, un mécanisme de sécurité fut implanté évolutivement. En plus de nécessiter l'appariement nucléotidique entre l'hybride ADN-ARNg, une seconde condition doit être remplie au travers du motif de reconnaissance du proto-espaceur (PAM) pour induire la DSB (126, 127). Cette séquence PAM est un motif de trois à cinq nucléotides, présente uniquement sur l'ADN cible. Elle est située directement en aval du proto-espaceur, soit directement du côté 3' de la portion ADN ciblé par l'ARNcr (127). C'est un motif qui sert de point de reconnaissance pour l'endonucléase exclusivement, permettant la distinction entre l'ADN du « soi bactérien » et le « non-soi ». La présence du motif PAM est un indicatif qu'il s'agit bel et bien d'ADN exogène, tandis que son absence désigne l'ADN du « soi » (127). Pour la protéine Cas9 de *Streptococcus pyogenes* (SpCas9), l'endonucléase Cas la plus communément employée en recherche, le motif PAM correspond au triplet NGG, où N représente n'importe quel nucléotide (127). Avec toutes les conditions d'hybridation respectées, le clivage des brins se produit à trois nucléotides en amont de la séquence PAM (127). Les multiples ARNg bactériens, couplés aux différents motifs PAM des endonucléases, sont méticuleusement établis conjointement pour n'avoir qu'une cible et

épargner l'ADN du « soi ». Le fonctionnement du système de défense, la structure de l'ARNg et le mécanisme de coupure sont tous illustrés dans la Figure 5.



**Figure 5. – Vue d'ensemble et fonctionnement du système CRISPR/Cas9**

**A.** Fonctionnement du système de défense procaryote CRISPR/Cas9. L'opéron Cas et le locus CRISPR encode l'information génétique des endonucléases Cas et des fragments ARNg. Après expression et traitement, leur complexation permet le ciblage et l'inactivation spécifique d'ADN viral envahissant **B.** Structure schématisée d'un ARNg typique. L'ARNcr d'une vingtaine de nucléotides sert au ciblage de l'ADN exogène. L'ARNtracr adopte une structure tridimensionnelle pour l'ancrage de l'ARNg dans la protéine Cas. **C.** Représentation schématique du mécanisme de coupure double brin. L'appariement entre l'hybride ADN-ARNg (Proto-espaceur – ARNcr) nécessite aussi la présence du motif PAM sur l'ADN cible. La DSB se produit trois nucléotides en amont de la séquence PAM. Figure adaptée à partir de Doudna et Charpentier, et la documentation produite par le laboratoire pionnier Doudna (126, 128).

### 1.8.1 Adaptation en tant qu'outil d'édition génomique

Tenant pour acquises toutes les subtilités de ce mécanisme de défense bactérien, ce dernier peut être détourné pour en faire un puissant outil d'édition génomique. Ainsi, un expérimentateur peut confectionner des ARNg synthétiques en éditant soigneusement la portion ARNcr pour le ciblage d'une séquence génique spécifique (126). Advenant une bonne conception de l'ARNg, et en délivrant tout le matériel avec succès dans un modèle d'étude d'intérêt; il est alors théoriquement possible de cibler tout gène, dans tout organisme, afin d'en provoquer l'altération (126). L'induction d'une DSB dans un gène cible sera immédiatement réparée, et ce de multiples façons. Le tout sera abordé plus en profondeur dans la prochaine sous-section 1.8.2.

La cargaison, ou le matériel à transférer, peut être délivrée sous diverses formes : soit un complexe ribonucléoprotéique Cas9-ARNg; soit ribonucléique avec un ARNm pour la Cas9 et un ARNg distinct; ou bien sous forme plasmidique (129). L'avantage de transférer directement le complexe ribonucléoprotéique Cas9/ARNg transcrit et traduit est qu'il sera prêt à agir aussitôt intégré dans le sujet cible. Cependant cette stratégie nécessite un important niveau de pureté de la protéine Cas9 hautement active, en plus d'avoir un poids moléculaire relativement élevé (130). En optant pour la voie ARN, il faut compter une étape de traduction préalable de la Cas9 dans l'organisme avant d'être effective. Le plus gros désavantage de cette forme est l'instabilité moléculaire du matériel ribonucléique (130). La formule plasmidique facilite « l'emballage » de la cargaison en une seule unité moléculaire, mais nécessite des étapes de transcription et de traduction additionnelles dans le sujet hôte. Un avantage des plasmides est qu'il arbore habituellement un rapporteur fluorescent, tel que la protéine fluorescente verte (GFP), qui sert de critère de sélection pour indiquer une transfection réussie du matériel (131). La nature moléculaire de la cargaison est dictée en fonction du mode de transfection choisi, qui lui est dépendant du sujet cible (par exemple : *in vivo* vs *in vitro*).

Les modes de livraison du matériel peuvent être groupés en trois grandes catégories : par voie physique, par vecteur viral ou par vecteur non viral (129). Les méthodes physiques comprennent des procédures telles que la micro-injection ou l'électroporation. Ces approches sont exclusivement employées avec embryons et cellules, et permettent une transfection du matériel sous toutes les formes moléculaires abordées (132, 133). L'utilisation de vecteurs viraux inclut

des virus adéno-associés (AAV), des adénovirus et des lentivirus. En combinant les trois, ils permettent l'infection d'un large éventail de cellules. Ils sont sollicités dans les contextes *in vivo* et *in vitro*. Les AVV et les adénovirus permettent l'empaquetage de matériel ADN, tandis que les lentivirus véhiculent de l'ARN (134). Finalement, les vecteurs non viraux englobent une multitude de procédés sollicitant liposomes, peptides pénétrant les cellules, nanoparticules lipidiques, nanoparticules inorganiques et autres polymères divers (130). Ils ne sont pas aussi bien implantés que les voies physiques ou les vecteurs viraux, mais il constitue un champ de recherche en pleine expansion. Leur principal avantage est la versatilité d'utilisation qu'ils offrent. Combinant tous les types de vecteurs non viraux, ils peuvent transporter le matériel CRISPR/Cas9 sous toutes ses formes moléculaires, en plus de permettre un certain pouvoir de conception chimique de leur surface, offrant des spécificités de ciblage cellulaire (130). Ils peuvent tout aussi bien être employés dans un cadre *in vivo*, qu'*in vitro*.

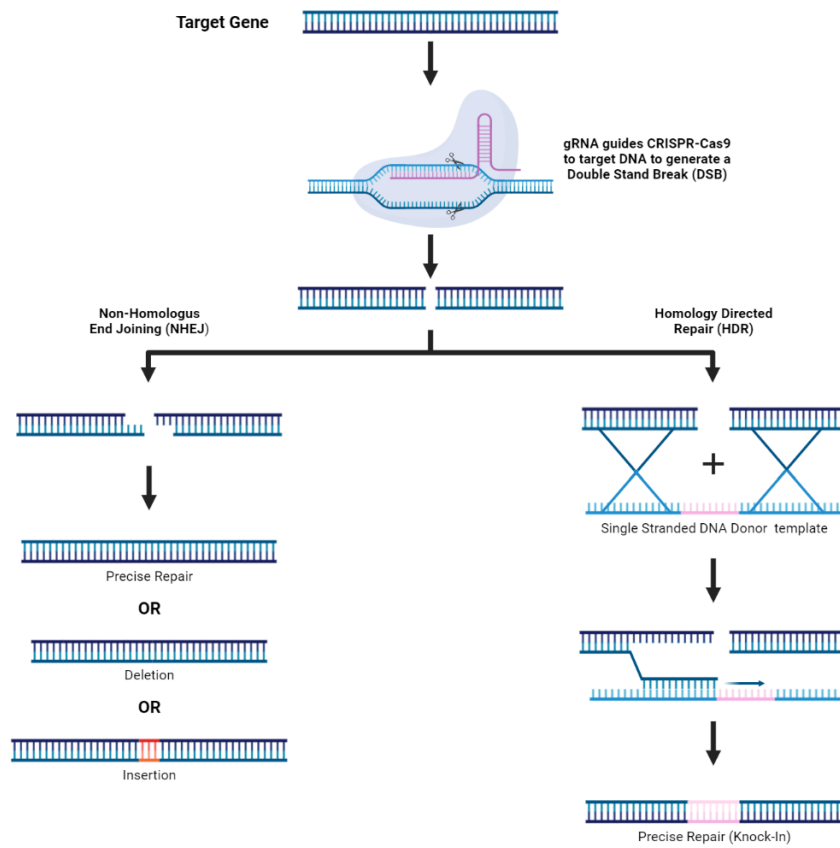
### **1.8.2 Réparation cassure double brin : Réparation NHEJ vs HDR**

Supposant une bonne conception de l'expérience et une transfection du matériel réussie, alors une DSB est introduite dans le gène cible du sujet. Cette rupture de l'ADN étant létale pour la cellule, celle-ci la raccommode suivant deux mécanismes de réparation se compétitionnant : Jonctions d'extrémités non homologues (NHEJ) et Réparation par recombinaison homologue (HDR) (126, 135). Étant la plus rapide à se déployer, la voie NHEJ est la plus prévalente, effectuant la religature des extrémités en un peu moins de 30 minutes, alors qu'elle prend plusieurs heures pour son homologue (136). Elle est cependant plus encline à incorporer des indels dans la séquence génomique originale. Tout dépendamment de la position et la gravité des défauts incorporés, cela peut mener à une répression partielle ou complète de l'expression du gène, prenant les appellations respectives de *Knockdown* ou *Knockout* (KO) (126, 137). Plus un gène est altéré tôt dans sa séquence génomique, plus les chances de générer un KO sont élevées. Elle peut exercer une édition précise de la DSB, mais il s'agit là d'un événement extrêmement rare (138).

Le deuxième phénomène de réparation, plus lent et plus rare, sollicite un brin d'ADN « sœur » comme matrice homologue pour une reconstruction précise de la cassure. Le brin d'ADN dit « donneur » doit posséder une région d'homologie avoisinant le site de la DSB afin d'être amarré à



proximité, et être sollicité comme gabarit durant le processus de réparation. Naturellement, la cellule va utiliser une chromatide sœur afin de régénérer la séquence originale fidèlement (139). Cependant, en confectionnant adéquatement un ADNss donneur avec des bras d'homologie et une séquence précise en son centre, il est possible d'incorporer une mutation candidate d'intérêt *Knock-In* (KI) dans le génome du modèle concerné (140). Cela permet ainsi de tester les impacts fonctionnels de ladite mutation dans un modèle d'intérêt. Habituellement, les ADNss donneurs employés font une longueur de 200 nucléotides, nécessitant des bras d'homologie d'une longueur minimale de 20 nucléotides à chacune des extrémités (141). Sachant qu'avant de religaturer la DSB, la cellule taillade les extrémités de la brisure pour réparer sur un tronçon plus long, cela laisse donc place à une plage de réparation de 160 nucléotides (141, 142). En théorie, il serait possible que le ciblage de l'ARNg soit distant du site d'édition d'intérêt de quelques dizaines de nucléotides. Par contre en pratique, l'idéal est de centrer tout le processus de réparation au maximum possible, avec des bras d'homologie faisant une cinquantaine de nucléotides de long (140). La Figure 6 présente schématiquement les deux voies de réparations se compétitionnant dans la cellule. Un point important à garder à l'esprit lors de la mise en place d'une expérience de KI : il faut s'assurer de supprimer le motif PAM génomique en même temps d'incorporer la mutation d'intérêt. Autrement, des restants du complexe ribonucléoprotéique pourraient recliver dans le génome indéfiniment. Une des stratégies est donc d'introduire une mutation silencieuse dans l'ADNss donneur, vis-à-vis du motif PAM génomique. Ce point sera abordé de nouveau dans la section 3.2.2 dans le processus de conception de l'expérience.



**Figure 6. – Compétition des voies de réparation : NHEJ vs HDR**

Suivant l'induction d'une DSB dans un gène cible, celle-ci peut soit être réparée par NHEJ ou HDR. NHEJ : mécanisme le plus prévalent, mais aussi le plus enclin à incorporer des indels, résultant dans la majorité des cas avec un KO du gène. HDR : phénomène plus lent et plus rare qui va réparer fidèlement la DSB en utilisant une matrice ADN homologue. Cette réparation résulte avec un KI précis dans le gène cible. Figure adaptée à partir de Lino *et al.* (129).

### 1.8.3 Principe de digestion d'hétéroduplexe

Une des façons de vérifier l'efficacité du complexe ribonucléoprotéique, quant à son aptitude à induire une DSB dans le génome ciblé, est d'effectuer un test de digestion d'hétéroduplexe (143). Comme mentionné dans la sous-section précédente, suite à une DSB produite par CRISPR/Cas9, la cellule est portée à la réparer le plus rapidement possible par NHEJ, incorporant des indels par le fait même. On peut donc tirer profit de ces défauts pour générer un hétéroduplexe, une région nucléaire double brin caractérisée par un mésappariement des paires de bases, plus ou moins

hybridées adéquatement (143). En amplifiant la région d'intérêt par PCR, puis en dénaturant et appliquant un traitement thermique refroidissant très spécifique aux amplicons, il est possible de régénérer l'hétéroduplexe déshybridé (143). Celui-ci se manifeste comme une protubérance sur les amplicons, pouvant être décelé et digéré enzymatiquement par l'Endonucléase 1 T7 (T7E1) (143). Lorsque révélé sur gel d'agarose, le résultat présentera une fragmentation de l'amplicon original, indiquant la bonne conception et le bon fonctionnement moléculaire du complexe Cas9-ARNg.

#### **1.8.4 Culture unicellulaire – Génération de clones**

Suite à la transfection du système d'édition génomique, les cellules forment une population hétérogène vis-à-vis de leur génotype. S'ensuivra une culture unicellulaire dans des plaques 96 puits, le but étant de générer des clones génétiquement identiques à partir d'une seule cellule mère initiale. Les récoltes et filtrations préalables de sécrétions cellulaires de myoblastes sont déterminantes lors de la prolifération à l'état unitaire. Supplémenter ces sécrétions dans le milieu de culture va stimuler la croissance des myoblastes individuels, dépourvus des différents stimuli signalétiques de leurs voisins habituels (144).

### **1.9 Hypothèse et objectifs**

Le variant identifié dans le gène *MLIP* constitue un candidat de choix pour investigation approfondie. Plusieurs points à son endroit laissent présager de son implication dans la myopathie distale du patient. Tout d'abord la nature du variant identifié, étant une mutation non-sens elle va causer une troncature prématurée avec des domaines protéiques absents. Ici, la mutation à la fin de l'exon 5 retirerait le crochet AT présent sur les exons 9 et 10 de *MLIP*. Également, dans la littérature actuelle, les transcrits contenant l'exon 4 sont également inclusifs pour l'exon 5, et avec la séquence NLS présente sur l'exon 4, la localisation d'isoformes nucléaires risque d'être affectée s'ils sont pleinement exprimés. L'expression quasi exclusive de *MLIP* au niveau du tissu musculaire est un point qui vient rejoindre la manifestation clinique du patient. Par ailleurs, une rapide analyse des données d'expression différentielle provenant du RNA-seq révèle un déséquilibre des niveaux d'expression, avec une baisse évidente de *MLIP* et par phénomène compensatoire, une hausse de celui de *LMNA*. Avec tout cela, il y a également le fait

que *LMNA* a été relié à des laminopathies lorsque dérégulé, dont des cardiomyopathies et dystrophies musculaires. Étant le partenaire protéique principal de notre candidat, une altération de *MLIP* est également suspectée de causer des pathologies à phénotypes musculaires.

Pour toutes ces raisons, nous avons donc émis l'hypothèse que le dysfonctionnement ou l'absence de *MLIP* cause une détérioration des fonctions musculaires et mène à une myopathie. Le but de ce projet est donc de valider la relation génotype-phénotype : prouvant la pathogénicité du variant *MLIP* identifié, au travers d'évaluation de données -omiques plus poussée et en sollicitant le LRS pour une meilleure caractérisation transcriptomique de *MLIP*. En parallèle, l'établissement d'un modèle cellulaire de myoblastes humains immortalisés, muté par l'entremise de la technologie CRISPR/Cas9, permettra une évaluation plus importante des différents impacts morpho-fonctionnels engendrés par le variant *MLIP*.

## **2 – Méthodologie**

### **2.1 Évaluation clinique et obtention des échantillons**

L'évaluation clinique du patient et le prélèvement de la biopsie musculaire furent réalisés par une neurologue de L'Institut-hôpital neurologique de Montréal (Le Neuro), la Dre Erin O'Ferrall. L'analyse histologique subséquente fut accomplie par un neuropathologiste du même institut, le Dr Jason Karamchandani. Le morceau de biopsie musculaire destiné à l'extraction des macromolécules d'intérêt a été transporté jusqu'au Centre de recherche du centre hospitalier de l'Université de Montréal (CRCHUM) sur glace sèche, et entreposé à -80.0°C jusqu'à son utilisation expérimentale. Le comité d'éthique de la recherche du Centre universitaire de santé McGill a approuvé l'étude. L'individu a signé un consentement éclairé autorisant l'analyse génétique dans un contexte de recherche.

#### **2.1.1 Extraction des macromolécules de la biopsie musculaire**

L'extraction séquentielle de l'ARN, l'ADN et les protéines a été réalisée en employant le réactif TRIzol (Invitrogen, 15596026), en prenant le soin de décontaminer la paillasse avec le RNaseZAP (Invitrogen, AM9782) au préalable. Une évaluation rapide par dosage spectrophotométrique est effectuée pour observer les qualités et quantités des macromolécules isolées avant leur entreposage respectif à -80.0°C, 4.0°C et -20.0°C.

**Tableau 2. – Séquences des amorces utilisées pour LRS, PCR et RT-PCR**

<b>Amorces LRS</b>
MLIP_CB1_BP07_F : 5' – CCGTGAC GTGTTACCGTGGGAATGAATCCTT TGGACTCCGAAGGGGAAGAT – 3'
MLIP_CB1_BP07_R : 5' – CCGTGAC GTGTTACCGTGGGAATGAATCCTT GGGCGAATTACTCCAGGCTT – 3'
MLIP_CB2_BP13_F : 5' – CCGTGAC AGAACGACTTCCATACTCGTGTGA TGGACTCCGAAGGGGAAGAT – 3'
MLIP_CB2_BP13_R : 5' – CCGTGAC AGAACGACTTCCATACTCGTGTGA GGGCGAATTACTCCAGGCTT – 3'
Conception : Portion tampon – Codes-Barres – Séquence d'hybridation
<b>Amorces PCR</b>
Amorce sens GCDF : 5' – GTGGCATTCTTCGCTTCTC – 3'
Amorce antisens GCDR : 5' – CCCAAAGTAGCTTGACAGTGG – 3'
GCD : <i>Genomic Cleavage Detection Forward/Reverse</i>
<b>Amorces RT-PCR</b>
Amorce sens : 5' – GGGAATTCGAAGCAAACAAA – 3'
Amorce antisens : 5' – GGGGACCTTGAAGGAGAATC – 3'

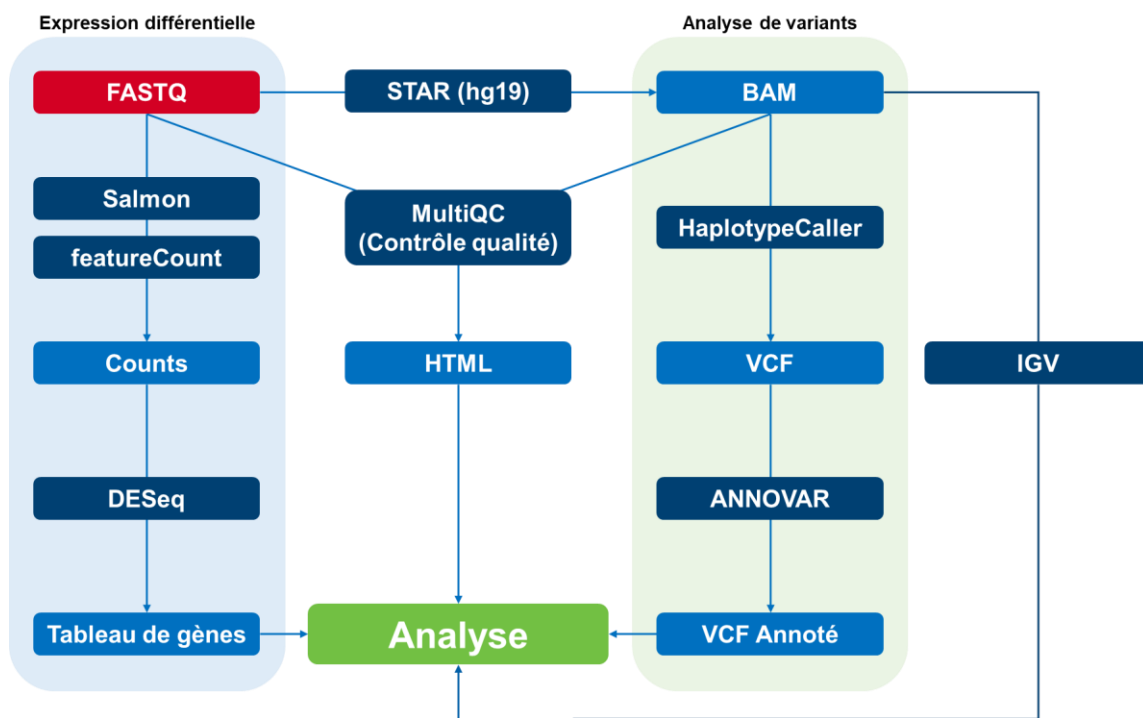
## 2.2 Techniques de séquençage employées

Trois approches de séquençage distinctes furent employées dans la réalisation de ce projet : le RNA-seq pour l'établissement du diagnostic génétique du patient; le LRS pour une caractérisation transcriptomique approfondie de *MLIP*; et le séquençage Sanger à des fins de validations des séquences d'amplicons PCR *MLIP*.

### 2.2.1 Séquençage transcriptomique (RNA-seq)

Avant la manipulation du matériel, la paillasse est décontaminée à l'aide du RNaseZAP (Invitrogen, AM9782). L'ARN total musculaire fut envoyé au Centre d'expertise et de services Génome Québec (CESGQ) pour séquençage transcriptomique en suivant les directives imposées, soit l'envoi d'une quantité minimale de 500ng d'ARN dans un volume minimal de 15µL. Un échantillon fortement concentré est dilué dans de l'eau certifiée dépourvue d'ARNase (Invitrogen, 10977015). L'échantillon est envoyé sur glace sèche, et un contrôle de qualité est

effectué par bioanalyseur (Agilent 2100 Bioanalyzer System) suivant sa réception. Un score RIN (valeur d'intégrité de l'ARN) supérieur à 6.5 est exigé pour le traitement ultérieur de l'échantillon : le présent échantillon possédant un score de 8.8. La librairie de séquençage est préparée à l'aide du kit *TrueSeq Stranded mRNA Library Preparation Kit* (Illumina, 20020595) et est ensuite séquencée sur un appareil Illumina HiSeq 2500 en utilisant des lectures appariées de 125pb. Les données brutes générées (fichiers FASTQ) sont ensuite traitées au travers d'un pipeline informatique, permettant ultimement leur analyse et interprétation. Le pipeline a été mis en place dans l'enceinte du laboratoire et apparaît à la Figure 7. Il comporte deux branches distinctes : une succession d'outils servant à la génération des données d'expression différentielle et l'autre pour l'identification de variants.



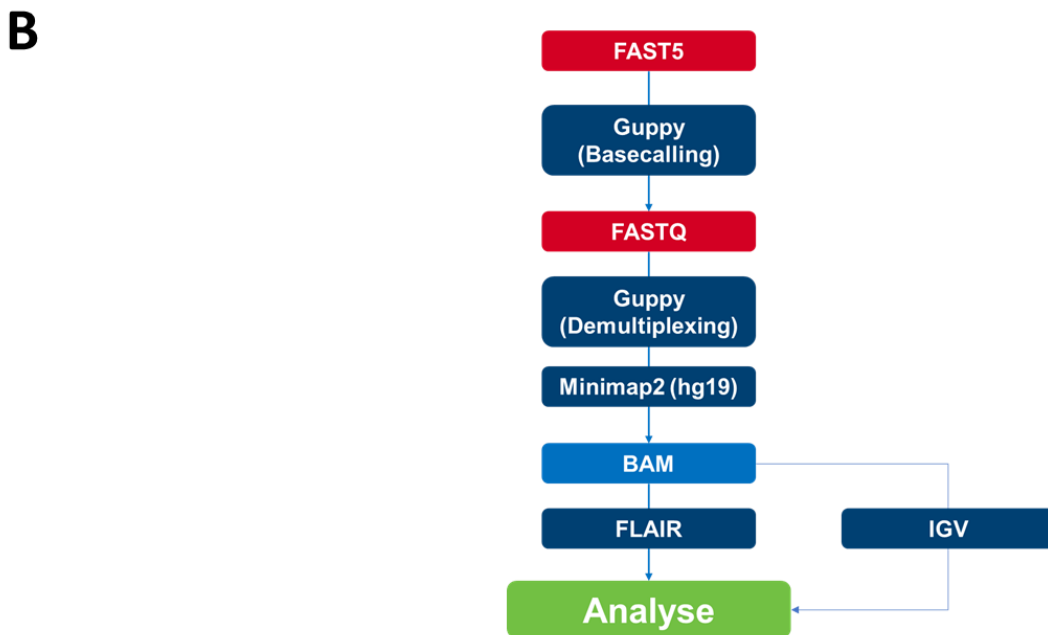
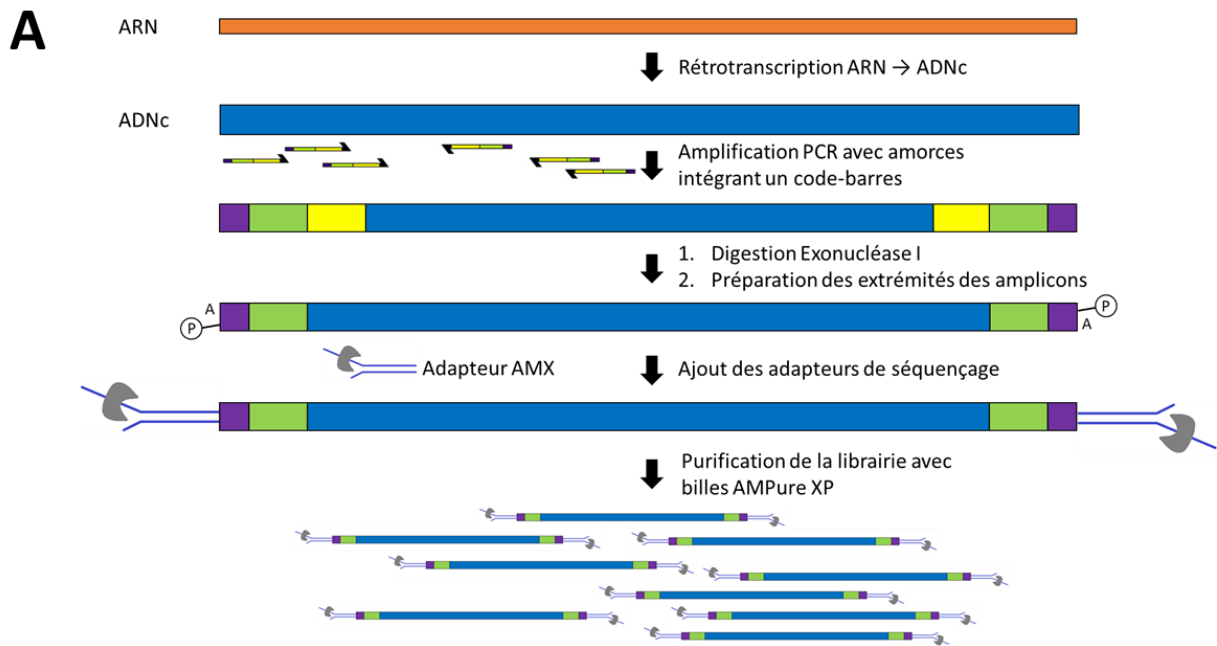
**Figure 7. – Schéma du pipeline bio-informatique RNA-seq**

Représentation schématique de la succession d'outils pour le traitement des données brutes du RNA-seq. D'un côté, une branche sert au traitement des données d'expression différentielle permettant l'interprétation des impacts fonctionnels. De l'autre, les outils servant à l'identification de variants pathogéniques. L'outil IGV est un logiciel indépendant permettant à l'utilisateur d'évaluer et visualiser manuellement les lectures des fichiers BAM.

### 2.2.2 Séquençage de longues lectures (LRS)

La préparation de la librairie de séquençage débute avec une rétrotranscription de 50ng d'ARN total musculaire grâce au kit *SuperScript VILO cDNA Synthesis Kit* (Invitrogen, 11754050), suivi d'une amplification PCR de 30 cycles avec la *LongAmp Taq DNA Polymerase* (NEB, M0323S). L'amplification emploie les couples d'amorces, MLIP\_CB1\_BP07\_F/R et MLIP\_CB2\_BP13\_F/R (Tableau 2), intégrant un code-barres permettant le multiplexage des amplicons patient-contrôle dans un même échantillon. Les séquences des amorces sont TGGACTCCGAAGGGGAAGAT et GGGCGAATTACTCCAGGCTT ciblant respectivement l'exon 3 et 11 de *MLIP*. Les préparations sont ensuite traitées avec l'Exonucléase I (NEB, M0293S) afin de dégrader et nettoyer de tout ADN simple brin libre, notamment des amorces non utilisées. Les extrémités des amplicons sont par la suite modifiées et ajustées avec l'enzyme *NEBNext Ultra II End Repair/dA-Tailing Module* (NEB, E7546S) afin de permettre l'attachement des adaptateurs de séquençage AMX (ONT, SQK-LSK109) par la ligase *NEBNext Quick Ligation Module* (NEB, E6056S). Une purification est effectuée avec les billes *AMPure XP Beads* (Beckman Coulter Life Sciences, A63881), suivi d'un dosage fluorimétrique grâce au QuBit4, permettant le multiplexage contrôle-patient suivant les concentrations spécifiques demandées, et ce dans un volume final de 12µL. La *flowcell* (ONT, FLO-MIN106) servant de chambre de séquençage est insérée dans un appareil de séquençage MinION et est préparée grâce au kit *Flow Cell Priming Kit* (ONT, EXP-FLP002) en suivant les directives du fabricant. La librairie est mélangée à 37.5µL de SQB et 25.5µL de LB (*Sequencing Buffer, Loading Beads*, ONT, SQK-LSK109) juste avant d'être chargée sur la *flowcell*. Le séquençage a été lancé pour une période de 24 heures. Les données brutes générées (fichiers FAST5) ont finalement été traitées au travers d'un pipeline bio-informatique, permettant ultimement leur analyse et interprétation. Le pipeline a été mis en place dans l'enceinte du laboratoire et apparaît dans la Figure 8 au côté de la préparation de la librairie LRS. Après l'appel de base, le démultiplexage patient-contrôle et l'alignement des échantillons par rapport au génome référence hg19, l'outil FLAIR permet une quantification des isoformes de *MLIP* pour l'évaluation des changements transcriptomiques.





**Figure 8. – Préparation de la librairie et pipeline bio-informatique LRS**

**A.** Illustration des étapes clés lors de la préparation de la librairie de séquençage. La librairie contrôle et la librairie du patient sont multiplexées en suivant les directives protocolaires du manufacturier. **B.** Schématisation du pipeline bio-informatique et la succession d’outils utilisés pour le traitement des données brutes du LRS. L’outil IGV est un logiciel indépendant permettant à l’utilisateur d’évaluer et visualiser manuellement les lectures des fichiers BAM.

### 2.2.3 Séquençage Sanger

Les échantillons nécessitant une validation par séquençage Sanger sont envoyés au CESGQ, en suivant les critères ci-après : échantillon PCR non purifié, quantité minimale se situant entre 5 et 25ng, dans un volume minimal de 20µL. Plus un amplicon PCR est long, plus sa concentration devrait être élevée pour un résultat de séquençage satisfaisant. Chaque échantillon est envoyé en duplicata, pour une lecture sens et antisens, couplé avec ses amorces respectives qui doivent être concentrées à 5µM dans un volume minimal de 10µL. Un échantillon fortement concentré est dilué dans de l'eau certifiée dépourvue de nucléase (Eau DEPC, Invitrogen, AM9906). Les résultats reviennent sous forme de chromatogrammes et leur analyse s'effectue grâce au logiciel GeneStudio.

### 2.3 Immunobuvardage (WB)

30µg de protéines musculaires ont été chargés dans un gel polyacrylamide à gradient 4-15% (Bio-Rad, 4561084). S'ensuit un transfert sur une membrane PVDF (Bio-Rad, 1620177), puis un blocage dans une solution PBS avec 5% de lait pendant 30 minutes. Les anticorps et dilutions utilisés vont comme suit : MLIP, lapin, anti-humain, polyclonal, 1 :2000 (Invitrogen PA5-72759); LMNA, lapin, anti-humain, polyclonal, 1 :1000 (NEB, 2026S); GAPDH, lapin, anti-humain, monoclonal, 1 :1000 (NEB, 5174S); anticorps conjugué à la peroxydase de raifort (HRP), chèvre, anti-lapin, polyclonal, 1 :10000 (Jackson ImmunoResearch, 111-035-114). Les anticorps primaires sont incubés pendant la nuit à 4.0°C dans des dilutions à base de PBS contenant 1% de BSA et 0.1% d'azoture de sodium. L'anticorps secondaire HRP est ensuite incubé dans les mêmes conditions, mais pour une période d'une heure. La révélation du signal se fait avec le substrat ECL (Invitrogen, 32106) et sa capture chemiluminescente est réalisée par l'appareil ChemiDoc System. Les résultats obtenus sont analysés à l'aide du logiciel Image Lab 6.0.1 de Bio-Rad.

### 2.4 Réaction en chaîne par polymérase (PCR)

Les réactions de PCR sont réalisées avec la polymérase *EasyTaq* (TransGen Biotech, AP111-01) en suivant les directives du manufacturier. Le mélange réactionnel de 50µL comporte 5µL d'*EasyTaq Buffer* 10X, 4µL de dNTPs 2.5mM, 0.6µL d'*EasyTaq DNA Polymerase*, une moyenne de 50ng d'ADN

purifié, 1µL pour chacune des amorces GCDF/GCDR (Tableau 2) à partir d'une solution de 10µM, et enfin complété avec de l'eau dépourvue de nucléases. Les cycles thermiques sont assurés par l'appareil 2720 Thermal Cycler d'Applied Biosystems. La température d'hybridation du couple d'amorces GCDF/GCDR est de 62.0°C, suivant la tendance : Dénaturation 94.0°C, 5min → 35x (94.0°C, 30sec → 62.0°C, 30sec → 72.0°C, 30sec) → Élongation finale 72.0°C, 10min et une préservation finale à 4°C. Les amplicons sont ensuite chargés et migrés sur gel d'agarose 1.0% pour la visualisation des fragments amplifiés.

### 2.4.1 Électrophorèse sur gel d'agarose

Un gel d'agarose 1.0% est composé de 0.250g de poudre d'agarose ultrapure (Invitrogen, 16500500) diluée et chauffée par intervalles au four à micro-ondes dans 25.0mL TBE 1X (Invitrogen, 15581044). L'agent révélateur, le *GelRed* 10 000X (Biotium, 41003), est ajouté à hauteur de 2.5µL dans la solution chaude avant polymérisation du gel. Les échantillons à charger sont composés de quelques centaines ng d'ADN, en général 1 à 3µL de matériel nucléique, 1µL de *BlueJuice* 10X (Invitrogen, 10816015) et un volume complémentaire d'eau DEPC (Invitrogen, AM9906) de sorte à atteindre 10µL au total. L'échelle moléculaire est composée de 1µL de *100bp DNA Ladder* (Invitrogen, 15628019), 1µL de *BlueJuice* 10X et 8 µL d'eau DEPC. La migration des échantillons est assurée par l'appareil *EC250-90* d'E-C Apparatus Corporation. Suite à l'électrophorèse, le gel est révélé sous rayons UV et le signal est capté par l'appareil ChemiDoc MP Imaging System de Bio-Rad. L'analyse et traitement des résultats s'effectuent à l'aide du logiciel Image Lab 6.0.1 de Bio-Rad.

### 2.5 PCR quantitative (qPCR)

Suivant la rétrotranscription d'ARNm musculaire par *SuperScript VILO cDNA Synthesis Kit* (Invitrogen, 11754050), quatre sondes TaqMan (Invitrogen) furent utilisées : Hs00951765\_m1 et Hs00370866\_m1 ciblant respectivement les jonctions d'exons 5-6 et 9-10 de *MLIP*. La sonde Hs00153462\_m1 se lie à la jonction 2-3 de *LMNA* et finalement Hs03004310\_g1 est spécifique à la jonction 1-2 du gène de référence *RPS29*, utilisé pour la normalisation des données. Les niveaux d'expressions sont évalués à l'aide des  $\Delta\Delta CT$  capturés et quantifiés par l'appareil QuantStudio 6/7 System.

## 2.6 PCR précédée d'une transcription inverse (RT-PCR)

L'ADNc a été obtenu par transcription inverse de l'ARN à l'aide du kit *SuperScript VILO cDNA Synthesis Kit* (Invitrogen, 11754050) et puis amplifié par PCR grâce à la polymérase *EasyTaq* (TransGen Biotech, AP111-01) avec la paire d'amorces : GGGAATTCGAAGCAAACAAA et GGGGACCTTGAAGGAGAATC. 300ng d'échantillons et 500ng de *1Kb Plus E-gel DNA Ladder* (Invitrogen, 10787018) furent transférés et migrés dans un E-gel 2% (Invitrogen, G401002).

## 2.7 Culture et entretien de la lignée cellulaire

La lignée cellulaire de myoblastes humains immortalisés provient du laboratoire du Dr. Eric Shoubridge du Neuro. Le milieu de culture employé est composé de 500.0mL de Sk-Max (Wisent, Cat. 301-061-CL), 10.0mL de supplément (Wisent, Cat. 301-061-XL), 100.0mL de sérum foetal bovin (FBS, Wisent, Cat. 080-150) et 5.0mL de pénicilline-streptomycine (Wisent, 450-200-EL). Les cellules sont cultivées dans des boîtes de pétri adhérentes de 100mm (Sarstedt, 83.3902) et à des conditions d'incubations de 37.0°C et 5.0% CO<sub>2</sub>, maintenues par l'incubateur CellXpert 170i de la compagnie Eppendorf. Des observations microscopiques sont effectuées au quotidien pour assurer le suivi de leur état. À une confluence cellulaire avoisinant les 70 à 90%, les cellules sont alors transférées dans un nouveau contenant de culture stérile afin de permettre leur croissance continue. L'ancien milieu de culture est alors retiré, les cellules sont lavées avec 4.0mL de tampon phosphate salin de Dulbecco (D-PBS, Wisent, 311-430-CL), puis détachées avec 2.0mL de trypsine 0.05% EDTA (Wisent, Cat. 325-042-CL) en les incubant à 37.0°C et 5.0% CO<sub>2</sub> pendant 3 à 4 minutes. En ajoutant 8.0mL de milieu de culture neuf aux cellules nouvellement détachées, le contenant est rincé avec la solution même avant de la transférer dans un contenant stérile neuf, suivant une dilution de 1/5. Le rythme de passage est établi tous les 2 ou 3 jours. Après 48 heures sans passage, le milieu de culture est remplacé avec du neuf afin d'éviter une acidification trop prononcée de celui-ci. Les cellules sont entretenues de la sorte jusqu'à leur utilisation expérimentale.

### **2.7.1 Électroporation des myoblastes**

La transfection du matériel pour l'édition génomique des myoblastes s'effectue au moyen de l'électroporation et est réalisée grâce à la machine Neon d'Invitrogen. Les paramètres d'électroporation modulables sont le voltage d'une pulsation (V), la durée d'une pulsation (ms) et le nombre de pulsations; ceux employés avec la lignée cellulaire sont de 1650V, 10ms et 3 pulsations. L'électroporation s'effectue en suivant les directives du protocole avec les quantités de cellules et de réactifs suivantes : 100 000 cellules, 2 $\mu$ g de Cas9 (Invitrogen, A36498), 400ng d'ARN guide et 0.5 $\mu$ L de la solution 100 $\mu$ M pour la matrice ADNss (IDT). Le tout est compris dans un volume totalisant 10 $\mu$ L, complété et suspendu dans du Tampon R, tampon du kit qui est utilisé avec les lignées cellulaires immortalisées (voir exception dans le protocole du kit).

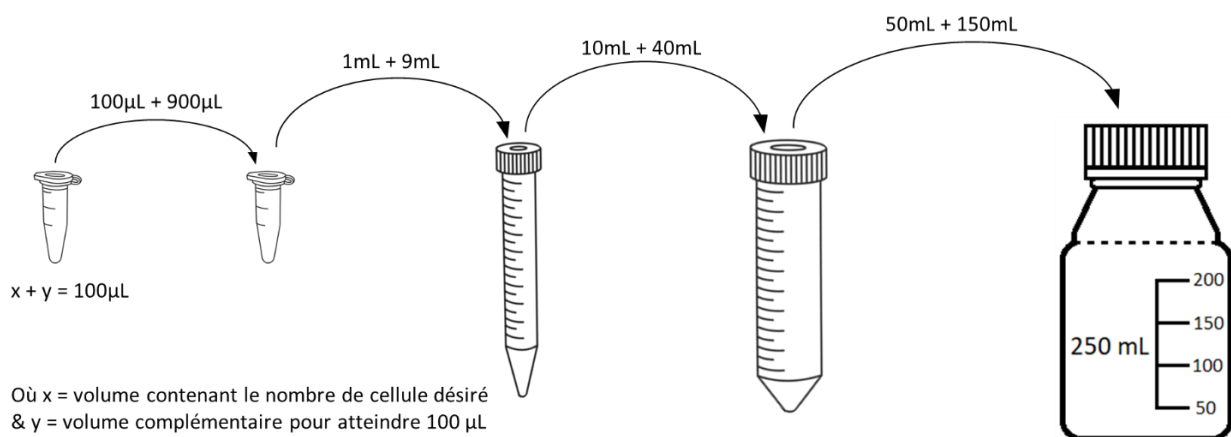
#### **2.7.1.1 Capture d'imagerie en microscopie fluorescente**

La vérification de l'efficacité d'électroporation s'effectue en transfectant 100 000 cellules avec 500ng de plasmide GFP pmEGFP-C1 (Addgene, 36312) et en examinant qualitativement les taux de survivabilité et de transfection. Les captures sont prises 4 jours après l'expérience d'électroporation, permettant aux cellules de pleinement récupérer et exprimer la protéine GFP. Les images ont été captées à l'aide du microscope Olympus CKX41, la source de lumière à fluorescence Olympus U-RFLT50 et l'appareil de capture Olympus DP20. La capture est réalisée en mode automatique avec un objectif 20X : un premier temps à la longueur d'onde d'excitation de la GFP soit 488nm, et une seconde fois sans aucun filtre d'éclairage. Finalement, une vérification préalable avec des cellules non-électroporées a été faite, afin d'attester que les myoblastes n'étaient pas auto-fluorescents.

### **2.7.2 Génération de clones par culture unicellulaire**

Trois à quatre jours suivant l'électroporation, une vérification préalable de l'état des cellules est effectuée avant de procéder avec la culture unicellulaire. Du milieu de culture frais est préparé tout juste avant l'expérience et supplémenté avec des sécrétions cellulaires filtrées de myoblastes. Les proportions de sécrétions devraient occuper 1/8 du volume final. Pour 20 plaques adhérentes 96 puits (Sarstedt, 83.3924), 200mL de milieu de culture (175mL frais/25mL sécrétions filtrées) sont préparés, comprenant 1600 myoblastes de sorte à avoir une

concentration finale de 8 cellules/mL. La population cellulaire est préparée par dilution séquentielle, suivant la méthode de la dilution limite illustrée dans la Figure 9. S'ensuit le remplissage des plaques avec la distribution de 100µL/puit à l'aide d'une pipette multicanal. Les plaques sont ensuite déposées à l'incubateur pour une période de deux semaines, laissant le temps aux premières colonies de se développer. Des vérifications quotidiennes sont effectuées, et à l'atteinte d'une confluence de 70-80%, les colonies sont subséquemment transférées dans des plaques de 48, 24, 12, 6 puits (Sarstedt, 83.3923, 83.3922, 83.3921, 83.3920) afin de permettre la croissance continue de la population. Après les plaques 48 puits, les colonies sont divisées dans deux puits distincts de plaque 24 puits, un servant à l'extraction d'ADN et la vérification du génotype, et l'autre pour le maintien de la croissance cellulaire.



**Figure 9. – Préparation de la population cellulaire par dilutions séquentielles**

Représentation du principe de dilution séquentielle ou dilution limite. La préparation du mélange de la sorte est déterminante pour la dispersion optimale de la faible population cellulaire dans le large volume de milieu. **Note** : Le 25mL de sécrétions cellulaires est ajouté au volume final de 150mL (125mL frais + 25mL sécrétions filtrées).

### **2.7.3 Extraction de l'ADN des clones**

L'extraction de l'ADN des clones est accomplie à l'aide du kit *PureLink Genomic DNA Kit* (Invitrogen, K182002), en suivant les procédures du protocole. L'isolation se base sur une purification par colonne; après les étapes de lyse cellulaire, de centrifugations et lavages de colonne, l'élution s'effectue avec 50µL de *PureLink Genomic Elution Buffer* en centrifugeant à température pièce (12 700RPM, 1 minute). L'éluant est retransféré sur sa colonne respective et recentrifugé avec les mêmes paramètres à des fins de rendements supérieurs. Un rapide dosage spectrophotométrique est effectué pour évaluer les qualités et quantités d'échantillons ADN extraits avant leur entreposage à 4.0°C ou -20.0°C (long terme).

## **2.8 Conception de l'édition génomique avec CRISPR/Cas9**

Trois expériences distinctes d'éditions génomiques CRISPR/Cas9 furent établies à l'aide de la plateforme en ligne Benchling. Elle permet à la fois la confection d'ARNg et de matrice ADNss pour la production d'un KI. Lors de la sélection des ARNg, des scores de *on-target* et de *off-target* supérieurs à 60 et 50 respectivement sont préconisés pour leur fidélité de ciblage. L'ADNss donneur proposé par la plateforme porte la mutation d'intérêt et supprime par le fait même la séquence PAM par mutation silencieuse. Les séquences des matrices ADNss sont ensuite traitées au travers de l'outil en ligne Silent Mutator, permettant l'introduction d'un site de restriction par mutation silencieuse. Ce site de restriction permet ainsi la vérification du KI au travers d'un test de digestion avec enzyme de restriction. Malgré une plage de réparation théorique de plusieurs dizaines de pb, il faut préconiser une conception CRISPR/Cas9 la plus centrée possible durant la confection de l'expérience. Toutes les séquences nucléiques du matériel utilisé apparaissent dans la Tableau 3.

**Tableau 3. – Séquences du matériel nucléique utilisé pour la conception des expériences d'édition génomique CRISPR/Cas9**

---

**Couples d'amorces servant à l'assemblage de la matrice ADN des ARN guides**

---

Amorce sens ARNg 1: 5' – TAATACGACTCACTATAGTTGGAACAGAAGGTCA – 3'

Amorce antisens ARNg 1: 5' – TTCTAGCTCTAAAACACACTGACCTTCTGTTCCA – 3'

Amorce sens ARNg 2: 5' – TAATACGACTCACTATAGCACATTGCTTTTGGAA – 3'

Amorce antisens ARNg 2: 5' – TTCTAGCTCTAAAACCTGTTCCTCAAAGCAATGT – 3'

Amorce sens ARNg 3: 5' – TAATACGACTCACTATAGCCCTACAAACACATTG – 3'

Amorce antisens ARNg 3: 5' – TTCTAGCTCTAAAACAAAGCAATGTGTTTGTAGG – 3'

---

Conception : Séquence du promoteur T7 – Extrémité 3' complémentaire entre la paire d'amorces – Séquence complémentaire à la portion ARNtrac de l'ARN guide.

---

**Matrice ADN servant à la transcription des ARN guides**

---

ARN guide 1:

5' – TAATACGACTCACTATAGTTGGAACAGAAGGTCAAGTGTGTTTTAGAGCTAGAA...GTCGGTGCTTTT – 3'

ARN guide 2:

5' – TAATACGACTCACTATAGCACATTGCTTTTGGAAACAGAGTTTTAGAGCTAGAA...GTCGGTGCTTTT – 3'

ARN guide 3:

5' – TAATACGACTCACTATAGCCCTACAAACACATTGCTTTGTTTTAGAGCTAGAA...GTCGGTGCTTTT – 3'

---

Conception : Séquence du promoteur T7 – Séquence cible des ARN guides – Séquence composant la portion ARNtrac de l'ARN guide.

---

**ADNss servant de matrice pour la génération de la mutation d'intérêt**

---

5' – TCTCTACTCTTGCTTTCTGGTTGCCATTCTAGCAATACAAGACCAAGTCAAGCTACAAGGCTTTTGCAGCAA  
TCCCTACAAACACATTGCTTTTGGAA**TAG**AAGGTCAGTGTTGACTCAAAAACAGTGAATTTTACATTTTAAATGT  
GATTTTAAAATGTTTTCTAAAATTTCTGTTGGGAAGATTCTGTTTTAAA – 3'

5' – ATCTCTACTCTTGCTTTCTGGTTGCCATTCTAGCAATACAAGACCAAGTCAAGCTACAAGGCTTTTGCAGCA  
ATCCCTACAAACACATTGCTTTTGGAA**TAG**AAGATCAGTGTTGGCTCAAAAACAGTGAATTTTACATTTTAAATG  
TGATTTTAAAATGTTTTCTAAAATTTCTGTTGGGAAGATTCTGTTTTAAA – 3'

5' – TACTCTTGCTTTCTGGTTGCCATTCTAGCAATACAAGACCAAGTCAAGCTACAAGGCTTTTGCAGCAATCCC  
TACAAACACATTGCTTCTTGA**TAG**AAGGTCAGTGTTGGCTCAAAAACAGTGAATTTTACATTTTAAATGTGATT  
TTAAAATGTTTTCTAAAATTTCTGTTGGGAAGATT – 3'

---

Mise en évidence : le codon natif passant de **CAG** à **TAG**, reproduisant la mutation candidate.

---



### 2.8.1 Synthèse des ARN guides

Les matrices ADN, à partir desquelles sont transcrits les ARNg, ont été construites en suivant les indications protocolaires du kit *GeneArt Precision gRNA Synthesis* (Invitrogen, A29377). Leur assemblage suit le schéma illustré à la Figure 10. Une paire d’amorces sens et antisens pour les ARN guides (Tableau 3) agissent comme matrice et amorce entre elles-mêmes avec une complémentarité à leur extrémité 3’ respective. Du côté 5’ de l’amorce sens se trouve une séquence pour le promoteur T7, tandis que l’amorce antisens possède une région complémentaire à la séquence palindromique conservée qui compose la portion ARNtracr de l’ARN guide. Après les premiers cycles de PCR, la matrice ADN est pleinement assemblée et est amplifiée subséquemment avec le couple d’amorces dites universelles, présentes dans le mélange réactionnel et fourni dans le kit. Les ARN guides sont ensuite transcrits, purifiés suivant les directives du kit et un dosage spectrophotométrique est réalisé afin d’évaluer qualité et quantité avant l’entreposage à -80.0°C.

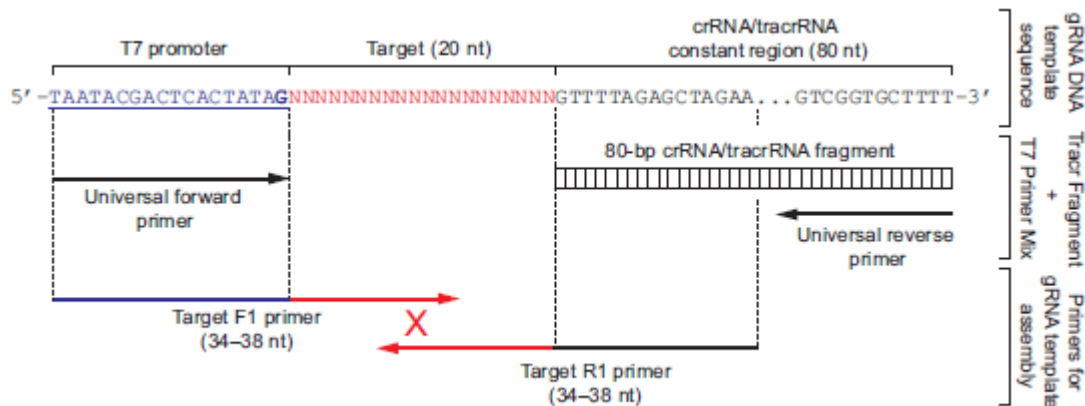


Figure 10. – Schéma d’assemblage des matrices ADN encodant les ARNg

Figure tirée du guide d’utilisateur accompagnant le kit Invitrogen A29377 (145).

### **2.8.2 Test de digestion d'hétéroduplexe**

Après transfection du complexe Cas9-ARNg, l'aptitude d'induire un DSB dans le génome des cellules est vérifié par test de digestion d'hétéroduplexe. L'ADN de la population cellulaire hétérogène est extrait, et la région d'intérêt est amplifiée par PCR avec le couple d'amorces GCDF/GCDR (Tableau 2) et les réactifs du kit *GeneArt Genomic Cleavage Detection Kit* (Invitrogen, A24372). Les amplicons subissent ensuite un traitement thermique spécifique servant à régénérer un hétéroduplexe sur un fragment d'ADN double brin, s'il y a lieu d'être. Le processus consiste en une décroissance thermique à rythme spécifique avec paliers allant comme suit : Dénaturation initiale 95.0°C, 5min → Réhybridation de 95.0°C à 85.0°C, -2.0°C/sec → Réhybridation de 85.0°C à 25.0°C, -0,1°C/sec → 4°C, ∞. L'essai de digestion est immédiatement réalisé à la fin du traitement thermique. Les amplicons à tester sont ensuite incubés avec 1µL de *Detection Enzyme* du kit (Invitrogen, A24372), tandis que les contrôles négatifs sont chargés avec 1µL d'eau certifiée dépourvue de nucléases. Une dernière incubation à 37.0°C pendant 1 heure est effectuée dans l'appareil 2720 Thermal Cycler d'Applied Biosystems, avant l'observation de la fragmentation sur gel d'agarose 1.0%.

### **2.8.3 Test de digestion avec enzyme de restriction**

Afin de vérifier le génotype des clones quant à l'incorporation de la mutation (KI), un test de digestion à l'enzyme de restriction est effectué en sollicitant Hpy188III du manufacturier NEB (R0622S). Le mélange réactionnel est composé de 1µg d'amplicons, 5U (1µL) d'enzyme de restriction Hpy188III, 5µL de tampon *CutSmart* 10X et le tout est complété avec de l'eau DEPC jusqu'à 50µL. Les contrôles négatifs ne contiennent pas d'enzyme de restriction, remplacée par 1µL d'eau DEPC supplémentaire. La digestion s'effectue en incubant à 37.0°C pendant une heure dans l'appareil 2720 Thermal Cycler d'Applied Biosystems. L'observation de la fragmentation s'effectue sur gel d'agarose 1.0%.

## 3 – Résultats

### 3.1 Article publié – Démonstration pathogénicité de *MLIP*

npj Genomic Medicine

**Novel homozygous nonsense mutation of *MLIP* and compensatory alternative splicing**

Jean Mezreani,<sup>1,2†</sup> Sébastien Audet,<sup>1,2†</sup> Florence Martin,<sup>1,2</sup> Jade Charbonneau,<sup>1</sup> Valérie Triassi,<sup>1,3</sup>

Eric Bareke,<sup>1</sup> Annie Laplante,<sup>1</sup> Jason Karamchandani,<sup>4,5</sup> Rami Massie,<sup>4,5</sup> Colin H Chalk,<sup>4,5</sup> Erin

O’Ferrall,<sup>4,5</sup> and Martine Tétreault <sup>1,2,3</sup>

† These authors contributed equally

1 CHUM Research Center, Montreal, Quebec, Canada

2 Department of Neurosciences, University of Montreal, Montreal, Quebec, Canada

3 Department of Bioinformatics, University of Montreal, Montreal, Quebec, Canada

4 Department of Neurology and Neurosurgery, Montreal Neurological Institute, Montreal, Quebec, Canada

5 Department of Pathology, Montreal Neurological Institute, Montreal, Quebec, Canada

Correspondence to: Martine Tétreault

Healthcare Center of the University of Montreal, Research Center Building

CRCHUM, 900 Saint-Denis

Montréal H2X 0A9, Quebec, Canada

E-mail: [martine.tetreault@umontreal.ca](mailto:martine.tetreault@umontreal.ca)

### 3.1.1 Abstract

Despite the growing accessibility of clinical sequencing, functional interpretation of variants remains a major hurdle to molecular diagnostics of mendelian diseases. We aimed to describe a new adult-onset myopathy with muscle weakness and hyperCKemia caused by a nonsense variant in muscular *LMNA*-interacting protein (*MLIP*). Following RNA-sequencing, differential expression analysis uncovered a significant downregulation of this gene, which had a surprisingly mild effect on *MLIP* protein expression. RT-PCR and long-read sequencing (LRS) both support an important transcriptome shift in the patient, where decreased *MLIP* levels are seemingly due to nonsense-mediated decay of transcripts containing the exon 5 mutation. Moreover, a compensatory mechanism upregulates the functionally lacking isoforms, and generates novel transcripts. These results support the recently discovered clinical implications of *MLIP* variants in myopathies, highlighting for the first time its relevance in adult-onset cases. These results also underline the power of LRS as a tool for the functional assessment of variants of unknown significance (VUS), as well as the definition of accurate isoform profile annotations in a tissue specific manner.

#### Keywords

Myopathy, Adult-onset, Long-read sequencing, MLIP, novel transcripts

### 3.1.2 Introduction

Myopathies and muscular dystrophies are a large group of progressive genetic neuromuscular diseases that emerge from dysfunctions and irregularities within the muscular tissue, affecting its general integrity, structure and molecular activity <sup>1</sup>. Gradual degeneration of the muscle fibers prompts variable clinical presentation, generally including but not restricted to; muscular weakness, cramps, stiffness, posture instability and ultimately leading to a probable loss of motor functions <sup>1</sup>. Due to the notable heterogeneity of these pathologies, the association of putative disease variants with phenotypes is highly complex. This hinders the ability to achieve an official diagnosis, as further genetic and molecular characterization is often required but strenuous. This drawback is amplified in genes that have yet to be extensively studied from a functional standpoint and have not yet been associated with muscle-related disorders.

Muscular LMNA-Interacting Protein (*MLIP*) has only recently been identified as a cause of autosomal recessive myopathy with rhabdomyolysis, myalgia and elevated serum creatine kinase <sup>2</sup>. The report highlights eight different variants in the *MLIP* gene within seven pediatric biallelic carriers, all of which are expected to cause major protein alterations: four frameshifts leading to premature stop codons, two stop-gains, as well as two alternative splicing which respectively induce an exon skipping and a premature stop codon <sup>2</sup>. While literature concerning the *MLIP* function is not extensive, it is generally thought to play a role in the regulation of transcriptional activators in pathways such as the *Akt/mTOR/FOXO1* cardiac stress response <sup>3,4</sup>, and is known to interact with lamin A/C (*LMNA*) <sup>5</sup>, which is a largely studied cause of muscle diseases <sup>6,7</sup>. Interestingly, all the investigated mutations localize near exon 4 or 9, resulting in a loss of the nuclear localization signal (NLS) and/or the AT-Hook DNA binding motif, respectively

<sup>5,8</sup>. It has been suggested that such major changes theoretically result in a considerable functional effect, therefore likely leading to dysregulations of various important biological pathways <sup>2</sup>.

Considering the group had access to a small cohort of patients with similar variants, they were able to demonstrate functional changes through the consistent reduction in mRNA levels of *MLIP*. The quantitative droplet digital PCR suggests nonsense-mediated decay, which should ultimately result in lowered protein levels and guided our initial hypothesis. In a single patient (Z46) exhibiting comparable but rather different phenotypes, a more in-depth functional assessment is required to assess the relevance of *MLIP* as the best candidate to explain the distal myopathy. Hence, on top of standard validation through Western Blot, quantitative PCR (qPCR) and reverse transcriptase PCR (RT-PCR), we propose long-read sequencing (LRS) as a potent clinical tool to identify accurate tissue-relevant transcripts, as well as being indicative of mRNA level changes in an isoform-specific manner.

### **3.1.3 Results**

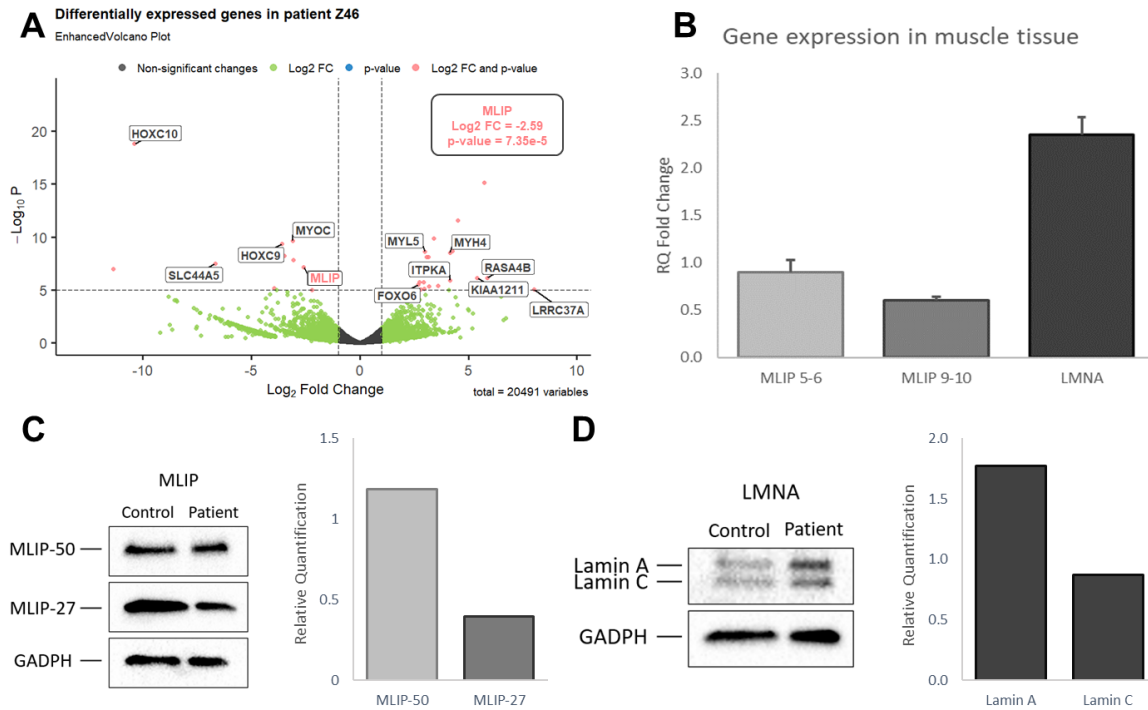
#### **Clinical examination**

Our patient presented around age 50 with distal (mainly dorsiflexors) more than proximal leg weakness and distal hand weakness. Despite a late-onset of debilitating symptoms, careful history revealed that he had consulted around age 5 for poor sports performance (unable to run as fast as his peers and to ice skate) but never received a clear diagnosis. The absence of muscle disorders or consanguinity in the familial history was also noted. By age 60 he had developed exaggerated lumbar lordosis and a Trendelenburg gait. Muscle pain and cramping were

observed but no rhabdomyolysis was reported. Evaluation of basal serum creatine kinase did reveal a significant increase (888U/L) that lands in the observed hyperCKemia range of 300-3000U/L<sup>2</sup>. Similar to the pediatric cases, our patient doesn't exhibit cardiac abnormalities, which could have been expected considering *MLIP* known role in adaptation to cardiac stressors<sup>3</sup>. Muscle biopsy showed only a single necrotic fiber, but many fibers with internal nuclei, fiber-size variability with hypertrophied fibers, type 1 fiber atrophy and type 2 fibers containing small, rimmed vacuoles. EMG demonstrated small polyphasic motor units with some fibrillations and positive sharp waves. ECG, ECHO and a 24H Holter monitor were also conducted but results were unremarkable.

### **RNA-sequencing**

Following inconclusive gene panels testing (Supplementary Table 1)<sup>9</sup>, RNA-sequencing served as a diagnostic tool for this complex case, where variant-calling data revealed a nonsense homozygous mutation in the exon 5 of *MLIP* (NM\_001281747.2: p.Gln762Ter, c.2284C>T). This finding was confirmed by Sanger sequencing of the patient's DNA. Upon performing a differential expression analysis using DESeq, a highly significant decrease can be observed in *MLIP* expression (Fig. 1A)<sup>10</sup>. An exon usage analysis was performed to identify alternative splicing in Z46 but only confirmed the global decrease of *MLIP* expression (Supplementary Figure 2). A pathway enrichment analysis, aimed at identifying additional validation targets and molecular changes, yielded no significant results regarding *MLIP* known functions or main interaction partners<sup>11,12</sup>.



**Figure 1. *MLIP* and *LMNA* respective expression in muscular tissue (A)** Volcano Plot highlighting the most differentially expressed genes in Z46 compared to controls following DESeq analysis. **(B)** RT-qPCR reveals mRNA expression levels for the targeted genes: *MLIP* 5-6 accounts specifically for transcripts containing the nonsense mutation, *MLIP* 9-10 covers all the known transcripts, and *LMNA* probe accounts for both lamin A and C. RPS29 is used for data normalization, error bars correspond to standard deviation (triplicata). **(C-D)** Western Blot analysis reveals the presence of four isoforms of *MLIP*, two of which can be quantified (27kDa and 50kDa). Both lamin types also appear for *LMNA* and were analyzed separately. *GADPH* is used as the reference gene for normalization of the blot in both instances.



## Expression data

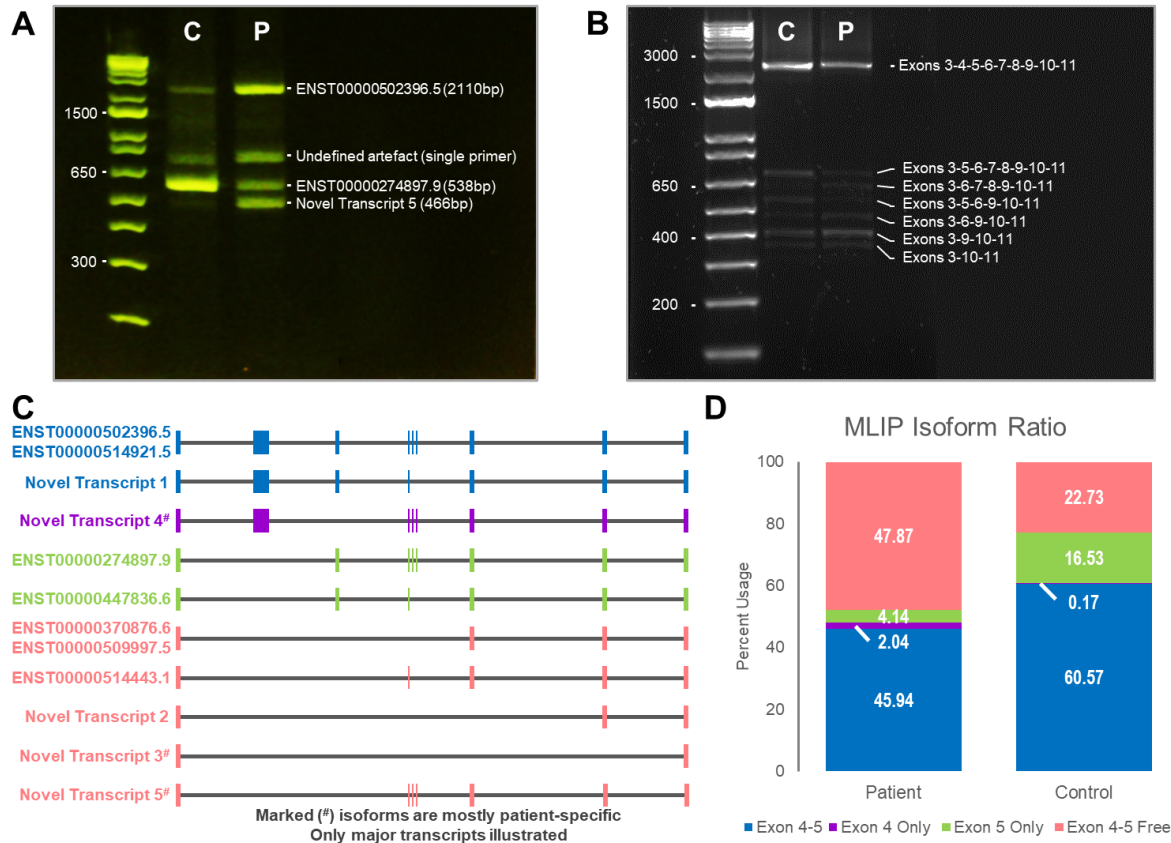
The downregulation of *MLIP* was consolidated using two qPCR probes, targeting either transcripts that specifically contain the nonsense variant, or accounting for all isoforms. The latter revealed a similar tendency to the RNA-sequencing: although more modest, a decrease of about 40% of *MLIP* transcripts is observed (Fig. 1B). Interestingly, a considerable upregulation of *LMNA* was also noted. Regarding the Western Blot, preliminary results showed a correlating downregulation of MLIP for the 27kDa isoform (ENST00000370876.6), which contains neither exon 4 or 5. A negligible change for the 50kDa (ENST00000274897.9), which theoretically carries the exon 5 mutation and should not be produced, was unexpected (Fig. 1C). Unfortunately, NLS containing isoforms (ENST00000502396.5 & ENST00000514921.5) could not be quantified, as the extraction was not specific to nuclear proteins, leading to very light bands for both samples (Supplementary Figure 3). Again, increased levels of *LMNA* can be observed, especially lamin A, exhibiting nearly two-fold increase in comparison to control (Fig. 1D).

## Isoform balance

RT-PCR was performed between the 3rd and 8th exon of *MLIP*, capturing most muscle transcripts while keeping the interpretation simple. Although qualitative, gel migration revealed a clear alteration of the patients' transcript balance: while the second to lowest band is the main isoform in the control, it is considerably decreased in Z46 compared to other visible transcripts, which appear upregulated (Fig. 2A). When correlating amplicons with the GTEx muscle isoforms repertoire, we supposed the presence of transcripts ENST00000502396.5,

ENST00000274897.9, and a novel isoform containing exons 3-6-7-8, making it the only fragment that does not carry the variant.

LRS subsequently allowed isoform-specific quantification of all known transcripts. Data reveal a near perfect correlation with RT-PCR results (Fig. 2B), where a drastic downregulation of isoforms containing exon 5 is milder when exon 4 is co-expressed, and the proportional expression of transcripts that do not carry the mutation is significantly increased (Fig. 2C-D). Additionally, we discovered that the patient seems to produce and conserve novel transcripts that contain exon 4, but not 5. Furthermore, they're not the only novel transcripts that our analysis revealed: while ENST00000502396.5 and ENST00000370876.6 are indeed the main muscle isoforms, ENST00000447846.6, ENST00000514433.1, as well as two other newly identified transcripts (labeled Novel Transcript 1 & 2) exhibit levels comparable to ENST00000274897.9 in both samples (Fig. 2C). Ratio quantification utilized data from all the isoforms, but only high-confidence transcripts are presented as potential novel isoforms, none of which exhibit alternative splicing that differs from canonical reading frames.



**Figure 2. Evaluation of nonsense variant effect on transcript balance of *MLIP*** (A) Migration of RT-PCR products on 2% E-gel following amplification between exons 3 and 8. 1Kb Plus E-gel DNA ladder was used for determination of fragment lengths. P = patient; C = Control (B) Qualitative visualization of 100ng of sequencing library on a 1% agarose gel with GelRed revealing agent. Migration was done prior to the end prep step, and 1Kb Plus DNA Ladder served for amplicons length evaluation. (C) Visual representation of major (high confidence) muscle transcripts identified with LRS data. Corresponding annotated transcripts are identified, and all isoforms are color-coded depending on grouping features. (D) Representation of the samples' relative transcripts balance, grouped by relevant primer features. Isoform quantification was performed by FLAIR, and final data contains 82.9k (patient) and 52.3k (control) reads.

### 3.1.4 Discussion

In this report, we used RNA-sequencing in a clinical setting after standard gene panels failed to identify candidates for the patient's pathology. While variant calling did not find pathogenic mutations in relevant myopathy-related genes (Supplementary Table 1), a novel homozygous variant causing an early truncation of *MLIP* on both alleles was predicted to have a highly negative effect on its function. Although there were very few studies about *MLIP* function, and no links to any pathology, numerous factors hinted at the relevance of this locus to myopathies. A muscle-specific expression pattern and a proven interaction with *LMNA* <sup>5</sup>, combined to the high in silico pathogenicity score as well as the decreased mRNA expression in transcriptomic data prompted further investigation of the variant. Very recently, an article identified *MLIP* as disease-causing in a cohort of patient exhibiting early-onset myopathy with rhabdomyolysis <sup>2</sup>. Similarly to the pediatric variants, the premature translation termination in exon 5 is predicted to importantly modify both the structure and the functions of the protein by inducing a biallelic absence of its AT-Hook DNA binding motif <sup>8</sup>. Indeed, current literature states that in addition to being present in two of the three main muscle isoforms, exon 5 is mutually inclusive to transcripts containing exon 4. Considering the latter contains the NLS <sup>5</sup>, the mutation reported in our patient could potentially hinder *MLIP* nuclear role. Nevertheless, it seems the localization of variants affects phenotypic presentation, as clinical examination findings are considerably different between our patient and the younger cohort.

One of the advantages of RNA-sequencing is the functional insight that can be gathered from expression data <sup>13</sup>. It has been hypothesized that dysfunctional *MLIP* potentially disrupts multiple molecular pathways, such as causing an overactivation of the Akt/mTOR or FOXO1

pathway in correlation to a reduced adaptive response to stressors<sup>3,4</sup>. Unfortunately, there was no significant enrichment of the genes related to those networks in our data. In a similar fashion, nothing noteworthy can be found in regard to both glycolysis and gluconeogenesis<sup>2</sup>. Observable molecular changes include the activation of the host immune system, which indicates probable progressive necrosis of unhealthy muscle cells<sup>14</sup>. The Syndecan-1 pathway also appears significantly upregulated, which might play a key role of protecting against cardiac phenotypes<sup>15</sup>. While our study gives interesting insights into potential molecular consequences of *MLIP* mutations, the analysis was performed on a single patient, and conclusive results surrounding pathway enrichments require more samples.

As the quantification of mRNA expression is an important step of the functional validation, we made sure to capture changes accurately by using two probes simultaneously. While a larger decrease could have been expected following the differential expression analysis, results correlate the general decrease of *MLIP* mRNA in muscle tissues. These non-significant observations support our hypothesis that while *MLIP* mRNA likely undergoes nonsense-mediated decay, the cells attempt to compensate for truncated proteins by specifically splicing transcripts to contain exon 4 and produce the nuclear protein. This would rationalize the considerable upregulation of *LMNA*, presumably as a compensatory mechanism for the deficiency of its direct partner<sup>5</sup>. Regarding protein levels, only two of the four visible isoforms can be quantified, but results once again correlate previous experiments: *MLIP* expression is decreased, causing upregulation of *LMNA* to compensate for the likely diminished yet necessary interaction between the nuclear proteins. The presence of unaltered *MLIP* proteins in the patient was unexpected. While the C-terminal antibody does not allow detection of truncated

proteins, normal bands in Z46 proves that muscle cells find a way to produce nearly full-length proteins despite the stop-gain. Two hypothetical explanations for this phenomenon are an elevated proportion of stop-codon read-through in response to reduced functional *MLIP* levels<sup>16</sup>, and alternative splicing allowing non-canonical excision of the exon 5 specifically<sup>17,18</sup>. These partial escape mechanisms, if observed, are likely to play a role in the milder phenotypes of our patient compared to pediatric cases.

It has been shown that distinct isoforms can have tissue-specific functions, hence the aim of evaluating changes in transcripts ratio<sup>19</sup>. The qualitative RT-PCR was key in constructing the hypothesis of conflicting cellular mechanisms: nonsense-mediated decay appears to oppose the upregulation of the functionally relevant *MLIP* transcripts in response to the deficiency of proteins capable of translocating to the nucleus. This came from visible downregulation of transcripts containing exon 5, while other isoforms are promoted in comparison. As transcripts containing the NLS theoretically always carry the variant, they require a substantial transcription upregulation to overcome concurrent mRNA decay.

LRS data confirmed our suppositions, clearly highlighting a significant decrease of transcripts containing the variant as well as a proportional increase of other transcripts, particularly those containing exon 4. In fact, not only is it observable in canonical isoforms, but Z46 expresses novel transcripts containing exon 4 without its usual splice partner. While those new isoforms represent a minority of the *MLIP* transcriptome, they confirm the hypothesis that compensatory alternative splicing allows production of nearly full-length proteins. Another unexpected result was that additional unreported non-canonical transcripts were present in non-negligible amounts for both samples. This suggests that their expression would not be restricted to our

study, and that the tissue-specific alternative splicing of *MLIP* might be even more extensive than what current literature suggests. While further assessment is required to confirm these findings, they prompt the relevance of performing LRS on more samples. Furthermore, we believe the approach is very powerful, and could uncover similar results in a broader range of pathologies, underlining its vast potential in clinical contexts. Indeed, while LRS currently offers less accurate base calling than its short read homolog, its true strength lies in the assembly of *de novo* tissue-specific transcript annotations, which hints towards a considerable improvement of current knowledge regarding existing transcriptomic splicing<sup>20</sup>. While long-read methods are still in an early developmental stage, community enthusiasm has led to consistent improvement surrounding the approach and related bioinformatics tools. Therefore, it would not be surprising if LRS applications reach a largescale, whole-transcriptome level in a not-so-distant future.

Overall, our report of a biallelic nonsense mutation supports the implication of *MLIP* as a causative gene for myopathies with muscle weakness and hyperCKemia. The important element we highlight is that clinicians should consider *MLIP* mutations outside of pediatric cases, as adult-onset phenotypes, albeit milder, could ensue from functional changes on the protein. Indeed, clinical heterogeneity is to be expected in *MLIP*-related disorders, mainly due to its complex transcriptomic architecture. To that regard, LRS shall help define more accurate tissue-specific transcript annotations. Along with future cellular modeling of the mutations, it should give additional insights on the protein molecular functions. We strongly believe our findings should further assist clinicians in the diagnosis of patients carrying *MLIP* variants, but also in the

assessment of functional effects of candidate variants, where available transcript annotations are potentially incomplete or inaccurate.

**\*\*\* La fin de l'article (Sections *Methods*, *References* et *Supplementary Material*) se trouve en fin du mémoire dans la section 7 – Annexe.**

**\*\*\* La numérotation des références dans le corps de l'article est distincte de celle du mémoire. Voir les références en annexe.**



## **3.2 Modélisation de la mutation candidate *MLIP* par CRISPR/Cas9**

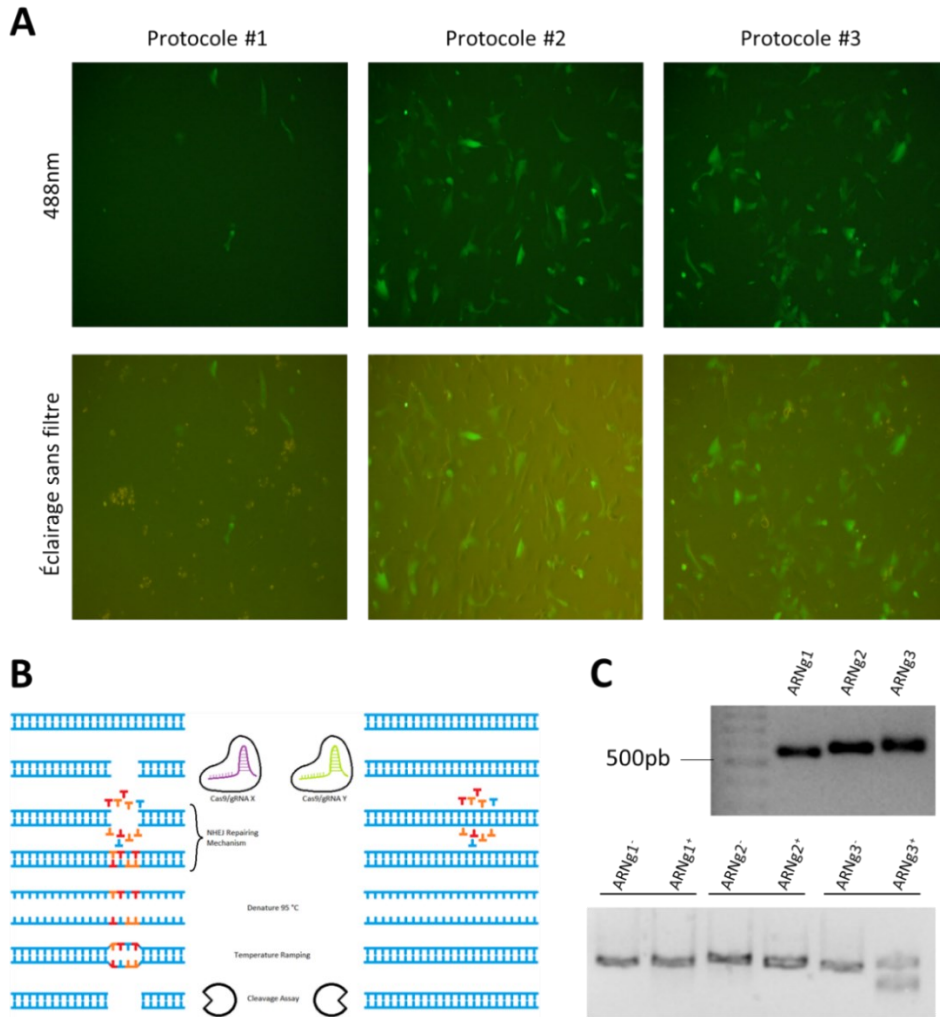
Suivant l'analyse des données -omiques du patient et l'identification du variant non-sens homozygote dans *MLIP*, les perturbations transcriptomiques subséquentement observées (Figure 1-2, Article) prouvent son implication dans la myopathie distale du patient. La suite logique des choses est donc d'enquêter sur son mécanisme pathogénique; la compréhension de l'origine du désordre constituant les fondations pour l'éventuel établissement de stratégies thérapeutiques. La modélisation de la mutation candidate dans une lignée de myoblastes humains immortalisés permettrait ainsi d'en apprendre davantage sur ses effets pathogéniques, en même temps de recueillir des informations relatives aux rôles moléculaires normaux de *MLIP* dans le muscle. Finalement, la modélisation cellulaire renforcera également l'association génotype-phénotype, suivant l'identification de perturbations cellulaires supplémentaires.

### **3.2.1 Optimisation des paramètres expérimentaux CRISPR/Cas9**

La mise en place d'expériences novatrices passe inévitablement par des phases d'optimisation. Ici, l'efficacité de la méthode de transfection et l'aptitude du complexe Cas9-ARNg à induire une DSB doivent être confirmées. La Figure 11 confirme la livraison de la cargaison par électroporation et la bonne conception et fonctionnement du complexe Cas9-ARNg3. La Figure 11A évalue qualitativement l'efficacité de transfection d'un plasmide GFP et le taux de survivabilité cellulaire pour trois protocoles distincts. Le second protocole (1650V, 1 pulsation, 20ms) fut retenu pour les manipulations ultérieures, octroyant les meilleurs rendements dans les deux catégories.

Avec les paramètres d'électroporation en main, les transfections du complexe Cas9 et des trois ARNg sont effectuées dans trois populations cellulaires distinctement. Les ADN totaux de chacune des instances sont récoltés et amplifiés par PCR avec le couple d'amorces GCDF/GCDR (Tableau 2). La Figure 11C illustre la bonne amplification de la région d'intérêt avec la présence d'un amplicon de 516pb, même après l'édition génomique. Ces mêmes amplicons sont repris, dénaturés, et réhybridés thermiquement de façon spécifique, avant d'être incubés avec ou sans enzyme de digestion (Endonucléase 1 T7) pour l'essai de digestion d'hétéroduplexe. La visualisation d'une fragmentation avec deux bandes de moindre intensité dans la colonne ARNg3<sup>+</sup>, prouve qu'il y a eu digestion de l'hétéroduplexe, sous-entendant un bon fonctionnement

du complexe Cas9-ARNg3 dans le génome des cellules. L'ARNg3 étant le seul à présenter une fragmentation sur le gel d'agarose, le troisième design CRISPR fut sélectionné pour l'expérience de KI.

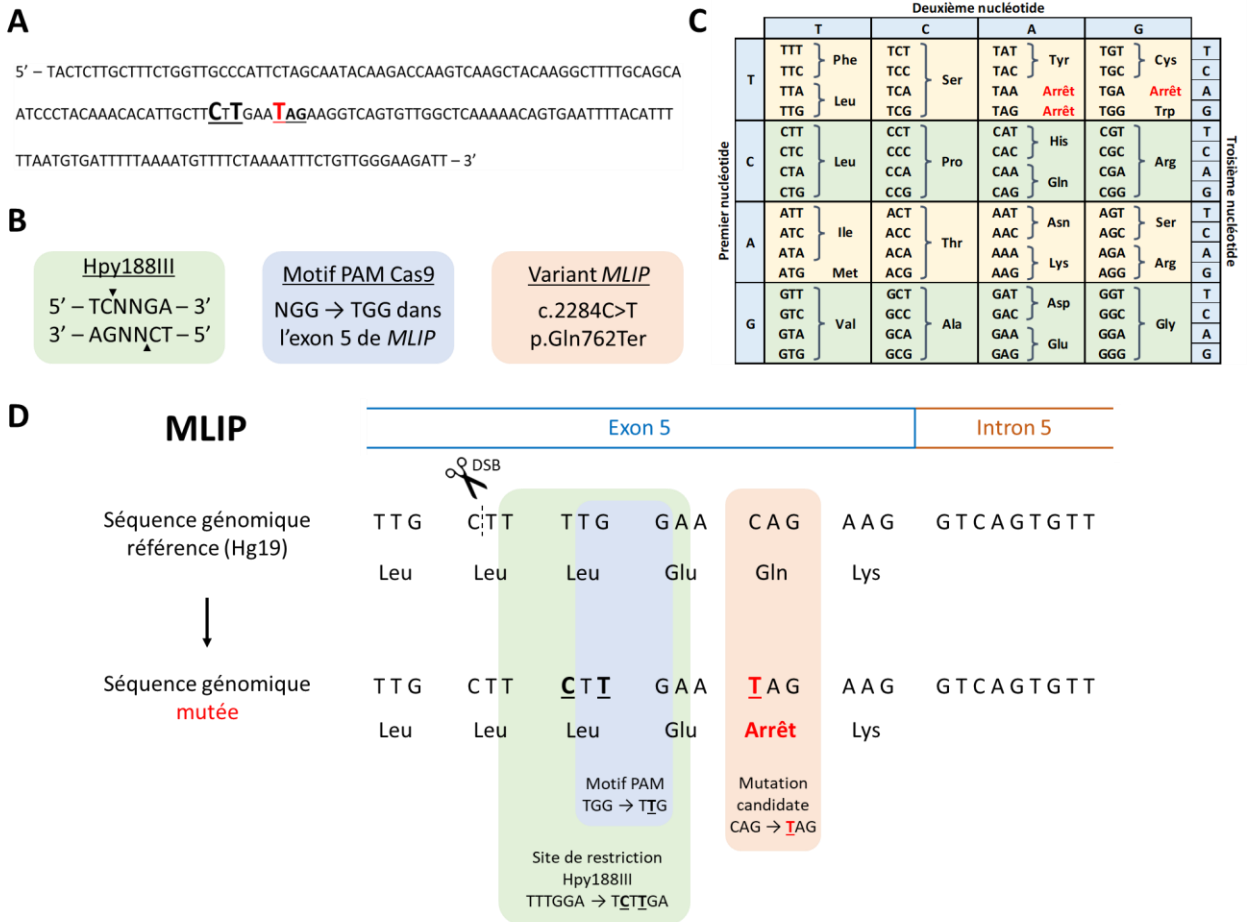


**Figure 11. – Optimisation des paramètres expérimentaux CRISPR/Cas9**

**A.** Trois protocoles d'électroporation distincts furent testés : 1) 1475V, 2 pulsations, 20ms; 2) 1650V, 3 pulsations, 10ms; 3) 1680V, 1 pulsation, 20ms. Les images sont capturées une première fois à 488nm, et une seconde fois sans filtre. **B.** Schématisation du principe du test de digestion d'hétéroduplexe, lors de la vérification du bon fonctionnement du complexe Cas9-ARNg. **C.** Amplification PCR des régions d'intérêt avec le couple d'amorces GCDF/GCDR (Tableau 2), suite à la transfection distincte de la Cas9 et des ARNg1, 2 et 3. Reprise des amplicons, application du traitement thermique et incubation avec (+) et sans (-) enzyme de restriction.

### 3.2.2 Conception de l'édition génomique KI de *MLIP*

Suite aux résultats obtenus dans la section 3.2.1, il n'y a que la troisième instance qui a fonctionné, la seule capable d'induire une DSB dans le génome des myoblastes (Figure 11C). L'expérience de KI pour *MLIP* est donc réalisée avec le troisième design, en y couplant cette fois-ci son ADNss donneur méticuleusement conçu (Tableau 3). La Figure 12A-B-D récapitule les modifications clés introduites pour assurer le succès de l'édition génomique et la possibilité de vérifier le génotype à l'aide d'un test de digestion à l'enzyme de restriction. La matrice ADNss assure une reconstruction spécifique et centrée de la DSB de sorte à : 1) introduire la mutation candidate identifiée dans *MLIP*, modifiant le codon CAG de la glutamine (Gln, Q) par le codon d'arrêt de traduction TAG; 2) introduire deux mutations silencieuses dans le codon original TTG, devenant CTT et préservant toujours la leucine (Leu, L) comme acide aminé. La première mutation silencieuse s'assure de supprimer le motif PAM génomique, TGG, juxtaposé à la séquence du proto-espaceur ciblé par l'ARNg3. Pour rappel, le but étant ici de freiner l'activité catalytique de la Cas9 après sa première coupure dans le génome. Finalement, la combinaison des deux mutations silencieuses sert à la conversion de l'hexanucléotide TTTGGA en séquence palindromique TCNNGA, générant un site de restriction reconnu et clivable par l'enzyme de restriction Hpy188III (146). Toutes ces mutations étant étroitement centralisées avec la position de la DSB, cela devrait résulter en des chances accrues d'incorporation des mutations lors du processus de réparation.



**Figure 12. – Conception du design KI CRISPR #3 de *MLIP***

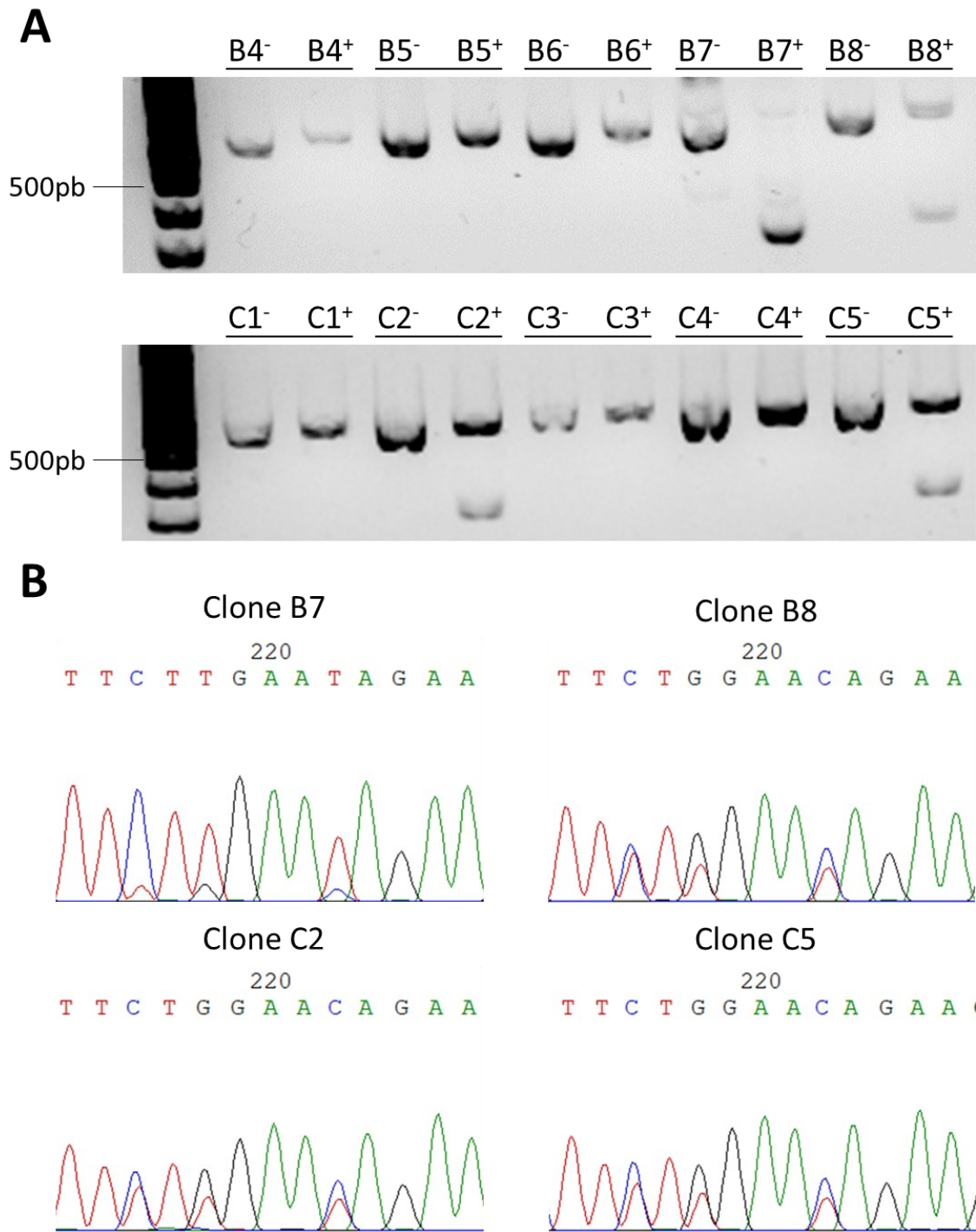
**A.** Représentation de la matrice ADNss arborant deux mutations silencieuses (en noir) et la mutation d'intérêt (en rouge). **B.** Objectifs visés par l'édition génomique de *MLIP* : génération d'un site de restriction enzymatique Hpy188III; suppression du motif PAM génomique TGG reconnu par l'endonucléase SpCas9; intégration de la mutation d'intérêt identifiée par RNA-seq. **C.** Code génétique présentant les différentes combinaisons de codons dans l'encodage des acides aminés. **D.** Illustration de l'édition génomique dans son ensemble : comparatif entre la séquence génomique *MLIP* originale et mutée; positionnement de la DSB; intégration des trois mutations clés à la fin de l'exon 5; et leurs conséquences sur la chaîne d'acide aminé résultante.

### 3.2.3 Vérification du génotype des clones

L'expérience du KI dans *MLIP* est réalisée en combinant les optimisations expérimentales et le design CRISPR #3. Suivant la transfection de la machinerie d'édition génomique, la population cellulaire hétérogène est recueillie et étalée dans des plaques de culture adhérentes 96 puits, de sorte à générer des clones par culture unicellulaire. Plusieurs dizaines de clones furent testés distinctement par test de digestion à l'enzyme de restriction Hpy188III. La présentation d'une fragmentation dans l'amplicon PCR original résulte à son envoi pour séquençage Sanger, à des fins de validations additionnelles et la confirmation du génotype du clone.

La Figure 13 présente quatre clones distincts ayant assimilé les différentes mutations abordées, suivant différents statuts de zygoté. Sur les gels d'agarose en Figure 13A, les clones B7, B8, C2 et C5 présentent tous une fragmentation de l'amplicon PCR original de 516pb. En ce qui a trait à B8, C2 et C5, il s'agit là d'une digestion partielle avec la présence de deux bandes en intensité réduite : une pour l'amplicon original et l'autre pour le fragment digéré. Cela laisse présager que les mutations génomiques n'ont été incorporées que sur un des deux allèles, conférant supposément le statut hétérozygote à ces trois clones. À l'inverse, le clone B7 présente une digestion complète de son amplicon avec une seule bande de digestion de forte intensité. Cette présentation suggère un statut homozygote pour ce dernier.

Les génotypes de ces quatre clones sont subséquemment confirmés par séquençage Sanger en Figure 13B. Avec la présence de doubles signaux aux positions 216, 218 et 222 des chromatogrammes, les clones B8, C2 et C5 s'apparentent tous bel et bien à un génotype hétérozygote. À l'occurrence du clone B7, les trois mutations discutées sont incorporées de façon unanime avec un signal dominant à chacune de ces positions. Ce résultat confirme le statut homozygote du clone, venant rejoindre le génotype du patient. Il faut toutefois noter la présence de légers bruits de fond pour le clone B7. Ces clones aux zygotés variables pourront être utilisés ultérieurement afin d'évaluer les impacts morpho-fonctionnels à différents niveaux.



**Figure 13. – Vérification du génotype des clones**

**A.** Amplicons PCR (GCDF/GCDR – Tableau 2) à partir des ADN isolés des clones, incubés avec (+) l'enzyme de restriction Hpy188III et sans (-) pour les contrôles. Révélation de la fragmentation sur gel d'agarose 1.0%. **B.** Chromatogrammes des séquences Sanger pour les clones présentant une fragmentation. Chromatogrammes analysés à l'aide du logiciel *GeneStudio*.

## 4 – Discussion

Ce projet de recherche est projeté par une analyse de données transcriptomiques chez un patient diagnostiqué avec une myopathie distale à apparition tardive, et l'identification d'un variant non-sens homozygote dans le gène *MLIP* chez ce dernier. Malgré une faible littérature à son endroit, avec un rôle moléculaire peu défini, et aucun variant pathogénique répertorié initialement, une multitude de pistes et indices ont mené à sa rétention pour des investigations approfondies. La mise en évidence éventuelle d'importants débalancements transcriptomiques de *MLIP* a permis de consolider la pathogénicité du variant, venant rejoindre l'association génotype-phénotype initialement postulée dans l'hypothèse, et conduisant à l'essor du projet vers une modélisation cellulaire de la mutation.

### 4.1 Pertinence de l'utilisation du RNA-seq comme outil diagnostique

À la suite de tests non concluants de panels de gènes, ne ressortant que deux variants hétérozygotes dans *RAPSN* et *SCN4A* non concordants avec les manifestations cliniques de la myopathie distale du patient, le RNA-seq a été l'outil décisif établissant le diagnostic moléculaire définitif du patient. Effectivement, au temps de sa découverte, aucun variant pathogénique dans le gène *MLIP* n'avait été répertorié dans la littérature, d'où l'inefficacité du panel de gène. Il était donc essentiel d'utiliser une approche NGS autre pour la résolution de ce cas précis, explorant tout l'étendu des gènes exprimés dans le tissu cible. Ayant eu accès à une biopsie musculaire, tissu atteint par le désordre, le choix d'opter pour le RNA-seq s'est imposé d'emblée. L'appel de variants a ressorti un nouveau variant homozygote non-sens dans le gène *MLIP* à la fin du cinquième exon. Apparaissant relativement tôt dans la séquence du transcrit et tronquant une grande portion protéique, dont le crochet AT présent sur les exons 9 et 10; cette altération est suspectée de grandement affecter les bonnes fonctions protéiques de *MLIP*. De surcroît, les renseignements relatifs aux niveaux d'expression génique apportés par le RNA-seq ont révélé une baisse globale de l'expression de *MLIP* et par phénomène compensatoire, une hausse de celui de *LMNA* (Figure 1, Article), sans doute pour pallier à l'interaction jugée nécessaire entre les deux

partenaires nucléaires. Ce résultat donne un premier aperçu quant à l'affectation fonctionnelle causée par le variant pathogénique.

#### **4.1.1 Pertinence de l'utilisation du LRS**

Suivant des analyses plus poussées vis-à-vis de l'expression tissulaire, l'ensemble des résultats penchent dans la même direction, avec une baisse d'expression globale observée à l'endroit de *MLIP* par qPCR (Figure 1B, Article). Un résultat qui n'était pas attendu était l'expression protéique relativement importante de *MLIP* observée par WB (Figure 1C-D, Article). Certes une décroissance est bien constatée, mais une détection de ce niveau n'était pas attendue; en partie à cause du codon d'arrêt prématuré dans l'exon 5 et le fait que l'anticorps primaire de détection se lie à la portion C-terminal de la protéine *MLIP*. Couplé à cela, un gel de RT-PCR (Figure 2A, Article) révélant une bande majoritaire supplémentaire chez le patient a poussé le questionnement davantage.

Bien que *MLIP* est connu pour avoir un patron d'expression transcriptomique et protéique complexe, la présence d'une bande supplémentaire chez le patient au niveau de l'ARNm et un profile protéique presque normal ont poussé à employer le LRS pour une caractérisation transcriptomique plus avancée de *MLIP*. En effet, avec les longues lectures proposées par la technologie couvrant essentiellement toutes les jonctions d'exons des multiples transcrits *MLIP* sur une même lecture, cela devrait conférer la meilleure représentation de la composition exonique et permettre une analyse fine des isoformes observées. C'est exactement ce qui a été obtenu. Le séquençage nanopore a permis de caractériser, quantifier et classifier efficacement le transcriptome complexe de *MLIP*, révélant un motif d'épissage alternatif spécifique au patient. Il s'avère que les cellules musculaires du patient cherchent à contourner ou minimiser la production de transcrits ARN arborant le cinquième exon muté (Transcrits bleus et verts, Figure 2C-D, Article). Conséquemment, on note une plus grande proportion de transcrits ne possédant pas la mutation (Transcrits roses, Figure 2C-D, Article) chez le patient. Un point très intéressant qui a pu être relevé, est la présence non négligeable d'un transcrit non canonique comprenant exclusivement l'exon 4 sans son partenaire d'épissage habituel, l'exon 5 (Transcrit mauve, Figure 2C-D, Article). Observé à la fois chez le contrôle et en plus grande proportion chez le patient, il s'agirait là d'un



autre effet résultant du mécanisme d'échappement partiel. L'ensemble de ces résultats consolident et établissent de manière plus quantitative les déviations transcriptomiques claires entre le patient et le contrôle, en plus d'apporter des éléments de réponses aux résultats précédents le LRS. Le mécanisme d'épissage compensatoire observé explique comment les protéines MLIP complètes (ou presque) peuvent être traduites et détectées par WB, malgré la mutation non-sens. Aussi, la bande additionnelle observée sur gel RT-PCR s'apparente à la combinaison d'exons 6-7-8, reflétant un des nouveaux transcrits non répertoriés dans la littérature. De surcroît, l'identification de nouveaux transcrits non répertoriés dans la littérature témoigne également de la puissance de caractérisation du LRS. La confirmation de tels résultats nécessiterait une évaluation plus approfondie sur plus d'échantillons, afin d'élucider tout l'étendu de l'épissage alternatif de *MLIP*. Cela n'est pas vrai que pour *MLIP*, mais pour bien d'autres candidats aussi. La force du LRS résidant principalement dans l'assemblage d'annotations *de novo* de transcrits, cet outil a le potentiel d'améliorer considérablement les connaissances actuelles concernant l'épissage transcriptomique (90).

## 4.2 Cohorte pédiatrique avec variants *MLIP*

Au temps de l'identification de la mutation non-sens chez le patient Z46, aucun autre variant pathogénique entourant *MLIP* n'avait été répertorié dans la littérature. Depuis, Lopes Abath Neto *et al.* ont relevé huit variants distincts comme causatifs de myopathies récessives avec rhabdomyolyse et myalgie chez une cohorte de sept patients pédiatriques; à la différence de notre patient présentant une myopathie distale sans rhabdomyolyse à apparition tardive (102). Ayant eu des résultats très pertinents avec le LRS, nous avons entamé des procédures de collaboration avec le groupe de recherche afin de reproduire les mêmes expériences avec leurs échantillons. Dans cette circonstance, il serait pertinent de porter attention à différents points clés ressortis avec l'analyse de notre patient, comme l'évaluation des niveaux de transcrits arborant ou non le variant pathogénique. Percevrions-nous également les mêmes nouveaux transcrits avec des combinaisons d'exons non répertoriés dans la littérature? Distinguerions-nous un phénomène d'épissage alternatif compensatoire, relatif à chacune des mutations, ou est-ce là un mécanisme d'échappement propre à notre patient seulement? Compte tenu de l'architecture transcriptomique complexe de *MLIP*, une importante activité d'épissage alternatif doit

impérativement y être inhérente. Les expériences de LRS permettront de confirmer ou non la présence de mécanismes de contournement dans la cohorte pédiatrique. S'il s'agit là d'une exclusivité du transcriptome du patient Z46 améliorant les niveaux de protéines MLIP complètes produites, cela pourrait être la raison qui explique les manifestations cliniques tardives de celui-ci? Globalement, la reproduction des expériences de LRS sur un plus grand nombre d'échantillons permettra un important enrichissement de la littérature entourant *MLIP*, tant sur un plan fondamental, que clinique. Le point clé que nous souhaitons souligner est l'importante hétérogénéité clinique qui peut être rencontrée autour de *MLIP*, les autres trouvailles connexes n'ayant répertorié que des défauts géniques causatifs dans un contexte pédiatrique.

#### **4.2.1 Absence de phénotype cardiaque**

Un point très intrigant dans tous ces cas cliniques est l'absence de phénotype cardiaque clair, compte tenu de l'expression tissulaire de *MLIP* et l'implication démontrée de *LMNA* altérée dans des cardiomyopathies diverses. Chez le patient Z46 aucune atteinte cardiaque n'avait été rapportée, tandis que du côté des patients pédiatriques, de minimes anomalies structurales, ventriculaires et aortiques avaient été identifiées chez la moitié de la cohorte (102). Malgré cela et avec l'état actuel de la littérature, on ne peut établir d'association évidente entre des variants *MLIP* et des phénotypes pathologiques cardiaques. Plusieurs éléments de réponses peuvent tendre à expliquer cette interrogation.

Il pourrait s'agir d'une question d'isoforme. En effet, comme mentionné préalablement, *MLIP* possède un transcriptome très complexe avec un grand nombre d'isoformes distinct, certain étant tissu spécifique (106). Cette répartition disparate du profil d'isoforme pourrait justifier l'atteinte musculaire et l'exemption cardiaque. La question de la dégénérescence musculaire n'est pas à écarter non plus. Malgré l'âge plutôt avancé du patient Z46 qui en témoigne autrement, c'est malheureusement une hypothèse qui pourrait encore affecter la cohorte pédiatrique. Dans un autre ordre d'idée, les niveaux d'activité d'épissage alternatif peuvent varier et agir avec des efficacités variables en fonction du tissu concerné (147). Jumelant ce concept au mécanisme de contournement observé dans les cellulaires musculaires du patient Z46, on pourrait hypothétiser une performance compensatoire variable dans les tissus, qui pourraient ici

atténuer les manifestations cardiaques spécifiquement. Finalement, il a été démontré que la dérégulation de *LMNA* peut engendrer une multitude de myopathies, affectant soit le tissu musculaire squelettique, soit cardiaque ou encore les deux en simultané (148-150). Au même titre que pour *LMNA*, nous pourrions faire face à un spectre de pathologie avec *MLIP*. Le faible échantillon de cas clinique demeure limitant quant à la pleine compréhension de ce nouvel acteur, d'un point de vue normal et pathologique.

L'hétérogénéité clinique dépeinte de *MLIP* apparaît être tout aussi complexe que son transcriptome ne l'est. Globalement, la pathogénicité générale des variants *MLIP* a bel et bien été démontrée, tant du côté pédiatrique que tardif. Il restera à préciser l'association génotype-phénotype avec l'évolution de la littérature et les analyses raffinées qui en découleront.

### **4.3 Modélisation CRISPR/Cas9**

Les multiples déviations transcriptomiques soulevées tendent donc à affermir l'implication du variant pathogénique dans la myopathie distale du patient. Par conséquent, le projet poursuit sa lancée sur une modélisation cellulaire de la mutation dans des myoblastes humains immortalisés. La production d'un modèle mutant permet une évaluation plus approfondie des impacts fonctionnels causés par le variant pathogénique.

L'intérêt premier de cette expérience est de générer un modèle mutant arborant un KI homozygote, reflétant le génotype du patient. Toutefois, il n'en demeure pas moins pertinent de travailler avec toutes les différentes instances du mutant en parallèle. Ici, l'obtention de résultats montrant les mutations d'intérêt incorporées de façon homozygote dans le clone B7 (Figure 13) témoigne de la bonne conception, optimisation et exécution de l'ensemble des manipulations entourant l'expérience d'édition génomique par CRISPR/Cas9. En effet, la digestion du fragment qui semble complète sur le gel (Figure 13A), et les intensités des signaux sur le chromatogramme dominant pour les mutations d'intérêt (Figure 13B), sous-entendent une édition réussie des deux allèles des myoblastes. À l'inverse, les clones B8, C2 et C5 semblent démontrer un statut hétérozygote avec une fragmentation partielle de l'amplicon initial (Figure 13A) et des signaux d'intensité partagée aux positions des mutations (Figure 13B). Bien que la réalisation de l'expérience semble s'être bien déroulée dans son ensemble, générant des mutants homozygotes

et hétérozygotes, une petite impureté persiste dans les résultats collectés à l'endroit du clone B7. Tant sur le gel que sur le chromatogramme, il y a présence d'un faible bruit de fond qui doit être investigué davantage. Il restera donc à vérifier la nature de cette faible impureté, ou alors de réextraire un clone pur à partir de cette colonie précise, afin d'avoir notre modèle mutant homozygote final.

Finalement, malgré la sélection des ARNg avec les meilleures spécificités de ciblage, il faut tout de même attester qu'aucune autre région génomique n'a été altérée par des effets hors cible. Pour ce faire, l'utilisation de logiciels de prédiction est de mise pour cette dernière tâche de vérification. Des outils de prédiction *in silico* tels que *Cas-OFFinder*, *Crisflash* ou *CasOT* fourniront les sites les plus susceptibles d'être affectés par des effets hors cible, en fonction de l'ARNg employé et le génome cible à l'étude (151-153). Il restera à amplifier et séquencer ces régions, afin d'attester de la qualité de l'édition génomique des mutants.

## **4.4 Optimisation du taux de rendement de KI**

L'édition génétique des myoblastes a été réalisée à l'ancienne, sans pistes stratégiques pour promouvoir le KI. Advenant que les impuretés ne peuvent être éclaircies, et que l'expérience était à refaire, de plus en plus de stratégies sont en développement afin d'augmenter les taux de rendement de KI. Les pistes sur lesquelles l'on peut jouer varient entre : la promotion de la voie HDR pour la réparation d'une DSB, et favoriser l'utilisation d'un brin ADNss donneur pour la réparation spécifique d'une DSB.

### **4.4.1 Promotion de la voie HDR**

Un premier paramètre consiste à jouer sur un cycle cellulaire actif, à travers la confluence cellulaire des cultures. Afin de garantir la division d'une cellule mère en deux cellules filles identiques, celle-ci exprime une machinerie variée assurant des mécanismes de surveillance qui contrôlent toute l'intégrité du processus (154). On compte entre autres une machinerie certifiant la réplication exacte de l'ADN, immobilisant le cycle cellulaire, et réparant l'ADN fidèlement par la voie HDR s'il advenait à y avoir une erreur (155). Ainsi, en réalisant l'expérience à plus faible confluence (en pleine prolifération), les cellules seront plus enclines à réparer la DSB via le

mécanisme HDR, augmentant les chances d'incorporer un KI (156). De surcroît, il a été démontré que les myoblastes en prolifération (et avant différenciation) possèdent une perméabilité membranaire plus permissive (157, 158). Cela est spécialement vrai au niveau du noyau qui est dans un état désassemblé lors de la division cellulaire, facilitant l'accessibilité du complexe Cas9-ARNg au génome (157).

Une autre méthode pour améliorer les rendements de KI repose dans l'utilisation de molécules précises affectant l'équilibre métabolique entre les deux voies de réparation NHEJ/HDR. L'inhibiteur SCR7, inhibiteur de l'ADN ligase IV impliquée dans la voie de réparation NHEJ, a été établi comme un composé prometteur, favorisant une réparation de la DSB par HDR (159). Conjointement, la molécule RS-1 est un stimulant de l'enzyme RAD51, une recombinase clé impliquée dans la voie de réparation HDR (160). Elle semble impacter davantage la balance en faveur de la voie de réparation HDR que SCR7 (161). Employés en parallèle, ils forment une paire avec un important potentiel pour une édition génomique KI réussie.

#### **4.4.2 Promotion de l'utilisation de la matrice ADNss donneuse**

À un stade plus avancé, des modifications covalentes sur l'endonucléase Cas9 et la matrice ADNss peuvent être éditées minutieusement de sorte à favoriser leur complexation en une unité Cas9(ARNg)-ADNss (162, 163). Le but d'une telle manipulation est de garantir la présence de l'ADNss donneur au site de la DSB. Suite au clivage dans le génome, l'ADNss est techniquement la première matrice avec une région homologue présente au site de coupure, et devrait théoriquement être la première à servir de gabarit pour la reconstruction spécifique (162, 163). À cette fin, la compagnie IDT propose sa formule Alt-R, dont la base moléculaire n'a pas été divulguée, mais qui repose passablement sur ce même principe (164).

#### **4.4.3 Analyse par courbe de fusion à haute de résolution**

De concert, dans l'optique d'accélérer les manipulations expérimentales, une méthode alternative pour la vérification rapide du génotype peut être envisagée. Il s'agit de l'analyse par courbe de fusion à haute résolution (HRM), tirant profit de la température de dénaturation de l'ADN (165). Un fragment d'ADN double brin possède une température de fusion précise, variant en fonction de sa longueur, sa composition nucléotidique et la concentration ionique

environnante (166). Fonctionnant avec un colorant intercalant qui fluoresce fortement lorsque lié à de l'ADN double brin, l'approche examine précisément la séparation des brins en temps réel, corrélant avec une chute du niveau de fluorescence (167). L'analyse HRM permet de reconstituer la séquence des amplicons PCR et confirmer le génotype des clones, accélérant immensément le processus de criblage (168). Son plus grand atout est la capacité de cribler pour tous les types de clones (KO, KI, et non édités) en simultanée; englobant toutes les manipulations actuelles de digestion enzymatique, électrophorèse sur gel et séquençage en une seule et même expérience.

## **4.5 Expansion du modèle cellulaire**

L'obtention du mutant permettra d'approfondir sur les impacts morpho-fonctionnels engendrés par le variant pathogénique au sein du modèle cellulaire. À cet effet, différentes instances du modèle cellulaire seront comparées en parallèle, soit : les myoblastes de type sauvage, les cellules électroporées sans aucune édition génomique acquise, des clones arborant un KO homozygote et hétérozygote, et finalement des clones présentant un KI homozygote et hétérozygote. Toutes ces variantes donneront la meilleure couverture, fournissant des renseignements par rapport aux rôles musculaires de MLIP, les impacts pathologiques, et comment le tout varie en fonction de la zygosité et la nature de la mutation acquise.

L'évaluation morpho-fonctionnelle reposera sur une multitude de points : un examen de la morphologie, prolifération et différenciation cellulaire; puis une investigation intracellulaire portant sur la localisation et l'interactome de MLIP.

### **4.5.1 Morphologie des myoblastes**

La morphologie typique de myoblastes sains est connue. La présence d'une mutation pourrait mener à diverses perturbations qui viendraient affecter cette notion. À cette fin, nous pouvons examiner l'indice de forme, le périmètre et la surface cellulaire pour l'identification d'anormalités morphologiques (169). Par exemple, avec l'introduction de la mutation dans le modèle cellulaire, une apparence plus large pourrait être attendue, résultat d'une division cellulaire ralentie (170). La morphologie des myoblastes fusionnant en myotubes doit aussi être évaluée minutieusement, étant donné qu'ils adoptent une conformation complètement différente de leur état unitaire. Ces

derniers passent au travers d'une restructuration complexe de leur cytoarchitecture : s'alignant étroitement côte à côte, adoptant une forme plus allongée, alignant leur noyau à la même hauteur, puis fusionnant à l'unisson en myotube (171, 172). Des défauts potentiellement observables correspondraient à une forme plus fine, un non-alignement des noyaux, un taux de fusion moindre (172). L'examen morphologique permettrait d'attester de tous ces traits précis.

#### **4.5.2 Prolifération cellulaire**

La vitesse de prolifération peut directement être impactée par l'introduction d'une mutation dans le génome cellulaire. Outre constater un ralentissement dans le rythme de passage des myoblastes ou des observations microscopiques qui suggèrent une cadence réduite de la division cellulaire, des approches plus quantitatives peuvent être employées. La résazurine est excellent indicateur pour des tests de viabilité cellulaire. À son état oxydé, la résazurine a une forte teinte bleutée et sa fluorescence intrinsèque est quasi nulle (173). Cependant, lorsque réduite irréversiblement en résorufine grâce aux propriétés rédox de la chaîne respiratoire de la mitochondrie, elle acquiert alors une vive coloration rosée et commence à fluorescer fortement dans le rouge (173, 174). Cela permet des mesures fluorimétriques exclusives à la résorufine, reflétant viabilité et population cellulaire (173, 174). Dans notre contexte, on pourrait s'attendre à des mesures fluorescentes de moindre intensité pour des cellules mutées, lorsque comparé à des cultures de type sauvage à un même temps donné.

#### **4.5.3 Différenciation cellulaire**

L'évaluation de l'aptitude de différenciation cellulaire constitue une étape cruciale lors d'analyses morpho-fonctionnelles. Ici, les myoblastes vont commencer à fusionner en grandes myofibres multinucléées (myotube) à l'atteinte d'une forte confluence cellulaire (175). La supplémentation de sérum de cheval (~2%) dans le milieu de culture aide à la complétion du processus (175). Il est possible d'attester du bon déroulement de la chose sur plusieurs plans. En plus des traits morphologiques spécifiques aux myotubes (Section 4.5.1), le développement de structures sarcomériques, l'unité contractile musculaire, est une autre conséquence de la différenciation des myoblastes (171). Son observation et le marquage spécifique de l'actine sarcomérique (actinine-alpha) est un bon indicatif de la transition en myotubes (171, 176).

Afin de bien distinguer les états cellulaires, on peut quantifier divers biomarqueurs. L'abondance de filaments de desmine aide à cerner la pureté des cultures de myoblastes, et la détection des facteurs de transcription MyoD et MefC2 signale l'enclenchement du processus de différenciation (177, 178). D'autre part, l'expression de gènes précis, tels que *MYF5*, *MYOG*, *ACTA1*, *MYH1*, *MYH3*, *MYH8*, sont couramment quantifiés pour étudier l'évolution de la différenciation en myotube (179). Les mesures sont idéalement réalisées sur deux fronts : une mesure des ARNm par qPCR pour évaluer l'expression génique, et un marquage immunohistochimique pour attester de la présence des protéines dans la cellule (179). L'évaluation progressive de ces derniers devrait révéler une régulation à la hausse en fonction du temps.

En plus d'utiliser les traits morphologiques (section 4.5.1) comme critères d'évaluation, le profil combiné de tous les biomarqueurs devrait démontrer : une certaine stabilité de la profusion de desmine; un pic d'expression des facteurs de transcription MyoD et MefC2; une régulation positive et croissante des gènes signatures de l'état de myotube; et finalement le développement d'actine sarcomérique. Dans le cadre des analyses impliquant le modèle mutant, des altérations dans les différents profils d'expression génique et/ou le développement de structures sarcomériques disparates pourraient survenir. Tous ces éléments devront être évalués rigoureusement afin d'identifier des défauts s'il y a lieu.

#### **4.5.4 Localisation intracellulaire de MLIP**

Sachant que les transcrits *MLIP* portant l'exon 4 avec la séquence NLS inclue également l'exon 5 et sa mutation non-sens, une certaine baisse et/ou délocalisation des isoformes nucléaires peut être suspectée. Le marquage intracellulaire par immunofluorescence est la méthode de choix pour la vérification de cette supposition, révélant la disposition *in vitro* des cibles moléculaires d'intérêt (180). Cette technique permettra de répondre et consolider plusieurs points : comment est-ce que la balance d'isoformes nucléaires et cytoplasmiques est affectée *in vitro* par la mutation non-sens. Voit-on un important changement avec une plus grande abondance des isoformes cytoplasmiques, suggérant que MLIP ne peut accomplir ses rôles nucléaires? Excluant la question de la localisation, constate-t-on nous une baisse générale *in vitro* des niveaux protéiques de MLIP, comme le suggèrent les expériences de RNA-seq, qPCR et WB?



#### 4.5.4.1 Interactome de MLIP

Par ailleurs, cette approche technique permet également d'explorer l'interactome diversifié de MLIP. En effectuant un marquage avec plusieurs teintes fluorescentes pour différentes entités moléculaires, on est en mesure d'évaluer plusieurs cibles dans un même montage. La colocalisation de deux signaux fluorescents souligne la proximité spatiale entre les deux acteurs marqués, et suggère une potentielle interaction moléculaire positive (181).

Les domaines de liaisons à la LMNA et ISL1 ne sont pas coupés par la mutation non-sens (Figure 4), mais un mauvais repliement protéique peut très bien abolir l'interaction avec ses partenaires moléculaires. Ainsi, on peut évaluer s'il y a colocalisation entre MLIP et la lamine de type A au niveau de la lamina nucléaire, ou encore s'il y a colocalisation au sein du noyau entre MLIP et le facteur de transcription ISL1. Avec l'absence théorique du crochet AT sur les protéines tronquées, on peut évaluer s'il y a superposition entre MLIP et certaines régions de la chromatine. Parallèlement, la réalisation de co-immunoprécipitation de protéines et l'immunoprécipitation de la chromatine permettront de valider toutes ses interactions à l'extérieur de la cellule (182-184). Toute cette collecte d'informations permettra de consolider la perte ou non de certaines des fonctions natives à MLIP.

#### 4.5.5 Expérience de sauvetage

L'exécution de toutes ces expériences va permettre de départager quantitativement les phénotypes normaux et pathologiques des myoblastes. Avec une représentation robuste de chaque état, une tentative de résolution des phénotypes pathologiques est envisagée pour élargir nos connaissances. À cet effet, des expériences de sauvetage à l'aide de composés chimiques permettant la lecture ribosomale des codons de terminaison prématurée seront réalisées. Un médicament tel que l'ataluren, aussi connu sous la dénomination PTC124, est couramment utilisé pour le traitement de la dystrophie musculaire de Duchenne (DMD) (185, 186). Celui-ci permet d'outrepasser des mutations non-sens au sein du gène de la dystrophine, rehaussant les niveaux d'expression de la protéine complète (185). Il a une sélectivité telle, qu'il n'affecte pas les codons normaux de terminaison encodés dans les autres transcrits ARNm (187). Cette médication est bien tolérée de manière générale chez les patients administrés, et permet

un regain substantiel de certaines fonctions des muscles striés, menant à une amélioration nette des symptômes éprouvés (186). En parallèle, l'utilisation de l'ataluren est présentement en train d'être évaluée pour le traitement d'autres désordres génétiques, avec entre autres plusieurs essais cliniques réalisés chez des individus souffrant de fibrose kystique (187, 188). Encore une fois, le composé est passablement bien toléré, mais l'amélioration de symptômes n'est pas aussi évidente que ce qui est observé du côté de la DMD (189).

Dans cette perspective, sachant que l'ataluren est bien établi pour le traitement d'une maladie musculaire, occasionnée par un type de mutation similaire à notre variant, et que son potentiel curatif est activement testé pour expandre son utilisation thérapeutique : il devient très pertinent d'évaluer si le tout est transposable dans le contexte de la myopathie distale du patient. En administrant le composé à notre modèle cellulaire, il sera possible d'en évaluer l'incidence sur les phénotypes pathologiques. Permettra-t-il une guérison complète, majeure ou mineure des symptômes, ou bien n'aura-t-il aucun effet? Si l'on voit des effets positifs, peuvent-ils être apparentés aux effets du mécanisme d'épissage alternatif de compensation observé chez le patient dans l'expérience de LRS? Somme toute, les tentatives d'expansion faites à l'endroit de l'ataluren en font un candidat prometteur à tester. Et advenant des résultats positifs et un essor du projet de recherche, l'enquête serait transposée vers l'étude d'un modèle murin mutant.

## 5 – Conclusion et perspectives

Dans son ensemble, énormément de progrès a été réalisé dans ce travail d'envergure, en même temps d'en être encore à ses débuts. La résolution du cas clinique et l'apport d'un diagnostic moléculaire définitif pour le patient atteint d'une maladie neuromusculaire rare ont été accomplis grâce au RNA-seq : identifiant le variant pathogénique pour la première fois, et montrant une expression génique de *MLIP* et *LMNA* affectée. La preuve de l'implication du variant dans la pathologie a été subséquemment démontrée avec les divers résultats capturés, reflétant les multiples perturbations transcriptomiques présentes chez le patient. Le LRS s'est avéré être un outil remarquable pour la caractérisation efficace du transcriptome complexe de *MLIP*, révélant un phénomène d'épissage alternatif compensatoire cherchant à contourner la production de

transcrits ARN arborant la mutation non-sens. Les percées faites dans ce projet témoignent une fois de plus du pouvoir diagnostique des applications NGS, surtout du RNA-seq, pour la résolution de cas clinique complexe. Elles mettent également à l'avant le pouvoir de caractérisation transcriptomique de la couverture à longues lectures, offertes par les techniques de séquençage de troisième génération. Ces trouvailles vont contribuer à l'avancée des notions dans la littérature entourant les myopathies : complétant un peu plus la classification de ces dernières, et pouvant notamment améliorer le taux de diagnostic clinique de maladies neuromusculaires, au travers de l'intégration de *MLIP* dans les panels de gènes testant pour ces groupes de pathologies. Elles illustrent également l'implication pathogénique de *MLIP* dans des cas à apparition tardive pour la première fois et de ce fait, mettent l'accent sur l'hétérogénéité clinique qui peut être rencontrée, fort probablement dû à l'architecture transcriptomique complexe de ce dernier.

D'autre part, la modélisation cellulaire du variant par CRISPR/Cas9 représente un modèle fiable pour une évaluation morpho-fonctionnelle plus approfondie des impacts engendrés par la mutation. Cette extension de l'étude permettra de récolter un supplément d'informations précieux, renseignant sur les rôles musculaires normaux et pathologiques de *MLIP*, embellissant davantage la littérature. Une meilleure compréhension de son mécanisme pathologique pourra paver la route pour l'établissement d'éventuelles stratégies thérapeutiques, notamment au travers des essais d'expériences de sauvetage qui seront testées.

Finalement, englobant toute l'enquête faite dans ce cas clinique précis, on est en droit d'envisager comment tout le processus pourrait être implanté comme filière de travail pour les investigations futures à venir au sein du laboratoire. L'optimisation des étapes de diagnostic, identification et investigation aura pour finalité d'accélérer l'établissement du lien corrélationnel entre un gène fautif et un trouble neuromusculaire rare. Proposer une prise en charge rapide et appropriée au patient, et réduire sa détresse psychologique en lui apportant un élément de réponse à sa situation, demeurent les priorités premières qui dirigent toute l'étendue de ce travail.

## 6 – Références bibliographiques

1. Gilbreath HR, Castro D, Iannaccone ST. Congenital myopathies and muscular dystrophies. *Neurol Clin.* 2014;32(3):689-703, viii.
2. Ahmed ST, Craven L, Russell OM, Turnbull DM, Vincent AE. Diagnosis and Treatment of Mitochondrial Myopathies. *Neurotherapeutics.* 2018;15(4):943-53.
3. Lilleker JB, Keh YS, Roncaroli F, Sharma R, Roberts M. Metabolic myopathies: a practical approach. *Pract Neurol.* 2018;18(1):14-26.
4. Gonzalez-Jamett AM, Bevilacqua JA, Diaz AMC. Hereditary Myopathies. In: Sakuma K, editor. *Muscle Cell and Tissue - Current Status of Research Field.* London, UK: IntechOpen; 2018.
5. Goto Y, Nonaka I, Horai S. A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature.* 1990;348(6302):651-3.
6. de Visser M. Late-onset myopathies: clinical features and diagnosis. *Acta Myol.* 2020;39(4):235-44.
7. Pfeiffer G, Elliott HR, Griffin H, Barresi R, Miller J, Marsh J, et al. Titin mutation segregates with hereditary myopathy with early respiratory failure. *Brain.* 2012;135(Pt 6):1695-713.
8. Finsterer J, Frank M. Potential causes of sudden cardiac death in nemaline myopathy. *Ital J Pediatr.* 2015;41:67.
9. Finsterer J, Stollberger C. Review of Cardiac Disease in Nemaline Myopathy. *Pediatr Neurol.* 2015;53(6):473-7.
10. Vill K, Blaschek A, Glaser D, Kuhn M, Haack T, Alhaddad B, et al. Early-Onset Myopathies: Clinical Findings, Prevalence of Subgroups and Diagnostic Approach in a Single Neuromuscular Referral Center in Germany. *J Neuromuscul Dis.* 2017;4(4):315-25.
11. McGrath ER, Doughty CT, Amato AA. Autoimmune Myopathies: Updates on Evaluation and Treatment. *Neurotherapeutics.* 2018;15(4):976-94.
12. Schmidt J. Current Classification and Management of Inflammatory Myopathies. *J Neuromuscul Dis.* 2018;5(2):109-29.
13. Valiyil R, Christopher-Stine L. Drug-related myopathies of which the clinician should be aware. *Curr Rheumatol Rep.* 2010;12(3):213-20.
14. Mangaraj S, Sethy G. Hoffman's syndrome - A rare facet of hypothyroid myopathy. *J Neurosci Rural Pract.* 2014;5(4):447-8.
15. Robinson-Papp J, Simpson DM. Neuromuscular diseases associated with HIV-1 infection. *Muscle Nerve.* 2009;40(6):1043-53.
16. Elagami MM, Ghrewati M, Khaddash I, Melki G. An Unusual Presentation of Influenza-Induced Myositis. *Cureus.* 2021;13(2):e13196.
17. Mills KR. The basics of electromyography. *J Neurol Neurosurg Psychiatry.* 2005;76 Suppl 2(Suppl 2):ii32-5.
18. Zipursky JS, Redelmeier DA. Muscle fasciculation detected by ECG. *BMJ Case Rep.* 2017;2017.
19. Pillen S, Arts IM, Zwartz MJ. Muscle ultrasound in neuromuscular disorders. *Muscle Nerve.* 2008;37(6):679-93.
20. Mohamed AA, Arifi AA, Omran A. The basics of echocardiography. *J Saudi Heart Assoc.* 2010;22(2):71-6.

21. Venturelli N, Tordjman M, Ammar A, Chetrit A, Renault V, Carlier RY. Contribution of muscle MRI for diagnosis of myopathy. *Rev Neurol (Paris)*. 2023;179(1-2):61-80.
22. Wolf PL. Abnormalities in serum enzymes in skeletal muscle diseases. *Am J Clin Pathol*. 1991;95(3):293-6.
23. Lilleng H, Abeler K, Johnsen SH, Stensland E, Loseth S, Jorde R, et al. Variation of serum creatine kinase (CK) levels and prevalence of persistent hyperCKemia in a Norwegian normal population. The Tromso Study. *Neuromuscul Disord*. 2011;21(7):494-500.
24. Joyce NC, Oskarsson B, Jin LW. Muscle biopsy evaluation in neuromuscular disorders. *Phys Med Rehabil Clin N Am*. 2012;23(3):609-31.
25. Talbot J, Maves L. Skeletal muscle fiber type: using insights from muscle developmental biology to dissect targets for susceptibility and resistance to muscle disease. *Wiley Interdiscip Rev Dev Biol*. 2016;5(4):518-34.
26. Nix JS, Moore SA. What Every Neuropathologist Needs to Know: The Muscle Biopsy. *J Neuropathol Exp Neurol*. 2020;79(7):719-33.
27. Vincent AE, Ng YS, White K, Davey T, Mannella C, Falkous G, et al. The Spectrum of Mitochondrial Ultrastructural Defects in Mitochondrial Myopathy. *Sci Rep*. 2016;6:30610.
28. Potapova NA. Nonsense Mutations in Eukaryotes. *Biochemistry (Mosc)*. 2022;87(5):400-12.
29. Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol*. 2009;16(2):107-13.
30. Lecker SH, Goldberg AL, Mitch WE. Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J Am Soc Nephrol*. 2006;17(7):1807-19.
31. Saier MH, Jr. Understanding the Genetic Code. *J Bacteriol*. 2019;201(15).
32. Luo LF. The degeneracy rule of genetic code. *Orig Life Evol Biosph*. 1988;18(1-2):65-70.
33. Zeng Z, Bromberg Y. Predicting Functional Effects of Synonymous Variants: A Systematic Review and Perspectives. *Front Genet*. 2019;10:914.
34. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463-7.
35. Cheng C, Xiao P. Evaluation of the correctable decoding sequencing as a new powerful strategy for DNA sequencing. *Life Sci Alliance*. 2022;5(8).
36. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. Fluorescence detection in automated DNA sequence analysis. *Nature*. 1986;321(6071):674-9.
37. Jorgenson JW, Lukacs KD. Free-zone electrophoresis in glass capillaries. *Clin Chem*. 1981;27(9):1551-3.
38. Kasper TJ, Melera M, Gozel P, Brownlee RG. Separation and detection of DNA by capillary electrophoresis. *J Chromatogr*. 1988;458:303-12.
39. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001;291(5507):1304-51.
40. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed*. 2013;98(6):236-8.
41. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;437(7057):376-80.

42. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456(7218):53-9.
43. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010;11(1):31-46.
44. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet*. 2008;24(3):133-41.
45. Ju J, Kim DH, Bi L, Meng Q, Bai X, Li Z, et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc Natl Acad Sci U S A*. 2006;103(52):19635-40.
46. Illumina. Illumina Sequencing Technology: Illumina Inc; 2010 [updated 2010 - 10 - 11; cited 2023 - 03 - 05]. Available from: [https://www.illumina.com/documents/products/techspotlights/techspotlight\\_sequencing.pdf](https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf).
47. Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res*. 2006;34(3):e22.
48. Frank M, Prenzler A, Eils R, Graf von der Schulenburg JM. Genome sequencing: a systematic review of health economic evidence. *Health Econ Rev*. 2013;3(1):29.
49. Mashayekhi F, Ronaghi M. Analysis of read length limiting factors in Pyrosequencing chemistry. *Anal Biochem*. 2007;363(2):275-87.
50. Bean LJH, Funke B, Carlston CM, Gannon JL, Kantarci S, Krock BL, et al. Diagnostic gene sequencing panels: from design to report-a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2020;22(3):453-61.
51. Ng KWP, Chin HL, Chin AXY, Goh DL. Using gene panels in the diagnosis of neuromuscular disorders: A mini-review. *Front Neurol*. 2022;13:997551.
52. Bevilacqua J, Hesse A, Cormier B, Davey J, Patel D, Shankar K, et al. Clinical utility of a 377 gene custom next-generation sequencing epilepsy panel. *J Genet*. 2017;96(4):681-5.
53. Narod SA, Salmena L. BRCA1 and BRCA2 mutations and breast cancer. *Discov Med*. 2011;12(66):445-53.
54. Marian AJ. Sequencing your genome: what does it mean? *Methodist Debakey Cardiovasc J*. 2014;10(1):3-6.
55. Seaby EG, Pengelly RJ, Ennis S. Exome sequencing explained: a practical guide to its clinical application. *Brief Funct Genomics*. 2016;15(5):374-84.
56. Niguidula N, Alamillo C, Shahmirzadi Mowlavi L, Powis Z, Cohen JS, Farwell Hagman KD. Clinical whole-exome sequencing results impact medical management. *Mol Genet Genomic Med*. 2018;6(6):1068-78.
57. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med*. 2017;9(386).
58. Scacheri CA, Scacheri PC. Mutations in the noncoding genome. *Curr Opin Pediatr*. 2015;27(6):659-64.
59. Karnuta JM, Scacheri PC. Enhancers: bridging the gap between gene control and human disease. *Hum Mol Genet*. 2018;27(R2):R219-R27.

60. Geyer PK, Green MM, Corces VG. Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in *Drosophila*. *EMBO J*. 1990;9(7):2247-56.
61. Vaz-Drago R, Custodio N, Carmo-Fonseca M. Deep intronic mutations and human disease. *Hum Genet*. 2017;136(9):1093-111.
62. Hollox EJ, Zuccherato LW, Tucci S. Genome structural variation in human evolution. *Trends Genet*. 2022;38(1):45-58.
63. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet*. 2016;17(4):224-38.
64. Lauer S, Gresham D. An evolving view of copy number variants. *Curr Genet*. 2019;65(6):1287-95.
65. Shi H, Zhou Y, Jia E, Pan M, Bai Y, Ge Q. Bias in RNA-seq Library Preparation: Current Challenges and Solutions. *Biomed Res Int*. 2021;2021:6647597.
66. Nicolau S, Choquet K, Bareke E, Shao YH, Brais B, O'Ferrall EK, et al. A Molecular Diagnosis of LGMDR1 Established by RNA Sequencing. *Can J Neurol Sci*. 2021;48(2):293-6.
67. Hamanaka K, Miyatake S, Koshimizu E, Tsurusaki Y, Mitsuhashi S, Iwama K, et al. RNA sequencing solved the most common but unrecognized NEB pathogenic variant in Japanese nemaline myopathy. *Genet Med*. 2019;21(7):1629-38.
68. Taniue K, Akimitsu N. Fusion Genes and RNAs in Cancer Development. *Noncoding RNA*. 2021;7(1).
69. Latysheva NS, Babu MM. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Res*. 2016;44(10):4487-503.
70. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*. 2015;112(17):5473-8.
71. Tetreault M, Bareke E, Nadaf J, Alirezaie N, Majewski J. Whole-exome sequencing as a diagnostic tool: current challenges and future opportunities. *Expert Rev Mol Diagn*. 2015;15(6):749-60.
72. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*. 2018;36(4):338-45.
73. Thompson JF, Milos PM. The properties and applications of single-molecule DNA sequencing. *Genome Biol*. 2011;12(2):217.
74. Van der Verren SE, Van Gerven N, Jonckheere W, Hambley R, Singh P, Kilgour J, et al. A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. *Nat Biotechnol*. 2020;38(12):1415-20.
75. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A*. 1996;93(24):13770-3.
76. Saha KK, Drndic M, Nikolic BK. DNA base-specific modulation of microampere transverse edge currents through a metallic graphene nanoribbon with a nanopore. *Nano Lett*. 2012;12(1):50-5.
77. Ding H, Anastopoulos I, Bailey ADt, Stuart J, Paten B. Towards inferring nanopore sequencing ionic currents from nucleotide chemical structures. *Nat Commun*. 2021;12(1):6545.

78. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol.* 2021;39(11):1348-65.
79. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol.* 2012;30(4):349-53.
80. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. *PLoS One.* 2021;16(10):e0257521.
81. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends Genet.* 2018;34(9):666-81.
82. Derrington IM, Butler TZ, Collins MD, Manrao E, Pavlenok M, Niederweis M, et al. Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci U S A.* 2010;107(37):16060-5.
83. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 2019;20(1):129.
84. Oxford-Nanopore-Technologies. How nanopore sequencing works: Oxford Nanopore Technologies; 2021 [updated 2023 - 02 - 03; cited 2023 - 03 - 05]. Available from: <https://nanoporetech.com/support/how-it-works#fullVideo&>.
85. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 2007;8(7):R143.
86. Leonardi T, Leger A. Nanopore RNA Sequencing Analysis. *Methods Mol Biol.* 2021;2284:569-78.
87. Tang AD, Soulette CM, van Baren MJ, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat Commun.* 2020;11(1):1438.
88. Liu Y, Rosikiewicz W, Pan Z, Jillette N, Wang P, Taghbalout A, et al. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.* 2021;22(1):295.
89. Barros-Silva D, Marques CJ, Henrique R, Jeronimo C. Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications. *Genes (Basel).* 2018;9(9).
90. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 2020;21(1):30.
91. Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, et al. Telomere-to-telomere assembly of a complete human X chromosome. *Nature.* 2020;585(7823):79-84.
92. Bahlo M, Bennett MF, Degorski P, Tankard RM, Delatycki MB, Lockhart PJ. Recent advances in the detection of repeat expansions with short-read next-generation sequencing. *F1000Res.* 2018;7.
93. Bruels CC, Littel HR, Daugherty AL, Stafki S, Estrella EA, McGaughy ES, et al. Diagnostic capabilities of nanopore long-read sequencing in muscular dystrophy. *Ann Clin Transl Neurol.* 2022;9(8):1302-9.
94. Maricelli JW, Lu QL, Lin DC, Rodgers BD. Trendelenburg-Like Gait, Instability and Altered Step Patterns in a Mouse Model for Limb Girdle Muscular Dystrophy 2i. *PLoS One.* 2016;11(9):e0161984.
95. Ahmady E, Deeke SA, Rabaa S, Kouri L, Kenney L, Stewart AF, et al. Identification of a novel muscle A-type lamin-interacting protein (MLIP). *J Biol Chem.* 2011;286(22):19702-13.



96. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
97. Ahmady E, Blais A, Burgon PG. Muscle Enriched Lamin Interacting Protein (Mlip) Binds Chromatin and Is Required for Myoblast Differentiation. *Cells*. 2021;10(3).
98. Moretti A, Lam J, Evans SM, Laugwitz KL. Biology of Isl1+ cardiac progenitor cells in development and disease. *Cell Mol Life Sci*. 2007;64(6):674-82.
99. Cattin ME, Wang J, Weldrick JJ, Roeske CL, Mak E, Thorn SL, et al. Deletion of MLIP (muscle-enriched A-type lamin-interacting protein) leads to cardiac hyperactivation of Akt/mammalian target of rapamycin (mTOR) and impaired cardiac adaptation. *J Biol Chem*. 2015;290(44):26699-714.
100. Huang ZP, Kataoka M, Chen J, Wu G, Ding J, Nie M, et al. Cardiomyocyte-enriched protein CIP protects against pathophysiological stresses and regulates cardiac homeostasis. *J Clin Invest*. 2015;125(11):4122-34.
101. Lallemand-Breitenbach V, de The H. PML nuclear bodies. *Cold Spring Harb Perspect Biol*. 2010;2(5):a000661.
102. Lopes Abath Neto O, Medne L, Donkervoort S, Rodriguez-Garcia ME, Bolduc V, Hu Y, et al. MLIP causes recessive myopathy with rhabdomyolysis, myalgia and baseline elevated serum creatine kinase. *Brain*. 2021;144(9):2722-31.
103. Salzer-Sheelo L, Fellner A, Orenstein N, Bazak L, Lev-El Halabi N, Daue M, et al. Biallelic truncating variants in the muscular A-type lamin-interacting protein (MLIP) gene cause myopathy with hyperCKemia. *Eur J Neurol*. 2022;29(4):1174-80.
104. Mezreani J, Audet S, Martin F, Charbonneau J, Triassi V, Bareke E, et al. Novel homozygous nonsense mutation of MLIP and compensatory alternative splicing. *NPJ Genom Med*. 2022;7(1):36.
105. Bermejo-Guerrero L, Arteché-Lopez A, de Fuenmayor Fernandez de la Hoz C, Hernandez-Lain A, Martin MA, Dominguez-Gonzalez C. A novel MLIP truncating variant in an 80-year-old patient with late-onset progressive weakness. *Brain*. 2022;145(10):e99-e102.
106. Cattin ME, Deeke SA, Dick SA, Verret-Borsos ZJA, Tennakoon G, Gupta R, et al. Expression of murine muscle-enriched A-type lamin-interacting protein (MLIP) is regulated by tissue-specific alternative transcription start sites. *J Biol Chem*. 2018;293(51):19761-70.
107. Huang ZP, Young Seok H, Zhou B, Chen J, Chen JF, Tao Y, et al. CIP, a cardiac Isl1-interacting protein, represses cardiomyocyte hypertrophy. *Circ Res*. 2012;110(6):818-30.
108. Hetzer MW. The nuclear envelope. *Cold Spring Harb Perspect Biol*. 2010;2(3):a000539.
109. Salina D, Bodoor K, Enarson P, Raharjo WH, Burke B. Nuclear envelope dynamics. *Biochem Cell Biol*. 2001;79(5):533-42.
110. Burke B, Stewart CL. The nuclear lamins: flexibility in function. *Nat Rev Mol Cell Biol*. 2013;14(1):13-24.
111. Lin F, Worman HJ. Structural organization of the human gene encoding nuclear lamin A and nuclear lamin C. *J Biol Chem*. 1993;268(22):16321-6.
112. Lin F, Worman HJ. Structural organization of the human gene (LMNB1) encoding nuclear lamin B1. *Genomics*. 1995;27(2):230-6.
113. Biamonti G, Giacca M, Perini G, Contreas G, Zentilin L, Weighardt F, et al. The gene for a novel human lamin maps at a highly transcribed locus of chromosome 19 which replicates at the onset of S-phase. *Mol Cell Biol*. 1992;12(8):3499-506.

114. Guerreiro I, Kind J. Spatial chromatin organization and gene regulation at the nuclear lamina. *Curr Opin Genet Dev.* 2019;55:19-25.
115. Turgay Y, Eibauer M, Goldman AE, Shimi T, Khayat M, Ben-Harush K, et al. The molecular architecture of lamins in somatic cells. *Nature.* 2017;543(7644):261-4.
116. Broers JL, Ramaekers FC, Bonne G, Yaou RB, Hutchison CJ. Nuclear lamins: laminopathies and their role in premature ageing. *Physiol Rev.* 2006;86(3):967-1008.
117. Bonne G, Di Barletta MR, Varnous S, Becane HM, Hammouda EH, Merlini L, et al. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nat Genet.* 1999;21(3):285-8.
118. Sullivan T, Escalante-Alcalde D, Bhatt H, Anver M, Bhat N, Nagashima K, et al. Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J Cell Biol.* 1999;147(5):913-20.
119. Fatkin D, MacRae C, Sasaki T, Wolff MR, Porcu M, Frenneaux M, et al. Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction-system disease. *N Engl J Med.* 1999;341(23):1715-24.
120. Kayvanpour E, Sedaghat-Hamedani F, Amr A, Lai A, Haas J, Holzer DB, et al. Genotype-phenotype associations in dilated cardiomyopathy: meta-analysis on more than 8000 individuals. *Clin Res Cardiol.* 2017;106(2):127-39.
121. Frock RL, Kudlow BA, Evans AM, Jameson SA, Hauschka SD, Kennedy BK. Lamin A/C and emerin are critical for skeletal muscle satellite cell differentiation. *Genes Dev.* 2006;20(4):486-500.
122. Burattini S, Ferri P, Battistelli M, Curci R, Luchetti F, Falcieri E. C2C12 murine myoblasts as a model of skeletal muscle development: morpho-functional characterization. *Eur J Histochem.* 2004;48(3):223-33.
123. Sharples AP, Stewart CE. Myoblast models of skeletal muscle hypertrophy and atrophy. *Curr Opin Clin Nutr Metab Care.* 2011;14(3):230-6.
124. Thurner M, Asim F, Garczarczyk-Asim D, Janke K, Deutsch M, Margreiter E, et al. Development of an in vitro potency assay for human skeletal muscle derived cells. *PLoS One.* 2018;13(3):e0194561.
125. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science.* 2010;327(5962):167-70.
126. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science.* 2014;346(6213):1258096.
127. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature.* 2014;513(7519):569-73.
128. Doudna JA. CRISPR Systems: Doudna Lab; 2018 [updated 2018 - 03 - 12; cited 2023 - 03 - 05]. Available from: [https://doudnalab.org/research\\_areas/crispr-systems/](https://doudnalab.org/research_areas/crispr-systems/).
129. Lino CA, Harper JC, Carney JP, Timlin JA. Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv.* 2018;25(1):1234-57.
130. Duan L, Ouyang K, Xu X, Xu L, Wen C, Zhou X, et al. Nanoparticle Delivery of CRISPR/Cas9 for Genome Editing. *Front Genet.* 2021;12:673286.
131. Petersen BL, Moller SR, Mravec J, Jorgensen B, Christensen M, Liu Y, et al. Improved CRISPR/Cas9 gene editing by fluorescence activated cell sorting of green fluorescence protein tagged protoplasts. *BMC Biotechnol.* 2019;19(1):36.

132. Elaswad A, Khalil K, Cline D, Page-McCaw P, Chen W, Michel M, et al. Microinjection of CRISPR/Cas9 Protein into Channel Catfish, *Ictalurus punctatus*, Embryos for Gene Editing. *J Vis Exp*. 2018(131).
133. Laustsen A, Bak RO. Electroporation-Based CRISPR/Cas9 Gene Editing Using Cas9 Protein and Chemically Modified sgRNAs. *Methods Mol Biol*. 2019;1961:127-34.
134. Xu CL, Ruan MZC, Mahajan VB, Tsang SH. Viral Delivery Systems for CRISPR. *Viruses*. 2019;11(1).
135. Zaboikin M, Zaboikina T, Freter C, Srinivasakumar N. Non-Homologous End Joining and Homology Directed DNA Repair Frequency of Double-Stranded Breaks Introduced by Genome Editing Reagents. *PLoS One*. 2017;12(1):e0169931.
136. Mao Z, Bozzella M, Seluanov A, Gorbunova V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)*. 2008;7(10):1765-71.
137. Rodgers K, McVey M. Error-Prone Repair of DNA Double-Strand Breaks. *J Cell Physiol*. 2016;231(1):15-24.
138. Guo T, Feng YL, Xiao JJ, Liu Q, Sun XN, Xiang JF, et al. Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing. *Genome Biol*. 2018;19(1):170.
139. Janssen JM, Chen X, Liu J, Goncalves M. The Chromatin Structure of CRISPR-Cas9 Target DNA Controls the Balance between Mutagenic and Homology-Directed Gene-Editing Events. *Mol Ther Nucleic Acids*. 2019;16:141-54.
140. Ranawakage DC, Okada K, Sugio K, Kawaguchi Y, Kuninobu-Bonkohara Y, Takada T, et al. Efficient CRISPR-Cas9-Mediated Knock-In of Composite Tags in Zebrafish Using Long ssDNA as a Donor. *Front Cell Dev Biol*. 2020;8:598634.
141. Schubert MS, Thommandru B, Woodley J, Turk R, Yan S, Kurgan G, et al. Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Sci Rep*. 2021;11(1):19482.
142. Zhao F, Kim W, Kloeber JA, Lou Z. DNA end resection and its role in DNA replication and DSB repair choice in mammalian cells. *Exp Mol Med*. 2020;52(10):1705-14.
143. Sentmanat MF, Peters ST, Florian CP, Connelly JP, Pruett-Miller SM. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci Rep*. 2018;8(1):888.
144. Juan-Colas J, Hitchcock IS, Coles M, Johnson S, Krauss TF. Quantifying single-cell secretion in real time using resonant hyperspectral imaging. *Proc Natl Acad Sci U S A*. 2018;115(52):13204-9.
145. Invitrogen. GeneArt™ Precision gRNA Synthesis Kit, USER GUIDE: Thermo Fisher Scientific Inc; 2016 [updated 2016 - 06 - 22; cited 2023 - 03 - 05]. Available from: [https://tools.thermofisher.com/content/sfs/manuals/geneart\\_precision\\_gRNA\\_synthesis\\_kit\\_man.pdf](https://tools.thermofisher.com/content/sfs/manuals/geneart_precision_gRNA_synthesis_kit_man.pdf).
146. Xu Q, Morgan RD, Roberts RJ, Blaser MJ. Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proc Natl Acad Sci U S A*. 2000;97(17):9671-6.
147. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. *Genome Biol*. 2004;5(10):R74.
148. Maggi L, Carboni N, Bernasconi P. Skeletal Muscle Laminopathies: A Review of Clinical and Molecular Features. *Cells*. 2016;5(3).

149. Crasto S, My I, Di Pasquale E. The Broad Spectrum of LMNA Cardiac Diseases: From Molecular Mechanisms to Clinical Phenotype. *Front Physiol.* 2020;11:761.
150. Iwahara N, Hisahara S, Hayashi T, Kawamata J, Shimohama S. A novel lamin A/C gene mutation causing spinal muscular atrophy phenotype with cardiac involvement: report of one case. *BMC Neurol.* 2015;15:13.
151. Bae S, Park J, Kim JS. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics.* 2014;30(10):1473-5.
152. Jacquin ALS, Odom DT, Lukk M. Crisflash: open-source software to generate CRISPR guide RNAs against genomes annotated with individual variation. *Bioinformatics.* 2019;35(17):3146-7.
153. Xiao A, Cheng Z, Kong L, Zhu Z, Lin S, Gao G, et al. CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics.* 2014;30(8):1180-2.
154. Barnum KJ, O'Connell MJ. Cell cycle regulation by checkpoints. *Methods Mol Biol.* 2014;1170:29-40.
155. Hustedt N, Gasser SM, Shimada K. Replication checkpoint: tuning and coordination of replication forks in s phase. *Genes (Basel).* 2013;4(3):388-434.
156. Liu M, Rehman S, Tang X, Gu K, Fan Q, Chen D, et al. Methodologies for Improving HDR Efficiency. *Front Genet.* 2018;9:691.
157. Goullee H, Taylor RL, Forrest ARR, Laing NG, Ravenscroft G, Clayton JS. Improved CRISPR/Cas9 gene editing in primary human myoblasts using low confluency cultures on Matrigel. *Skelet Muscle.* 2021;11(1):23.
158. Jackson MF, Hoversten KE, Powers JM, Trobridge GD, Rodgers BD. Genetic manipulation of myoblasts and a novel primary myosatellite cell culture system: comparing and optimizing approaches. *FEBS J.* 2013;280(3):827-39.
159. Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol.* 2015;33(5):543-8.
160. Lee AR, Park JH, Shim SH, Hong K, La H, Park KS, et al. Genome stabilization by RAD51-stimulatory compound 1 enhances efficiency of somatic cell nuclear transfer-mediated reprogramming and full-term development of cloned mouse embryos. *Cell Prolif.* 2021;54(7):e13059.
161. Song J, Yang D, Xu J, Zhu T, Chen YE, Zhang J. RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat Commun.* 2016;7:10548.
162. Savic N, Ringnalda FC, Lindsay H, Berk C, Bargsten K, Li Y, et al. Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *Elife.* 2018;7.
163. Aird EJ, Lovendahl KN, St Martin A, Harris RS, Gordon WR. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun Biol.* 2018;1:54.
164. Usher I, Ligammari L, Ahrabi S, Hepburn E, Connolly C, Bond GL, et al. Optimizing CRISPR/Cas9 Editing of Repetitive Single Nucleotide Variants. *Front Genome Ed.* 2022;4:932434.
165. Erali M, Wittwer CT. High resolution melting analysis for gene scanning. *Methods.* 2010;50(4):250-61.
166. Khandelwal G, Bhyravabhotla J. A phenomenological model for predicting melting temperatures of DNA sequences. *PLoS One.* 2010;5(8):e12433.

167. Slomka M, Sobalska-Kwapis M, Wachulec M, Bartosz G, Strapagiel D. High Resolution Melting (HRM) for High-Throughput Genotyping-Limitations and Caveats in Practical Case Studies. *Int J Mol Sci.* 2017;18(11).
168. Samarut E, Lissouba A, Drapeau P. A simplified method for identifying early CRISPR-induced indels in zebrafish embryos using High Resolution Melting analysis. *BMC Genomics.* 2016;17:547.
169. Bruyere C, Versaevel M, Mohammed D, Alaimo L, Luciano M, Vercruyse E, et al. Actomyosin contractility scales with myoblast elongation and enhances differentiation through YAP nuclear export. *Sci Rep.* 2019;9(1):15565.
170. Su TT, O'Farrell PH. Size control: cell proliferation does not equal growth. *Curr Biol.* 1998;8(19):R687-9.
171. Berendse M, Grounds MD, Lloyd CM. Myoblast structure affects subsequent skeletal myotube morphology and sarcomere assembly. *Exp Cell Res.* 2003;291(2):435-50.
172. Barro M, Carnac G, Flavier S, Mercier J, Vassetzky Y, Laoudj-Chenivesse D. Myoblasts from affected and non-affected FSHD muscles exhibit morphological differentiation defects. *J Cell Mol Med.* 2010;14(1-2):275-89.
173. Prabst K, Engelhardt H, Ringgeler S, Hubner H. Basic Colorimetric Proliferation Assays: MTT, WST, and Resazurin. *Methods Mol Biol.* 2017;1601:1-17.
174. Gong X, Liang Z, Yang Y, Liu H, Ji J, Fan Y. A resazurin-based, nondestructive assay for monitoring cell proliferation during a scaffold-based 3D culture process. *Regen Biomater.* 2020;7(3):271-81.
175. Cheng CS, El-Abd Y, Bui K, Hyun YE, Hughes RH, Kraus WE, et al. Conditions that promote primary human skeletal myoblast culture and muscle differentiation in vitro. *Am J Physiol Cell Physiol.* 2014;306(4):C385-95.
176. Sjoblom B, Salmazo A, Djinic-Carugo K. Alpha-actinin structure and regulation. *Cell Mol Life Sci.* 2008;65(17):2688-701.
177. Tabakov VY, Zinov'eva OE, Voskresenskaya ON, Skoblov MY. Isolation and Characterization of Human Myoblast Culture In Vitro for Technologies of Cell and Gene Therapy of Skeletal Muscle Pathologies. *Bull Exp Biol Med.* 2018;164(4):536-42.
178. Owens J, Moreira K, Bain G. Characterization of primary human skeletal muscle cells from multiple commercial sources. *In Vitro Cell Dev Biol Anim.* 2013;49(9):695-705.
179. Stern-Straeter J, Bonaterra GA, Kassner SS, Zugel S, Hormann K, Kinscherf R, et al. Characterization of human myoblast differentiation for tissue-engineering purposes by quantitative gene expression analysis. *J Tissue Eng Regen Med.* 2011;5(8):e197-206.
180. Im K, Mareninov S, Diaz MFP, Yong WH. An Introduction to Performing Immunofluorescence Staining. *Methods Mol Biol.* 2019;1897:299-311.
181. Adler J, Parmryd I. Colocalization analysis in fluorescence microscopy. *Methods Mol Biol.* 2013;931:97-109.
182. Lee C. Coimmunoprecipitation assay. *Methods Mol Biol.* 2007;362:401-6.
183. Rao VS, Srinivas K, Sujini GN, Kumar GN. Protein-protein interaction detection: methods and analysis. *Int J Proteomics.* 2014;2014:147648.
184. Das PM, Ramachandran K, vanWert J, Singal R. Chromatin immunoprecipitation assay. *Biotechniques.* 2004;37(6):961-9.

185. Sheikh O, Yokota T. Developing DMD therapeutics: a review of the effectiveness of small molecules, stop-codon readthrough, dystrophin gene replacement, and exon-skipping therapies. *Expert Opin Investig Drugs*. 2021;30(2):167-76.
186. McDonald CM, Campbell C, Torricelli RE, Finkel RS, Flanigan KM, Goemans N, et al. Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017;390(10101):1489-98.
187. Welch EM, Barton ER, Zhuo J, Tomizawa Y, Friesen WJ, Trifillis P, et al. PTC124 targets genetic disorders caused by nonsense mutations. *Nature*. 2007;447(7140):87-91.
188. Pibiri I, Lentini L, Melfi R, Gallucci G, Pace A, Spinello A, et al. Enhancement of premature stop codon readthrough in the CFTR gene by Ataluren (PTC124) derivatives. *Eur J Med Chem*. 2015;101:236-44.
189. Sermet-Gaudelus I, Boeck KD, Casimir GJ, Vermeulen F, Leal T, Mogenet A, et al. Ataluren (PTC124) induces cystic fibrosis transmembrane conductance regulator protein expression and activity in children with nonsense mutation cystic fibrosis. *Am J Respir Crit Care Med*. 2010;182(10):1262-72.

## **7 – Annexe**

Ci-joint la fin du manuscrit de l'article présenté au début du Chapitre 3.

### **Methods**

#### **Patient**

The proband was evaluated by an experienced neurologist (E.O.) and had a muscle biopsy as part of his clinical workup. The McGill University Health Centre Research Ethics Board approved the study. The proband signed an informed consent authorizing genetic analysis in a research setting.

#### **RNA-Sequencing**

RNA was extracted from muscle biopsies using TRIzol. The sequencing library was prepared using the TruSeq stranded mRNA library preparation kit and sequenced on an Illumina HiSeq 2500 using 125bp paired-end reads. Alignment was performed using STAR against a reference genome (Hg19) before variant calling with GATK (v.3.7) <sup>21,22</sup>. VCF annotation utilized ANNOVAR databases and custom scripts <sup>23</sup>. Read counts were obtained using featureCounts prior to DESeq differential expression quantification. EnhancedVolcano was used to produce the plot <sup>24</sup>, DEXseq enabled the exon usage analysis <sup>25</sup>, and gene enrichment analysis was performed using GSEA <sup>10-12,26</sup>.

#### **Reverse Transcription Polymerase Chain Reaction**

cDNA was obtained following reverse transcription with SuperScript Vilo mix (Invitrogen, 11754050), and amplified with EasyTaq polymerase (TransGen Biotech, AP111-01) using custom

primers GGGAATTCGAAGCAAACAAA; GGGGACCTTGAAGGAGAATC. 300ng of samples were loaded on a 2% E-gel with 500ng of 1Kb Plus E-gel DNA ladder.

### **Quantitative Polymerase Chain Reaction**

qPCR utilized four different TaqMan probes (Invitrogen): Hs00951675\_m1 and Hs00370866\_m1 respectively target the exon junctions 5-6 and 9-10 of *MLIP*. Hs00153462\_m1 is specific to junction 2-3 of *LMNA*. Hs03004310\_g1 binds junction 1-2 of housekeeping gene *RPS29* for data normalization. Expression levels were assessed through relative  $\Delta\Delta CT$  quantification from the QuantStudio 6/7 System.

### **Western Blot**

Proteins were obtained from muscle tissues following TRIzol processing. 30 $\mu$ g of proteins were loaded on a gradient 4-15% polyacrylamide gel (Bio-Rad, 4561084). Samples were transferred on a PVDF membrane and blocked in 5% milk-PBS (30 minutes). Antibodies used in this report: MLIP rabbit-polyclonal (Invitrogen PA5-72759, 1:2000 dilution); LMNA rabbit-polyclonal (NEB 2026S, 1:1000 dilution); GAPDH rabbit-monoclonal (NEB 5174S, 1:1000 dilution); HRP anti-rabbit (Jackson ImmunoResearch 111-035-114, 1:10000 dilution). Primary antibodies incubation was done overnight at 4°C with dilutions in 1% BSA-0.1% sodium azide-PBS, and followed by 1-hour incubation with secondary antibody solution. Proteins revelation with an ECL substrate kit was followed by chemiluminescence capture through the ChemiDoc System, and images were analyzed with the Image Lab 6.0.1 software.



## **Long-Read Sequencing**

Library was prepared from total muscle RNA, reverse transcribed with the SuperScript Vilo mix and amplified (30 cycles) using the LongAmp *Taq* DNA Polymerase (NEB, M0323S). Two sets of barcoded primers targeted exons 3 (TGGACTCCGAAGGGGAAGAT) and 11 (GGGCGAATTACTCCAGGCTT). The amplicons were incubated with the Exonuclease I (NEB, M0293S) for single-stranded DNA cleanup. The library was end prepped with the NEBNext Ultra II End Repair/dA-Tailing Module to allow AMX adapter ligation with the NEBNext Quick Ligation Module ligase (ONT, SQK-LSK109). Purification with AMPure XP beads preceded the final 1X dsDNA QuBit4 fluorometric quantification, which allowed concentration-specific multiplexing of samples. Library was loaded for on a flow cell (FLO-MIN106) for MinION sequencing over a 24-hour period. The final data were processed with Guppy high-accuracy base calling and demultiplexing followed by the FLAIR pipeline <sup>27,28</sup>.

## **Data availability**

Data that support the findings of this study are available from the corresponding author upon request. The sequencing data of this study were stored in NCBI Bio Project (PRJNA809363).

## **Acknowledgements**

We would like to thank the patient for his willing participation in this study. We also thank Dr. Simon Grandjean Lapierre for his MinION Sequencing system, as well as Floriane Point and Dr. Karine Choquet for their helpful insights on setting up the LRS protocol. This project was supported by The Courtois Foundation, Fondation du Grand Défi Pierre Lavoie and the Rare Disease Foundation supported this study. S.A. received a master's degree fellowship from the

Canadian Institute of Health Research and F.M from the Neuroscience department of Université de Montréal. M.T. received a Junior 1 salary award from the Fond de recherche du Québec – Santé.

### **Authors' contribution**

J.M. and S.A. contributed to the work equally. J.M: Data acquisition, analysis, and interpretation (RT-qPCR, RT-PCR, Western blot & LRS). Wrote the manuscript. S.A: Data acquisition, analysis, and interpretation (RNA-sequencing, RT-PCR, Western blot & LRS). Wrote the manuscript. F.M: Data acquisition, analysis, and interpretation (RNA-sequencing & RT-qPCR). Reviewed the manuscript. J.C: Data acquisition, analysis, and interpretation (RT-PCR & Western blot). Reviewed the manuscript. V.T: Bioinformatic pipelines (LRS). Reviewed the manuscript. E.B: Bioinformatic pipelines (RNA-sequencing). Reviewed the manuscript. A.L: Data acquisition and analysis (Western blot). Reviewed the manuscript. J.K: Pathological evaluation of muscle biopsy. Reviewed the manuscript. R.M, C.H.C. and E.K.F.: Clinical evaluation. Reviewed the manuscript. M.T: Study design and data interpretation. Supervision of JM, SA, FM, JC, & VT. Wrote the manuscript.

### **Competing interests**

The authors declare no competing interests.

## References

- 1 Shieh, P. B. Muscular dystrophies and other genetic myopathies. *Neurol Clin* **31**, 1009-1029, doi:10.1016/j.ncl.2013.04.004 (2013).
- 2 Lopes Abath Neto, O. *et al.* MLIP causes recessive myopathy with rhabdomyolysis, myalgia and baseline elevated serum creatine kinase. *Brain* **144**, 2722-2731, doi:10.1093/brain/awab275 (2021).
- 3 Huang, Z. P. *et al.* Cardiomyocyte-enriched protein CIP protects against pathophysiological stresses and regulates cardiac homeostasis. *J Clin Invest* **125**, 4122-4134, doi:10.1172/JCI82423 (2015).
- 4 Cattin, M. E. *et al.* Deletion of MLIP (muscle-enriched A-type lamin-interacting protein) leads to cardiac hyperactivation of Akt/mammalian target of rapamycin (mTOR) and impaired cardiac adaptation. *J Biol Chem* **290**, 26699-26714, doi:10.1074/jbc.M115.678433 (2015).
- 5 Ahmady, E. *et al.* Identification of a novel muscle A-type lamin-interacting protein (MLIP). *J Biol Chem* **286**, 19702-19713, doi:10.1074/jbc.M110.165548 (2011).
- 6 Charniot, J. C. *et al.* Functional consequences of an LMNA mutation associated with a new cardiac and non-cardiac phenotype. *Hum Mutat* **21**, 473-481, doi:10.1002/humu.10170 (2003).
- 7 Muchir, A. *et al.* Activation of MAPK pathways links LMNA mutations to cardiomyopathy in Emery-Dreifuss muscular dystrophy. *J Clin Invest* **117**, 1282-1293, doi:10.1172/JCI29042 (2007).
- 8 Huang, Z. P. *et al.* CIP, a cardiac Isl1-interacting protein, represses cardiomyocyte hypertrophy. *Circ Res* **110**, 818-830, doi:10.1161/CIRCRESAHA.111.259663 (2012).
- 9 González-Jamett, A. M., Bevilacqua, J. A. & Díaz, A. M. C. in *Muscle Cell and Tissue* (ed Kunihiro Sakuma) Ch. 6, (2018).

- 10 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 11 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 12 Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273, doi:10.1038/ng1180 (2003).
- 13 Widyastuti, H. P., Norden-Krichmar, T. M., Grosberg, A. & Zaragoza, M. V. Gene expression profiling of fibroblasts in a family with LMNA-related cardiomyopathy reveals molecular pathways implicated in disease pathogenesis. *BMC Med Genet* **21**, 152, doi:10.1186/s12881-020-01088-w (2020).
- 14 Dalakas, M. C., Alexopoulos, H. & Spaeth, P. J. Complement in neurological disorders and emerging complement-targeted therapeutics. *Nat Rev Neurol* **16**, 601-617, doi:10.1038/s41582-020-0400-0 (2020).
- 15 Chaterji, S., Lam, C. H., Ho, D. S., Proske, D. C. & Baker, A. B. Syndecan-1 regulates vascular smooth muscle cell phenotype. *PLoS One* **9**, e89824, doi:10.1371/journal.pone.0089824 (2014).
- 16 Li, C. & Zhang, J. Stop-codon read-through arises largely from molecular errors and is generally nonadaptive. *PLoS Genet* **15**, e1008141, doi:10.1371/journal.pgen.1008141 (2019).
- 17 Sibley, C. R., Blazquez, L. & Ule, J. Lessons from non-canonical splicing. *Nat Rev Genet* **17**, 407-421, doi:10.1038/nrg.2016.46 (2016).
- 18 Chiba, M., Ariga, H. & Maita, H. A Splicing Reporter Tuned to Non-AG Acceptor Sites Reveals that Luteolin Enhances the Recognition of Non-canonical Acceptor Sites. *Chem Biol Drug Des* **87**, 275-282, doi:10.1111/cbdd.12656 (2016).

- 19 Cattin, M. E. *et al.* Expression of murine muscle-enriched A-type lamin-interacting protein (MLIP) is regulated by tissue-specific alternative transcription start sites. *J Biol Chem* **293**, 19761-19770, doi:10.1074/jbc.RA118.003758 (2018).
- 20 Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**, 30, doi:10.1186/s13059-020-1935-5 (2020).
- 21 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 22 Brouard, J. S., Schenkel, F., Marete, A. & Bissonnette, N. The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. *J Anim Sci Biotechnol* **10**, 44, doi:10.1186/s40104-019-0359-0 (2019).
- 23 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).
- 24 Blighe, K. & Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling (Bioconductor from within R. R version 4.0.3. , 2020).
- 25 Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008-2017, doi:10.1101/gr.133744.111 (2012).
- 26 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
- 27 Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* **20**, 129, doi:10.1186/s13059-019-1727-y (2019).
- 28 Tang, A. D. *et al.* Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat Commun* **11**, 1438, doi:10.1038/s41467-020-15171-6 (2020).

# Supplementary Material

## Supplementary Table 1. Genetic Analysis performed for patient Z46.

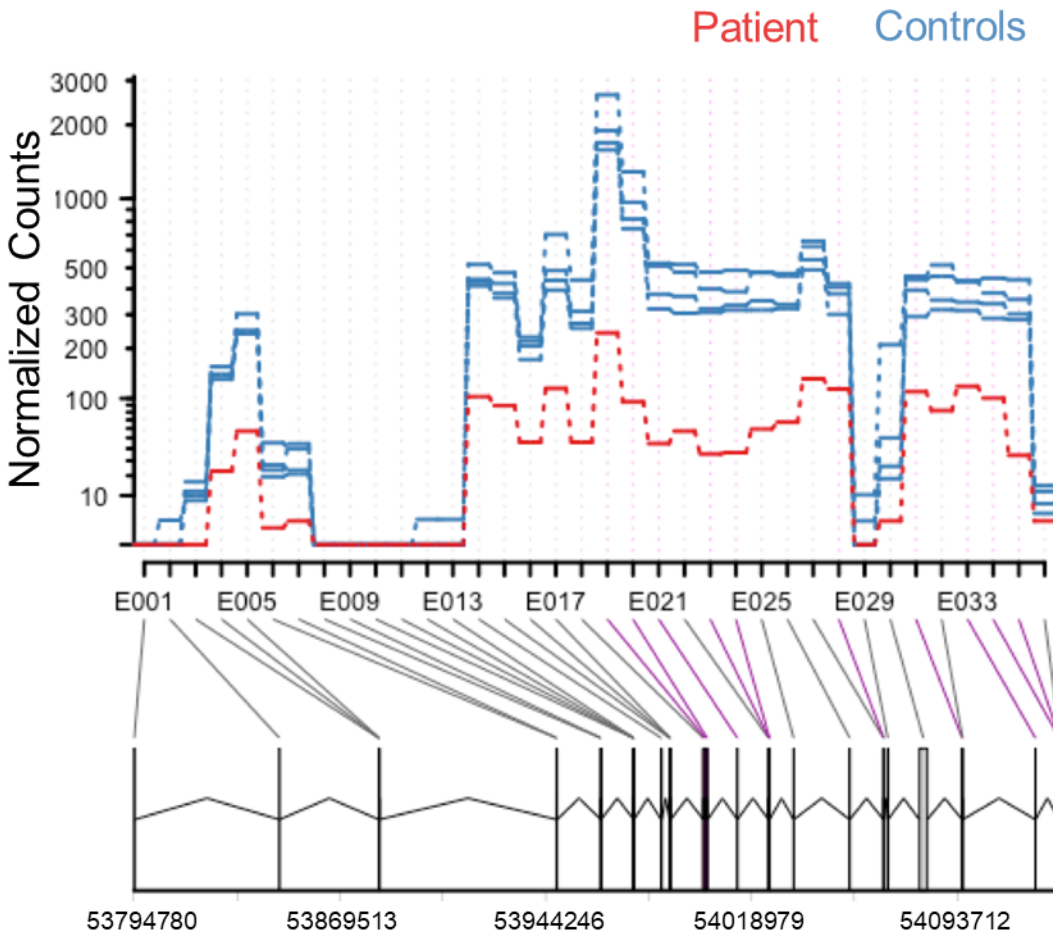
Gene	Mutation	Zygosity	Status	Associated Disease(s)	Inheritance(s)	Matching Phenotype	Source
RAPSN	c.264C > A	Heterozygous	Pathogenic / Likely Pathogenic	Congenital Myasthenic Syndrome 11 (CMS11)	Autosomal Recessive	No	Gene Panel <sup>1</sup> & RNA-seq
SCN4A	c.4270 G > A	Heterozygous	Pathogenic / Likely Pathogenic	Hyperkalemic Periodic Paralysis (HYPP) - <del>Potassium</del> Congenita (PMC)	Autosomal Dominant - Autosomal Dominant	No	Gene Panel <sup>1</sup> & RNA-seq
MLIP	c.2284 C > T	Homozygous	VUS	-	-	-	RNA-seq
TTN	c.38002 A > G	Heterozygous	VUS	Dilated Cardiomyopathy 1G (CMD1G) - Limb-Girdle, type 2J (LGMD2J)	Autosomal Dominant - Autosomal Recessive	No	RNA-seq
GMPPB	c.893 G > A	Heterozygous	Likely Pathogenic	Muscular Dystrophy-Dystroglycanopathy Type A14 - C14 (MDDGA14 - MDDGC14)	Autosomal Recessive - Autosomal Recessive	No	RNA-seq
LMOD3	c.39_41Del	Heterozygous	Benign	Nemaline Myopathy 10 (NEM10)	Autosomal Recessive	No	RNA-seq
AMPD1	c.303-1 G > A	Heterozygous	VUS	Myopathy Due to Myoadenylate Deaminase Deficiency (MMDD)	Autosomal Recessive	Moderate	RNA-seq
KIF21A	c.3581 C > G	Heterozygous	Benign / Likely Benign	Congenital Fibrosis of the Extraocular Muscles (FEOM)	Autosomal Dominant / Recessive	No	RNA-seq

**Negative results from gene panel and RNA-seq analysis:** AARS2; ACADS; ACADVL; ACTA1; ADCK3; AFG3L2; AGK; AGL; AGRN; AIFM1; ALDOA; ANO5; ATP2A1; B3GALNT2; B3GNT1; BAG3; BIN1; C10ORF65; CACNA1S; CAPN3; CAV3; CGDCT8; CFL2; CHAT; CHCHD10; CHKB; CHRNA1; CHRNB1 - CHRND; CHRNE; CLCN1; CNBP; CNTN1; COA5; COL4A1; COL6A1; COL6A2; COL6A3; COLO; COQ2; COQ6; COQ8; COX10; COX14; COX6B1; CPT2; CRYAB; DES; DMD; DMPK; DNA2; DNAJB6; DNME2; DOK7; DUX4; DYX1F; EARS2; EMD; ENO3; FARS2; FASTKD2; FHL1; FKBP; FKTN; FLNC; FOXRED1; GAA; GBE1; GFER; GFMI; GHPT1; GNE; GIDC2; GYGI; GYS1; HADHA; HADHB; HCCS; HINT1; HNRNP1; HNRNP2B1; HRAS; HSPG2; ISCU; ISPD; ITGA7; KABTBD13; LAMA2; LAMP2; LARGE; LDB3; LMNA; LPIN1; MATR3; MEGF10; MGME1; MPV17; MRPL3; MRPS16; MRPS22; MTFMT; MTM1; MTMR14; MTO1; MUSK; MYF6; MYH14; MYH2; MYH7; MYH7B; MYL2; MYOT; NEB; OPAL1; PABPN1; PDSS1; PDSS2; PFKM; PGAM2; PGK1; PHKA1; PLEC; PNPLA2; PNPT1; POLG; POLG2; POMGNT1; POMT1; PRKAG2; PRKAG3; PTRF; PUS1; PYGM; RBCK1; RMND1; RRM2B; RYR1; SARS2; SCOL1; SCOL2; SDHAF1; SEPN1; SEPT9; SGCA; SGCB; SGCD; SGGC; SGK196; SIL1; SLC22A5; SLC25A4; SLC37A4; SLC52A1; SLC52A2; SMCHD1; STIM1; SUSL1; SYNE1; SYNE2; TAZ; TCAP; TK2; TNNT1; TNPO3; TPM2; TPMB; TRIM3; TRMU; TSFM; TUBB3; TUFM; TYMP; VCP; YARS2

**Additional negative results (RNA-seq analysis only):** BYN1, CLC1, COO10, CYBA, CYBB, DAG1, DMM2, DNPK, ETF, ETFH, HNRPDL, KBTBD13, KLHL4, KLHL40, KLHL41, LBD3, LIMS2, MYPN, OCTN2, POGD1, POPDC1, SPEC, TIA1, TOMP1, TOR1AIP1, TRAPP11, ZNF9

<sup>1</sup> Panel from Medical Neurogenetics LLC. Intron/exon junction coverage ≥ 99.0

**Supplementary Figure 2.**



**Supplementary Figure 2.** Exon usage of *MLIP*.

Exon usage analysis was conducted using DEXSeq, where Ensembl transcripts were collapsed.

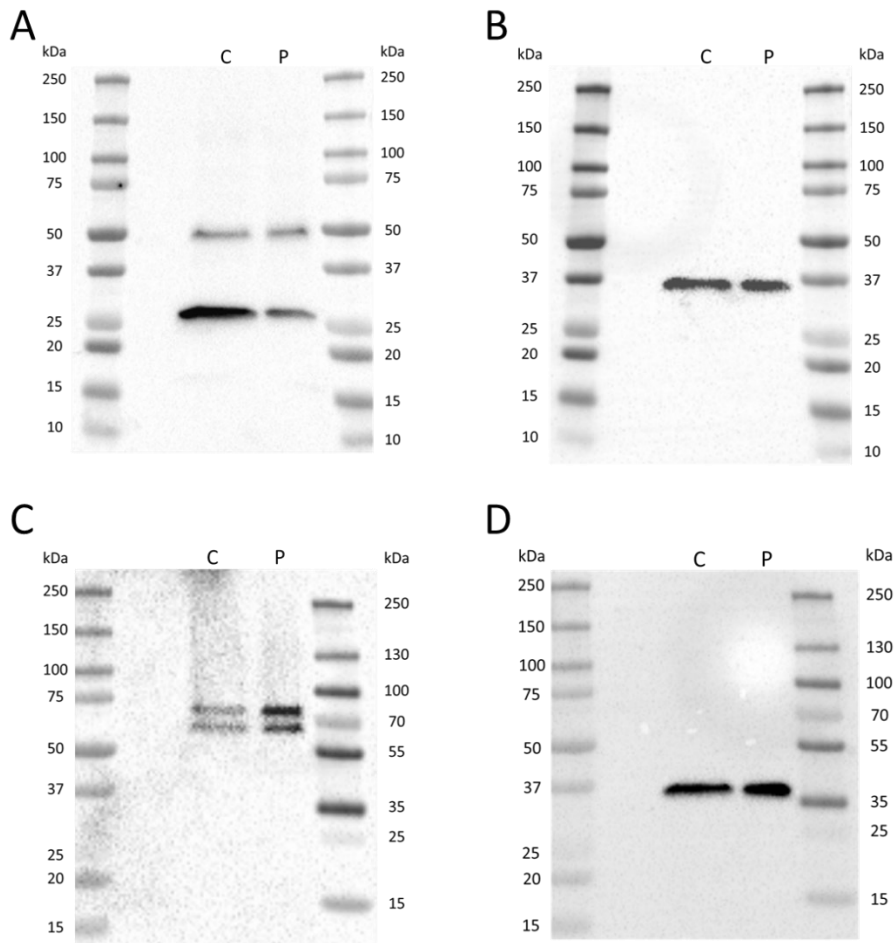
The flattening approach generates a singular transcript with 36 exons, where exon 5 of *MLIP*

corresponds to either exon 18, 19 or 20 in the representation. When comparing the affected

individuals with the healthy samples, a general downregulation of *MLIP* levels can be noted ( $\approx$  3-

5 fold), and is slightly emphasized around exon 5 ( $\approx$  8-10 fold).

**Supplementary Figure 3.**



**Supplementary Figure 3. *MLIP* and *LMNA* respective expression in muscular tissue**

(A) Western Blot analysis reveals the presence of four isoforms of *MLIP*, two of which can be quantified (27kDa and 50kDa). (C) Western Blot analysis reveals the presence of the two isoforms of *LMNA*, lamin A (75kDa) and lamin C (65kDa) were analyzed separately. (B, D) *GAPDH* is used as the reference gene for normalization. Full unaltered membranes are shown.