Université de Montréal

# Identification and functional analysis of novel pathogenic variants in patients with undiagnosed myopathies

Par

Jennifer Hauteclocque

Department of Neurosciences, University of Montréal, Faculty of Medicine

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

Juin 2021

Université de Montréal

Unité académique: Department of Neurosciences, University of Montréal, Faculty of Medicine

*Ce mémoire intitulé*

**Identification and functional analysis of novel pathogenic variants in patients with undiagnosed myopathies**

*Présenté par*

**Jennifer Hauteclocque**

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

**Dre. Inge Meijer**
President-rapporteur

**Dre. Martine Tétreault**
Directrice de recherche

**Dre. Claire Bénard**
Membre du jury

# Résumé

« Myopathie héréditaire » est un terme générique pour les maladies génétiques rares caractérisées par une faiblesse musculaire et une hypotonie avec ou sans atrophie musculaire. Les personnes atteintes d'une forme légère peuvent présenter des contractures, une scoliose, une hyporéflexie ou des caractéristiques dysmorphiques, et les plus sévères peuvent être accompagnées de symptômes cardiaques ou respiratoires pouvant s'avérer mortel. Alors que les méthodes de séquençage de nouvelle génération basées sur l'ADN ont considérablement accéléré la découverte de gènes responsables de maladies rares, de nombreux patients demeurent sans diagnostiques génétiques. L'une des principales raisons de ce problème est le grand nombre de variants de signification inconnue identifiés, où l'impact biologique est peu ou pas connu. Ce mémoire de maîtrise contient trois projets distincts dont l'objectif global est d'augmenter le rendement diagnostique pour les patients atteints de myopathies héréditaires rares. La première étude porte sur trois frères et sœurs atteints d'une dystrophie musculaire non diagnostiquée. Une combinaison de techniques « omic » a été utilisée pour identifier un variant faux-sens dans le gène *IARS* accompagné d'un déséquilibre allélique spécifique aux tissus musculaires. L'inhibition de *iars-1* chez le *C. elegans* a entraîné une désorganisation progressive du muscle de la paroi corporelle, mais sans perte significative de la motilité. Ainsi, nous avons conclu que *iars-1* joue clairement un rôle dans l'organisation des myotubes. La pathogénicité du variant, cependant, nécessite une enquête plus approfondie. La deuxième étude porte sur une femme présentant une myopathie statique congénitale se manifestant par une faiblesse proximale et distale. En utilisant le séquençage de l'ARN, nous avons identifié pour la première fois un profil d'expression génique compatible avec une prédominance des fibres musculaires de type I, focussant l'intérêt sur un variant dans le gène *RYR1*. La troisième étude englobe une cohorte de vingt-huit patients porteurs de la même mutation *RYR1*, mais présentant une hétérogénéité clinique significative. Des modèles « knock-in » de *C. elegans* pour les études deux et trois ont démontrés des changements en transmission synaptique, la durée de vie, la taille corporelle et la locomotion. Ainsi, nous avons conclu que les deux variants identifiés dans *RYR1* ont probablement également des conséquences cliniques chez les porteurs humains. En fin de compte, ces études mettent en évidence l'utilité du séquençage de l'ARN en tant qu'outil de diagnostic complémentaire, capable de restreindre la liste de candidats potentiellement pathogéniques, ainsi que le pouvoir du *C. elegans* en tant que modèles pour des tests rapides et coordonnés de variants candidats.

**Mots-clés** : Myopathies, Dystrophies, Séquençage de nouvelle génération, Séquençage ARN, Interprétation de variants, *Caenorhabditis elegans*, génétique, muscle.

# Abstract

"Hereditary myopathies" is an umbrella term for rare inherited diseases characterized by muscle weakness and hypotonia with or without muscle atrophy. Individuals with a mild affliction may present with contractures, scoliosis, hyporeflexia or dysmorphic features, while those more severely affected may present cardiac or respiratory involvement that could prove deadly. While traditional DNA-based next-generation sequencing techniques have greatly accelerated discovery of genes causing rare diseases, many patients remain without a known genetic cause. The main reason for this diagnostic shortfall is the vast number of variants of unknown significance identified whose biological functions are unknown. This master's thesis contains three separate projects with an overarching goal to increase the diagnostic yield of patients with rare hereditary myopathies. The first study focuses on three siblings with an undiagnosed muscular dystrophy. A combination of "omic" techniques were used to identify a missense variant as well as a muscle-specific allelic imbalance in the gene *IARS* leading to the exclusive expression of the mutant allele. *Iars-1* knock-down in *C. elegans* resulted in progressive disorganization of the body wall muscle but with no significant loss of motility. Thus, we concluded that *iars-1* likely plays a role in the organization of myotubes. The pathogenicity of the variant, however, requires further investigation. The second study involves a woman with a congenital static myopathy exhibited as proximal and distal weakness. Using RNA-sequencing, we identified for the first time a gene expression profile consistent with type I fiber predominance in the proband which guided the search for the causative *RYR1* variant. The third study encompasses a cohort of twenty-eight patients who carry the same *RYR1* mutation but display significant clinical heterogeneity. Knock-in models of *C. elegans* for both studies demonstrated altered synaptic transmission, lifespan, body size and locomotion. Thus, we concluded that both variants identified in *RYR1* likely have consequences for human carriers as well. Ultimately, these studies highlight the utility of RNA-seq as a complimentary diagnostic tool capable of narrowing the search for novel pathogenic mutations as well as the value of *C. elegans* as models for rapid and coordinated testing of candidate variants.



**Key words**: Myopathy, dystrophy, next-generation sequencing, RNA-sequencing, variant interpretation, *Caenorhabditis elegans*, genetics, muscle

# Table of Contents

## List of Tables

## List of Figures

### Main text



### Functional analysis of an RYR1 variant underlying a myopathy with variable expressivity



### Case study: A novel RYR1 variant associated with type I fiber predominance identified using RNA-sequencing

## List of Abbreviations

| | |
|---|---|
| aa-AMP | aminoacyl adenylate |
| agars | aminoacyl tRNA synthetase |
| ABD | anticodon-binding domain |
| Ach | acetylcholine |
| anti-ss | antisynthetase syndrome |
| AP | action potential |
| *C. elegans* | Caenorhabditis elegans |
| CCD | catalytic core domain |
| cDNA | complimentary DNA |
| CF | Common Fund |
| CFTD | congenital fiber type disproportion |
| CGC | Caenorhabditis Genetics Center |
| CM | congenital myopathy |
| CMD | congenital muscular dystrophy |
| CNM | centronuclear myopathy |
| CRISPR | clustered regularly interspaced short palindromic repeats |
| crRNA | CRISPR RNA |
| DGE | differential gene expression |
| DMD | Duchenne's muscular dystrophy |
| *DNMT1* | DNA cytosine-5-methyltransferase 1 |
| *DNMT2* | DNA cytosine-5-methyltransferase 2 |
| DR2 | divergent region 2 |
| dsRNA | double stranded RNA |
| EC | excitation-contraction |
| EV | empty vector |
| GFP | green fluorescent protein |
| GRIDHH | growth retardation, impaired intellectual development, hypotonia and hepatopathy |
| gRNA | guide RNA |
| GTEx | Genotype-Tissue Expression Project |
| HDR | homology directed repair |
| *IARS* | isoleucyl-tRNA synthetase |
| INDEL | insertion/deletion |
| IPTG | isopropyl β-D-1-thiogalactopyranoside |
| LoF | loss of function |
| MAF | minor allele frequency |
| MH | malignant hyperthermia |
| MHS | malignant hyperthermia susceptibility |
| MMD | multi-minicore disease |
| MNV | mulltinucleotide variant |
| mRNA | messenger RNA |
| NGM | nematode growth media |

| | |
|---|---|
| NGS | next-generation sequencing |
| NHEJ | non-homologous end joining |
| NMJ | neuromuscular junction |
| *NRAP* | nebulin related anchoring protein |
| NTD | N-terminal domain |
| *OBSCN* | obscurin |
| PAM | protospacer adjacent motifs |
| PPi | pyrophosphate |
| RISC | RNA-induced silencing complex |
| RNAi | RNA interference |
| RNAi-TGS | RNAi-induced transcriptional gene silencing |
| RNA-seq | RNA-sequencing |
| rRNA | ribosomal RNA |
| RT | reverse-transcription |
| RYR1 | ryanodine receptor 1 |
| RYR1-RM | RYR1-related myopathy |
| *SEPN1* | selenoprotein 1 |
| SERCA | sarcoplasmic reticulum calcium ATPase |
| SNV | single nucleotide variant |
| ssRNA | single stranded RNA |
| tracrRNA | trans-activating crRNA |
| tRNA | transfer RNA |
| TTN | titin |
| VC | vital capacity |
| VUS | variant of unknown significance |
| WES | whole exome sequencing |
| WGS | whole genome sequencing |
| wt | wild-type |
| α-DG | α-dystroglycan |

## Acknowledgements

This master's thesis is the denouement of a long journey of diverse challenges that came into fruition as a result of the boundless support of my supervisor, colleagues, family, and friends. This experience has afforded me an exceptional opportunity to pursue research in a laboratory under the guidance of experts while reveling in a joyful life there.

I would first like to thank my supervisor, Martine Tétreault, without whom I would never have had the courage to take on such an endeavor. As a fresh undergraduate student, inevitably, some experiments did not go according to plan; however, these instances of disappointment and self-doubt were always met with kindness, patience, and a positive outlook for new learning opportunities. You truly foster a respectful and compassionate laboratory dynamic revered by your team and that any student would be lucky to be a part of. Thank you, Martine, for entrusting me as your first student and for helping me grow into a stronger, wiser, and more confident scientist.

I would also like to thank my colleagues including past, current, and returning members of "Team Tétreault". Thank you, Jean and Sébastien, for elevating the laboratory experience with your goofiness and sarcasm (respectively) in between helping me with experiments, reports, translations, and more. The depth of my gratitude for our friendship is impossible to describe in words. I would also like to thank Adrien for entrusting me as his mentor and for all his help on the *C. elegans* experiments. I learned as much from you as you did from me, and I positively believe that your incredible work ethic will get you anywhere you want to be in life regardless of the circumstances. Though not an official member of our lab, I would also like to thank Julie for gracing us with her omniscience regarding laboratory techniques. Your down-to-earth personality, good taste in music, and infinite wisdom will be sorely missed. Finally, I would also like to thank Nab, Camille, Lovatiana, Jade, Renaud, Annie, Marjorie, Valerie, Camberly, Moustafa and Gaël for an unforgettable laboratory experience. I can't wait to see what you all accomplish.

All the members of Parker lab also deserve a very special thank you for all the support and encouragement they gave me in the design and setup of my *C. elegans* experiments. Sarah, Ericka, and Dr. Parker, I greatly appreciate all your help and mentorship. From nothing, you built up a solid foundational knowledge of worms and related laboratory techniques. Ericka, your unrelenting enthusiasm to spice up each holiday and birthday is something to be admired, and the Parker lab is extremely lucky to have you. Another thank-you to Audrey, for the major role you played in the fluorescent microscopy experiments, your approachability, and your willingness to help others. Finally, thank you Gilles, for your WormLabs expertise and for allowing me to unknowingly mispronounce your name for the better part of our friendship.

Last but certainly not least, I would like to thank my friends and family for all their love and encouragement. I gratefully acknowledge your patience and benevolence, Austin, and thank you for making it easier to put on a brave face through some of the more stressful times.

I would also like to thank my siblings, Al, Jay, Louis, Phil, and Meag, and my parents who listened, or at least pretended to listen, to the different challenges and intricacies of my research life. Though my masters took me many years, you always encouraged me that home would be waiting for me when I finished. It looks like it was possible after all, and I made it!

*I dedicate this work to my cherished brother, Louis, who has been my biggest cheerleader throughout my degree. In your eyes, no accomplishment was too small for celebration and your pride in me has helped me discover my own.*

# Chapter 1: Introduction

## 1.1 Current status of hereditary myopathies

Hereditary myopathies are a heterogeneous group of neuromuscular disorders caused by genetic mutations primarily affecting skeletal muscle structure and function. Though all myopathies typically manifest as hypotonia and static or progressive muscle weakness, subgroups are categorized by distinct albeit often overlapping histological, clinical, and genetic features. Two major subgroups include non-dystrophic myopathies and dystrophic myopathies (dystrophies). Histological features of non-dystrophic myopathies range from fiber-type disproportion and myofibrillar disorganization to the presence of rods, cores, and central nuclei among other structural anomalies (1). Meanwhile, dystrophic myopathies are classified by fibrosis, wasting, and atrophy (1). Although myopathies can develop at any stage of life, the most severe cases are generally found in early onset forms known as congenital myopathies (CMs) (2,3) affecting up to 1 in 25,000 people (3, 4). These disorders demonstrate recessive, dominant, or X-linked modes of inheritance, and are generally separated into four categories: 1) core myopathies; 2) nemaline myopathies; 3) centronuclear myopathies; 4) congenital fiber type disproportion myopathy. Those severely afflicted may exhibit signs of feeding and respiratory difficulties, delayed motor milestones and on rare occasions, cardiac involvement (5). The involvement of the respiratory muscles is often used as an indicator of prognosis (1). Otherwise, more mild symptoms may include contractures, scoliosis, hyporeflexia and dysmorphic features (5).

Disease severity has a significant impact on quality of life as some individuals may only experience reduced endurance and fatigue while others are wheelchair bound. The International Standard Care Committee for Congenital Myopathies agrees that optimal care for patients with CMs entails structured disclosure of disease diagnosis, prognosis, risks, and treatment, consistent follow-up care with repeated anticipatory guidance, and specialty care as required (6). Yet, as many physicians are unfamiliar with such rare disorders, quality of care is highly variable. One gap in patient care arises from the approximate 40% of patients who remain without a molecular diagnosis due to our limited understanding of the

pathomechanisms involved in CMs (7, 8). Filling this gap is important for accurate prognosis, optimizing disease management, and informing reproductive choices (9). Furthermore, early diagnosis usually allows the individual to seek care and support sooner, maintain independence longer and has been associated with psychological benefit and improved quality of life (10).

Another gap in patient care is a complete lack of FDA approved treatments (11), despite promising preliminary data. For example, while some medical conditions have shown improvement with physical exercise, its effectiveness in neuromuscular disease and congenital myopathies is, to the best of our understanding, not harmful (12, 13). Similarly, the antioxidant N-acetylcysteine (NAC) (14) successfully reduced oxidative stress in cultured myotubes from patients with RYR1-related myopathies (RYR1-RMs) *ex vivo* (15); however, oral administration of NAC was ineffective in human patients (16). Further investigation into the fundamental mechanisms underlying CMs is necessary for the identification of effective treatments and pharmacological agents that could remedy this shortcoming. In the meantime, patient care is mainly focused on management of individual symptoms including drug therapy, genetic counselling, nutritional support, assisted breathing, orthopaedic treatments, and an array of therapies that address physical, speech, respiratory, or occupational difficulties.

**1.2 Diagnostic approaches**

Currently, histological and morphological features of muscle biopsies are the primary basis for classification of hereditary myopathies; however EMG, MRI, muscle ultrasound, serum creatine kinase levels and nerve conduction studies may also be used to exclude other diagnoses (7). Nonetheless, significant histological overlap between different myopathies or a lack of defining histological features often complicates diagnosis (7). As a result genetic tests are increasingly used in a clinical setting. Traditional approaches, such as linkage analysis and homozygosity mapping followed by candidate gene sequencing, required a large number of individuals and are less suited for heterogeneous disorders, small families, or sporadic cases (17). Nonetheless, linkage analysis studies are responsible for the identification of more than 3,000 genes underlying Mendelian diseases (17). Today, next-generation sequencing (NGS) is replacing older techniques as it becomes more and more standardized in a clinical environment. While the human genome project took over a decade to complete through

Sanger sequencing (18), the same can be achieved in a single day by NGS due to the novel ability to sequence millions of DNA fragments in parallel. One of the most common applications of NGS is the gene panel which offers an affordable (19), sensitive (20), and rapid approach to sequencing known-disease causing genes and gene regions for pathogenic variants. While limited to what is known of disease-causing genes, the data output of gene panels is simple to interpret and the diagnostic rate is >50% when used for primary diagnostics for CMs (8). This yield suggests that a large portion of pathogenic variants have already been discovered. For the remaining variants, NGS techniques such as whole genome sequencing (WGS) and whole exome sequencing (WES) are gradually building traction as a tool for gene discovery, especially for rare disorders, in laboratories across Canada, Europe, and the United States (21). However, the widespread adoption of NGS and WES into a clinical environment is impeded by the cost of sequencing, the time-consuming process for analysis, and the need for large amounts of computational storage space (3). The broader sequencing scope encompassed by these techniques also poses an ethical dilemma due to the increased risk of incidental findings.

Compared to WGS, WES only sequences the coding regions of the genome and so requires enrichment of these target regions by PCR amplification and probe targeted hybridization. The variable efficiency of these steps on an exome-wide scale leaves WES vulnerable to uneven sequence depth and incomplete coverage (22). Indeed, the proportion of false-positive single nucleotide variants is much higher for WES (78%) than for WGS (17%) largely due to gaps in coding region coverage (22). However, the price of better coverage, higher specificity and broader variant detection implied by WGS is the challenge of interpreting variants from a larger dataset (22). In fact, our ability to identify genetic variants has far surpassed our ability to interpret their functional effects. As a result, NGS techniques often result in the detection of an overwhelming number of variants of unknown significance (VUS), which may explain the diagnostic shortfall observed in CMs. This is especially true for common CM-associated genes like *RYR1* (ryanodine receptor 1) or *TTN* (titin) that are some of the largest genes in the human genome and are especially prone to polymorphisms (7, 8). Another explanation is the fact that WES, as the more commonly used technique, only captures about 1% of the genome (23) where approximately 85% of disease-causing variants are thought to be

found (24). Thus the intronic and non-coding regulatory elements such as promoters, enhancers, and microRNAs harbouring up to 25% of pathogenic variants are intrinsically overlooked (7).

### 1.3 RNA-sequencing and gene discovery

RNA-sequencing (RNA-seq) is an alternative application of NGS that sequences the actively expressed transcriptome of a particular tissue or condition and is hence able to bypass the costly exome-enrichment step required by WES (25, 26). Similar to WES, RNA-seq provides variant calling in the coding regions of the genome, but captures only what is actively transcribed in the target tissue. This includes known and unidentified mRNA transcripts, non-coding RNAs, microRNAs, PIWI-interacting RNAs, and tRNAs depending on the protocol used (26, 27). When using peripheral blood mononuclear cells, this amounts to about 40% of all coding single nucleotide variants (SNVs) or 81% of peripheral blood SNVs (28). This tissue-specificity necessitates sequencing of biological materials with disease-representative expression patterns. Fortunately, this is seldom a problem in heritable myopathies as muscle biopsy is relatively readily available (9). Alternatively, sequencing data can also be procured from databases such as the Common Fund's (CF) Genotype-Tissue Expression Project (GTEx) (29). Tissue specificity is a limiting factor for some diseases, but its striking impact on diagnostic yield and the ability to narrow the search for disease-causing variants is otherwise a valuable asset (30).

RNA-seq possesses a powerful advantage over other NGS techniques in that it provides insight into gene expression, allele-specific expression, gene splicing, and fusions (31, 32) which can be used to interpret the significance of VUSs (31-34). For example, a study by Cummings *et al.* demonstrated in a cohort of 50 patients with undiagnosed heritable myopathies that RNA-seq data contradicting predictions of aberrant splicing for a particular variant is a valuable means of ruling out non-pathogenic VUSs (31). This is especially beneficial in CMs where the causative genes are particularly polymorphic. This study also illustrated how the detection of aberrant splicing events observed in patients but not controls was a significant contributor to the identification of eight pathogenic non-coding variants. Thus RNA-seq can be used to identify

the functional effect of an elusive pathogenic variant, whether it be in the coding or non-coding regions, which can then be used to guide the identification of the variant itself.

The combination of RNA-seq and DNA-based sequencing approaches represent a powerful diagnostic duo. Between 9-30% of pathogenic variant affecting RNA-expression or processing are found in non-coding regions (30), and would require DNA-sequencing to be identified. Alternatively, RNA-seq applied as a complimentary diagnostic tool greatly boosts diagnostic resolution in cases where WES, WGS, and/or gene panel have been inconclusive. For instance, in a previous study clinical genetic testing including Sanger sequencing and gene panel failed to find a pathogenic variant in a 6-year-old boy with a clinical and histological presentation consistent with Duchenne's muscular dystrophy (DMD) (34). However, employment of RNA-seq and RT-PCR in combination identified a 51 base pair insertion and the causative intronic variant in the gene *DMD*. Similarly, in Cummings' study, RNA-seq helped achieve a diagnosis in 17 of 50 (35%) previously unsolved cases of myopathic or dystrophic patients that underwent WES or WGS (31). A similar yield (36%) was obtained when RNA-seq was applied to 25 cases of monogenic neuromuscular disorders that were exome and/or gene panel negative (30). Despite its clearly demonstrated utility in these anecdotal studies, RNA-seq is still an unrefined technology. In order for it to expand its clinical utility consistent standards for assay optimization, analytics, bioinformatics, and regulation are required along with supporting evidence of reproducibility (27). In the meantime, RNA-seq will continue to be used in academia where the optimal RNA-seq pipeline differs from laboratory to laboratory depending on the goals of the research (35). Nonetheless, a generic roadmap for RNA-seq procedures includes RNA extraction, library preparation, sequencing, alignment, and analysis.

### 1.3.1 RNA extraction, library preparation and alignment

The first step of RNA-sequencing is the extraction of RNA from the sample. Researchers are typically interested in messenger RNA (mRNA) because it acts as the intermediary between genes and proteins. However, mRNA makes up only about 1-2% of total RNA whereas ribosomal RNA (rRNA) comprises about 90% of total RNA (35). Therefore, in order to have sufficient mRNA for detection, one option is to enrich for mRNA using selection of poly(A). In Eukaryotes, almost all mRNA transcripts contain poly(A) tails that allow them to be selected for

against other types of RNA. Poly(A) selection is thought to provide the best gene coverage and quantification accuracy (36), however plentiful and high-quality mRNA is required (35, 37). When sample quantities are limited, another option is to deplete samples of rRNA (35).

Many commercial sequencing kits available (TruSeq® RNA Sample Preparation V2, Universal Plus mRNA-Seq, and NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina®) combine RNA capture with cDNA library preparation through the following general workflow: RNA capture, cDNA synthesis, end-repair, adaptor ligation, library enrichment (Fig. 1) (36). After poly(A) selection or rRNA depletion, mRNA must be chemically fragmented into pieces of a certain size determined by the sequencing platform (e.g. Illumina pair-end reads 100-125bp) (37). The challenge however is to achieve random fragmentation such that all regions of RNA are equally represented (37). Afterwards reverse-transcription (RT) using random hexamer primers is performed to make complimentary DNA (cDNA) and adaptors are ligated to the 3' and 5' ends of each fragment. In order to maintain directionality, adaptors for each end must be different from one another or the antisense strand must be labelled with dUTP and degraded enzymatically before amplification (37, 38). As such accurate quantification of the sense strand responsible for translation is achieved. Finally, several cycles (8-12) of PCR are performed to amplify the library. Due to the variability of size and composition of cDNA fragments differential amplification must be corrected or accounted for (37). Furthermore, the abundance and quality of amplicon fragments produced in the library as well as genome complexity have direct effects on the sensitivity of RNA-seq for detecting true variants (39, 40). Increasing read depth, the number of times a nucleotide is read in a certain position of the genome, allows for the detection of rare variants found on lowly expressed transcripts. However, too much read depth dramatically increases the number of false positives detected (41). Therefore, read depth is an important consideration that depends on the research goals.

RNA-sequencing platforms quantitatively sequence the millions of fragments generated from the library preparation step and output the data in the form of FASTQ files (42). Each fragment is trimmed and processed to form a read, and each read must reach a quality threshold to carry on to the next step. The remaining reads are then aligned to a reference genome using alignment tools such as STAR, TopHat, Bowtie, or Trinity (25, 43, 44).

Alternatively, when no reference genome is available reads can be assembled *de novo* using alignment tools such as SOAPdenovo-Trans, Trinity or Oases (35). Hereafter, we describe the role of RNA-seq in the analysis of genetic variants, differential expression, alternative splicing, and allelic imbalances.

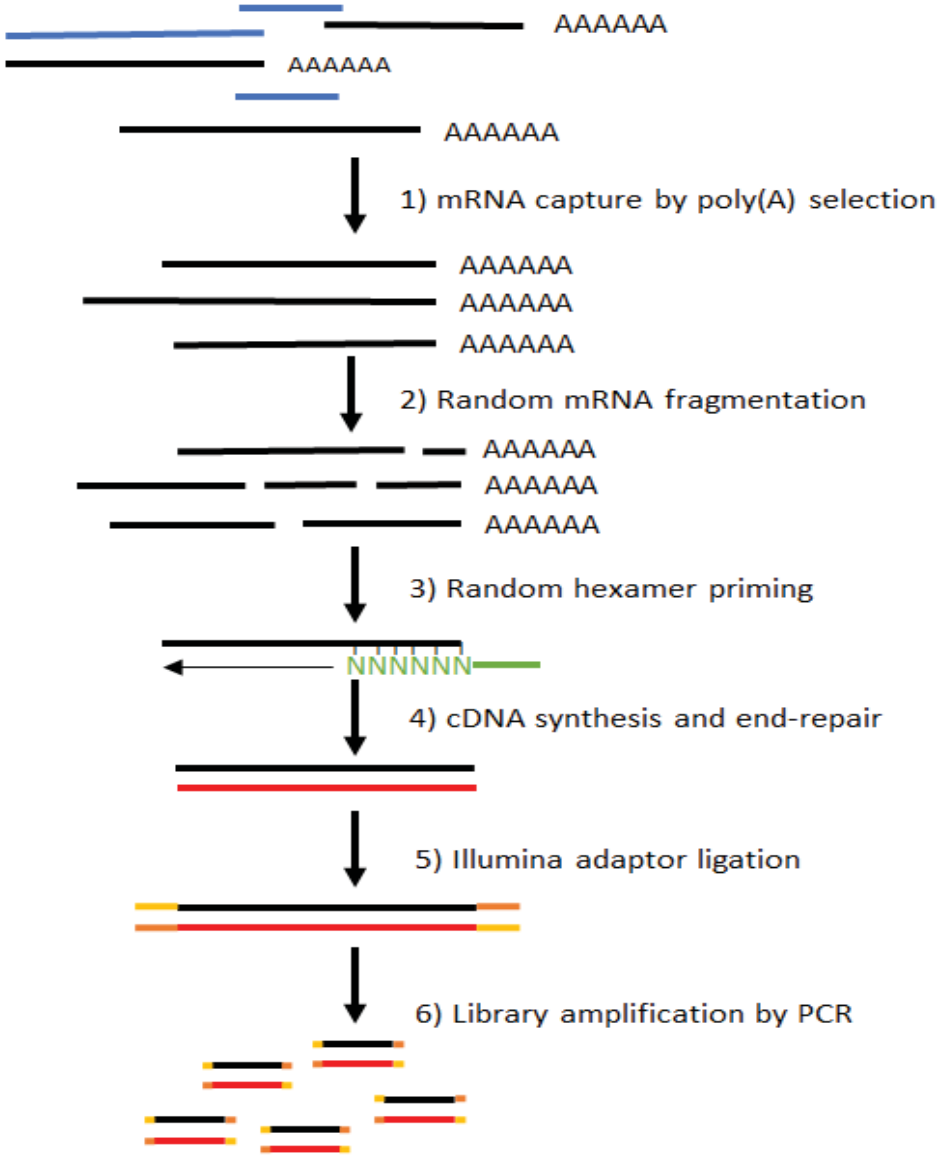



**Figure 1. RNA-seq workflow**
Generic workflow of RNA-sequencing library construction consisting of RNA capture, cDNA synthesis, end-repair, adaptor ligation, and library enrichment. Blue=non-mRNA, green=hexamer primer, red=cDNA strand, yellow=5' adaptors, orange=3' adaptors.

### 1.3.2 Variant calling

After alignment, the newly constructed transcriptome is compared to a reference genome to identify variants in the form of SNVs, multi-nucleotide variants and small insertions or deletions (INDELs) (39). RNA-seq is able to detect about 70-82% of coding variants detected by whole genome-sequencing (WGS) (39, 45). Various annotation tools such as ANNOVAR (46) or VEP (47) can then be used to assign functional information to each variant with the goal of linking variants to the patient's phenotype (17). These annotations are used to filter for the most probable variants for manual analysis. The annotations and filters used will depend on the disease under study, mode of inheritance and the research goal. Common annotations include the type of variant (nonsense, stop gain, frameshift, missense, synonymous), the allele frequency in public databases (ExAC (48), The 1000 Genomes Project (48), GnomAD (49)), pathogenicity predictions (CADD (50), Mutation Taster (51), Polyphen (52), etc.), and disease-associated phenotypes (OMIM (53), ClinVar (54)).

### 1.3.3 Differential expression

Currently, RNA-seq is the primary option for differential gene expression (DGE) analysis (55). RNA-seq has been reported to be as accurate as DNA microarrays and correlated very well with qRT-PCR data (56). However, RNA-seq has a significantly larger dynamic range of abundance than the aforementioned approaches allowing for the detection of very highly and very poorly expressed transcripts (56-58). DGE tools can be used to estimate the magnitude and significance of differentially expressed genes between conditions and can correct for multiple testing (59). EdgeR (60) and DESeq2 (61) are commonly used tools that use regression models and normalize data according to sequence depth, gene length, and/or RNA composition in order to estimate differences in gene expression (59). Optimally, patient expression profiles are compared to those of healthy individuals matched for relevant factors such as age and sex. A filtering process, often involving fold change and p-value, may be applied in order to prioritize those for manual investigation (61). The list of differentially expressed genes can also be clustered by associated disease, biological pathway, tissue expression and more using online tool such as DAVID Bioinformatic Tools (62). Defining differential expression signatures relative to a corresponding disease can provide vital clues on how to best care for a patient. In the past,

DGE analysis has been used to identify long term effects of certain lifestyle choices (63), to identify therapeutic targets (64-66), to test the efficacy of drugs (67), and to identify biomarkers for disease (65, 68). In the context of hereditary myopathies, expression profiling has been used for the discovery of novel biomarkers for disease, disease progression and therapeutic efficacy (69, 70) as well as for the identification of affected gene pathways, mechanisms and therapeutic targets (70, 71).

### 1.3.4 Alternative splicing

Many isoforms of mRNA with different function can be produced from a single gene by skipping one or more exons in a naturally occurring process called alternative splicing. Alternative splicing can be influenced by epigenetic factors (72) and is controlled by splicing machinery that depends on genetic sequences to mark intron/exon junctions (73). It is widely accepted that alternative splicing is responsible for proteomic diversity within and among tissue types (74, 75). While alternative splicing is important for cell development and differentiation (72), mutations affecting splicing are believed to be at the root of 15-50% of human genetic diseases (76). A single RNA-seq experiment can not only identify novel and rare transcripts, isoforms and alternative splice sites, but can also quantify those identified (77) without the need of prior knowledge of splice junctions (78). While RNA-seq can detect variants affecting intron-exon junctions, its ability to detect deep intronic variants affecting splicing regulators is null except in the case of intron retention (34). This means that RNA-seq would detect the expression of alternatively spliced transcripts but might miss the causative variant in which case PCR, Sanger sequencing or targeted NGS can be used to sequence the surrounding areas.

There are two ways in which RNA-seq can be used to identify alternative splicing events: a transcript-focus analysis in order to reconstruct the transcriptome and estimate the concentration values of each transcript (43) and an event-based approach focused directly on the analysis of splicing events. These algorithms will categorize the list of different events in five groups: exon skipping, alternative 3’ splice site, alternative 5’ splice site, mutually exclusive exons and intron retention (43). While RNA-seq can accurately identify alternative splicing events, RT-PCR is still often used to validate these findings. Using RNA-seq to identify pathogenic phenomena can facilitate the identification of pathogenic variants and can hence

ameliorate molecular diagnosis (79). Furthermore, the identification of alternative splicing events has led to the development of drugs targeting splicing events, machinery, and variants as well as the development of splicing-associated therapies (80).

### 1.3.5 Allelic imbalance

In general, homologous alleles are expressed equally within a given cell; however, studies on X-chromosome inactivation, genomic imprinting, and random mono-allelic expression have challenged this concept (81). When two alleles of a given gene are differentially expressed, meaning either one allele is lowly expressed in comparison to the other or one allele is not expressed at all, this phenomenon is known as allelic imbalance. Allelic imbalance can be normal such as with X-chromosome inactivation and is even thought to play a role in cellular identity and neuronal diversity (82). However, related phenomena such as loss of heterozygosity where only one of the parental alleles is expressed due to the loss of the other (83), haploinsufficiency where alleles cannot produce sufficient quantities of protein to reach functional biological dosage (84), and even subtle allelic imbalances (85) have all been associated with disease and disease predisposition.

RNA-seq can detect allelic imbalances in two ways. The first method targets heterozygous sites and compares the expression of both alleles. This is done by calculating read-count ratios for the minor-allele. A median ratio is calculated in genes having 5 or more variants, and this value is compared against controls in order to characterize the imbalance. This method has been successfully employed to diagnose a subset of patients with a neuromuscular disease (30). The design of this approach, however, begets certain limitations. Because only heterozygous sites are targeted, a severe allelic imbalance or loss of heterozygosity could be dismissed as a homozygous allele. Furthermore, identifying the causative variant may be challenging if it is located on the lowly or null expressed allele due to RNA-seq's dependence on expression levels. The second approach, while more expensive, overcomes these challenges by combining RNA-seq with WGS or WES. Allelic imbalances can be detected by targeting alleles that are heterozygous at the DNA level and homozygous at the RNA level. With this approach, it is necessary to loosen the allelic ratio parameters used to define homozygosity at the RNA-level in order to capture more subtle allelic imbalances. In the

past, identification of allelic imbalances has contributed to our understanding of variable clinical expressivity and incomplete penetrance and has led to better prognosis and more precise genetic counselling (84, 86). Furthermore, identification of various allelic imbalances in a multitude of diseases has also led to better markers for disease and disease predisposition (85). Finally, the study of allelic imbalance could also provide insight into more complex disease mechanisms such as X inactivation, imprinting, chromosomal rearrangements and uniparental disomy (30).

### 1.4 Functional analyses to validate candidate variants

The American College of Medical Genetics and Genomics (87) provides a set of weighted criteria required for the classification of novel variants into five tiers: pathogenic, likely pathogenic, uncertain significance, likely benign and benign (87). Types of variant evidence include population, functional, segregation, *de novo*, allelic, computational, and predictive data. After identifying a candidate mutation, functional analyses of model systems play a critical role to enhance evidence for the genotype-phenotype relationship. Selection of the most appropriate model will depend on the goals and timeframe of the research project. Notably, the method by which a heritable disease is translated into the model (i.e. CRISPR/Cas9, RNA interference (RNAi), transgenics) will have a significant impact on research time.

Typically, the most important criterion is the recapitulation of physiological conditions observed in affected humans in order to optimize representation. *In vitro* cell cultures, particularly those derived from the affected patients under study, are the most genetically similar models available, pose no risk of death or discomfort to animals, and have the potential to be high throughput, but lack the biological ultrastructure of living models. It is important to consider the expression pattern of a gene and the cellular function under study when selecting the model cell type (88). A challenge working with *in vitro* models is the capture of all the main pathological features of a disorder in an environment very different from that of the human body. For instance, Kraeva *et al* found that a physiologically representative model of the heterozygous mutation (p.R4892Q) associated with central core disease was only achieved when patient RYR1 was expressed in RYR1-/- (dyspedic) myotubes and not RYR1+/- myotubes, despite the dominant nature of the mutation in humans (89).

A severe disease phenotype may also facilitate observation of minute changes in response to experimental conditions. The dyspedic mouse, containing two disrupted alleles in *RYR1* (ry153/ry153) is the most widely reported model for RYR1-RMs within the past decade (90). These RYR1-null mice demonstrate a birth lethal defect, lack of excitation-contraction coupling, abolishment of the contractile response, and gross abnormalities of the skeletal muscle (91) and have greatly contributed to the characterization of RYR1 structure and function. However, RYR1-RMs in humans are typically much less severe with malignant hypothermia, a sensitivity to medication used during general anesthesia, being one of the mildest conditions. Similarly, the golden retriever muscular dystrophy canine is a naturally occurring model of DMD with more similar etiology and clinical features to humans compared to the more commonly used Mdx mouse model (92). The canine model has been particularly useful in identifying adverse side effects to treatments that did not present in mice, and its large size, outbred nature, and very apparent muscle weakness facilitates observation of scalable variables such as cell proliferation or the immune response (92). Generally generation time, brood size, and body size along with the increased expense of purchasing and maintaining larger animals restricts the number of available test subjects which in turn limits the number of testable variables and the power of statistical analyses (92).

Despite being farther on the evolutionary tree, non-mammalian animals who share a high level of genetic conservation with humans are also frequently used as models for human disease. Zebrafish are longstanding vertebrate models for developmental studies due to their transparency during the embryonic and larval stage, rapid development, affordability, and ease of care. These fish have an orthologous gene for about 70% of known human genes (93), and their genetic tractability through gene knock-down, knock-out or overexpression has accelerated our understanding of the origin, composition and consequences of abnormal histological structures (94), abnormal muscle function (95) and efficacy of new and pre-established treatments for muscle diseases (96).

As more and more variants are discovered, functional analysis is ever the rate limiting step for both discovery and diagnosis (9). Simple organisms play an increasingly important role as first-pass models when testing for pathogenicity of novel variants. Farther down the

evolutionary tree, invertebrate models like the *Caenorhabditis elegans* (*C. elegans)* nematode have very short generation times and lifespan of about 2 and 21 days respectively allowing for even more rapid experimental turnaround time at any stage of development. The use of *C. elegans* is advantageous for several reasons. First, *C. elegans* are hermaphrodites and can thus mate with themselves, typically producing large broods of offspring in the expected Mendelian ratios (97). Generally, these worms prefer to self-reproduce allowing for the easy propagation of disease mutations and of worms unable to mate due to a mutation or otherwise (98). Second, the optical transparency of the body wall facilitates visualization of the internal structures by microscopy (98). Third, despite the evolutionary distance, *C. elegans* possess 95 striated body wall muscles with highly structurally, compositionally, and functionally conserved sarcomeres that act as the equivalents of vertebrate skeletal muscle (98). These tissues lack satellite cells for muscle regeneration and remain mononucleated rather than fusing together making for a simple and attractive model of muscle tissue, especially for hereditary myopathies (98). Finally, in the case of variant discovery, the time lost and financial fallout is minor compared to most other models should the pathogenicity hypothesis be disproven. While *C. elegans* have many quantifiable phenotypes used to interpret the variant effects, the biological relevance of these parameters is not always known.

### 1.4.1 The genetic tractability of *C. elegans*

Already *C. elegans* have contributed to the discovery of several important cellular processes including the role of caspases in programmed cell death (99), the first microRNA and its mRNA target (100, 101), and the first genes to mediate axon guidance (102). In fact, *C. elegans* were the first organism to have all 302 neurons and their synapses mapped in completion (103). These animals have also contributed to the discovery of important genetic techniques such as RNA interference (RNAi) (104) and the use of green fluorescent protein (GFP) as a biological marker for gene expression (105). These techniques are still widely used today. Currently, *C. elegans* promise much potential as a model animal capable of rapid and coordinated testing of novel genetic variants and potential therapeutic strategies. Indeed, personalized models of human pathogenic variants represented in *C. elegans* have applications in evaluating the efficiency of drugs at various doses, detecting negative side effects, identifying

affected pathways of a drug response, and informing the basis for future pharmacological treatment. The value of *C. elegans* as models for human disease stems from their genetic tractability owing to an abundance of available genetic tools for both forward and reverse genetics (106).

For example, RNA interference is a sequence-specific technique developed in *C. elegans* that relies on an antisense mechanism for gene silencing (104). Double stranded RNA (dsRNA) is known to be more effective and selective than single stranded RNA (ssRNA) (104), and can be introduced into the worm by microinjection, feeding, or soaking (107). The feeding technique capitalizes on the use of an RNAi competent OP50 strain of *E. coli*, HT115(DE3), which contains two important genetic modifications that allow for the production and maintenance of dsRNA (108). First, an IPTG-inducible T7 RNA polymerase allows for the generation of dsRNA from dsDNA in the presence of IPTG (109). Second a deletion in the RNase III (*rnc)* allele prevents dsRNA from being degraded (108). Therefore, knock-down of a target gene is achieved by cloning target cDNA in between the aforementioned phage T7 polymerase promoters (107), transforming the vector into the HT115 *E. coli* strain, and then feeding the bacteria to the worms. When dsRNA is ingested by the worm, it is bound by the worm protein RDE-1 and cleaved of its passenger strand to form single stranded small interfering RNA (siRNA). Afterwards recruitment of the RNA-induced silencing complex (RISC), which includes RDE-1, to the siRNA enables binding and cleavage of targeted endogenous mRNA (110). Fortunately, an RNAi library of bacterial strains corresponding to 86% of *C. elegans* genes has been established and is commercially available (111). Furthermore, transgenic strains of *C. elegans* have been established to enable tissue-specific susceptibility to RNAi thus making mosaic analysis of gene function possible. For example, because RDE-1 is instrumental to RNAi, transgenic strains of *C. elegans* have been established to express RDE-1 in the muscle or the hypodermis (112). Similarly, tissue-specific strains have also been established for neuronal cell subtypes (113). These strains are especially valuable when global knock-down of a particular gene masks relevant disease phenotypes by inducing sterility, lethality or other severe pleiotropy (113)

CRISPR/Cas9 gene editing is another important biological tool that was deducted from the discovery of palindromic segments of DNA in *E. coli* now known as Clustered Regularly

Interspaced Short Palindromic Repeats (CRISPR) (114, 115). CRISPRs generally consist of noncontiguous identical repeat sequences separated by variable sequences called spacers that are frequently found adjacent to operons of *cas* genes (116, 117). In 2007, Barrangou *et al.* were the first to demonstrate that CRISPR and *cas* genes are involved in microbial immunity (116). They observed that after exposure to virulent bacteriophages, *S. thermophilus* bacteria incorporate new spacers at the CRISPR1 locus with striking sequential likeness to fragments of viral DNA (coined protospacers (118)) used in the assay. These new spacers granted phage-resistance in a manner that correlated with the spacer sequence fidelity and the number of new spacers introduced. Furthermore, introduction of a novel spacer identical to the protospace resulted in resistance while removal of a spacer resulted in restored viral sensitivity. In 2012, Jinek *et al* deduced the mechanism by which a dual-RNA complex derived from spacers can be used to guide highly-specific Cas9 endonuclease cleavage of protospacer DNA, destroying the virus (117). In a natural context, dual-RNA is composed of two separate RNAs with partial complementarity to support binding to one another. The first component, CRISPR RNA (crRNA), is transcribed from the spacer sequence and is important to cleavage specificity. The second component is t*rans*-activating crRNA (tracrRNA) which plays a role in crRNA maturity in order to enable binding to Cas9. Jinek *et al* concocted a trcrRNA-crRNA chimera known as guide RNA (gRNA) whose spacer sequence could be modified to target virtually any DNA sequence in all cell types including human (117). Thus, an alternative methodology to exploit cas systems for RNA-programmable genome editing was born.

Modern CRISPR/Cas9 applications rely on the presence of an effector enzyme and gRNA. While gRNA can be modified to recognize and pair with any sequence of about 20 nucleotides in length, digestion is dependent on specific sequences at the 3’ end of target DNA known as protospacer adjacent motifs (PAMs) (118). Recognizable PAM sequences are unique to each effector; that most commonly used being the Cas9 derived from *S. pyogenes* (SpCas9) which recognises the PAM sequence NGG. Some Cas9 systems have even been engineered with a nuclear localization sequence to facilitates entry into the nucleus (119). Paired Cas9-gRNA scans the genome for the appropriate PAM sequences before testing for gRNA complementarity. In fact, sequences matching the gRNA but lacking a PAM sequence are ignored thus increasing

specificity and rapidity of target recognition in large DNA molecules (120). Once bound to the target DNA sequence, Cas9-gRNA introduces a double stranded break that must be repaired in order to avoid cell death. Endogenous mechanisms of the cell will repair the break by non-homologous end joining (NHEJ) resulting in the direct ligation of broken strands at risk of introducing insertions or deletions of variable length. Frequently, these indels jeopardize gene function by disrupting regulatory elements or introducing a frame-shift, especially when introduced early in the gene. Thus, the canonical NHEJ pathway is ideal for generating knock-out models (121).

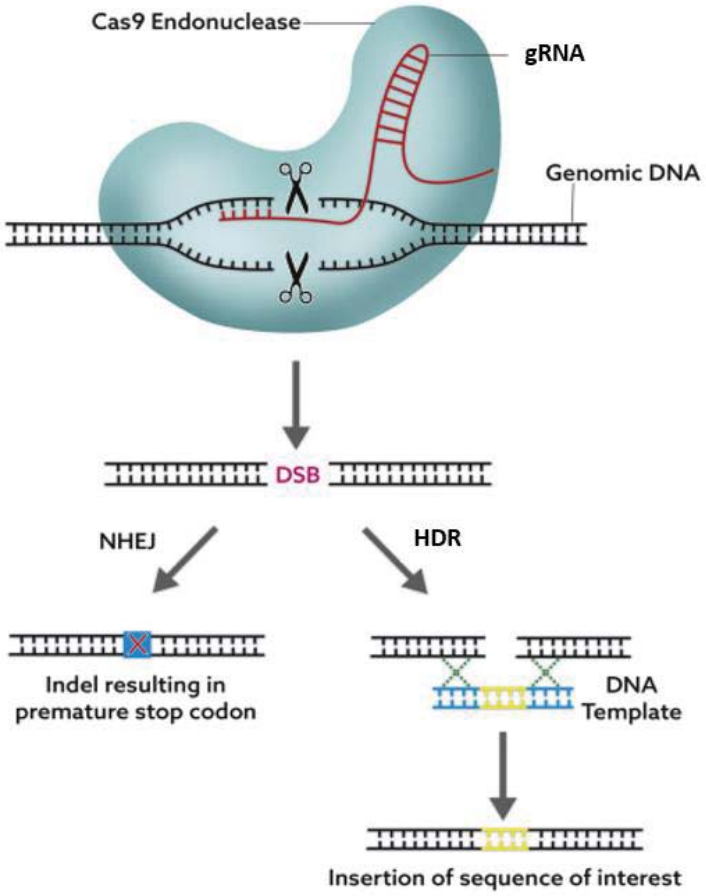



**Figure 2. CRISPR/Cas9 mediated gene editing**
Small guide RNA directs the Cas9 complex to the target site on DNA. The complex makes a double stranded break and pursues one of two pathways. In the absence of a DNA template, non-homologous end joining (NHEJ) introduces an indel mutation that often results in a frameshift mutation. In the presence of a DNA template, homologous directed repair introduces a desired sequence to form recombinant DNA. This image was adapted from previously published work (122).

The homology directed repair (HDR) pathway relies on 5'-3' sectioning of the double-stranded break allowing for the formation of single stranded overhangs (123). Sectioning allows for a complementary DNA sequence to bind forming a repair template for DNA polymerase to synthesize the missing parts. Therefore, modification of donor DNA allows for insertion of desired mutations into the genome of the host. Hence the HDR pathway is ideal for generating precise knock-in models (Fig. 2). In general, HDR has a very low efficiency compared to the NHEJ pathway and requires cells to be in a specific phase of the cell cycle (124). Thus, many subjects must be screened in order to identify instances of a successful CRISPR/Cas9 experiment. In *C. elegans*, identification of successful editing is facilitated by co-CRISPR of off-target dominant mutations with palpable phenotypes such as rolling (125) or protein fluorescence (126). For reasons unknown, about 40% of broods yield no edits while 20%, known as "jackpot broods", demonstrate striking editing efficiency (20-60%) (126). Successful edits of the gene of interest, whether paired with the visual phenotype or not, are only found in broods that demonstrate the off-target phenotype (126). Therefore, targeted screening of the siblings of animals carrying the palpable phenotypes within jackpot broods greatly narrows the search for desired mutants.

# Chapter 2: Hypothesis and Objectives

While gene panel and next-generation DNA-sequencing techniques have dramatically accelerated gene discovery for rare hereditary myopathies many remain without a genetic cause. This represents a significant burden to both patients and the healthcare system. Therefore, our primary aim is to increase the diagnostic yield of rare hereditary myopathies by applying RNA-sequencing to uncover disease-causing variants and genetic phenomena in individuals who lack a genetic diagnosis despite exhaustive investigation using traditional approaches. Our secondary aim is to validate the pathogenicity of these variants in *C. elegans* models. By these means we have identified and validated the pathogenicity of a muscle-specific allelic imbalance of *IARS* (isoleucine tRNA synthetase) in favour of a mutant allele and of two variants in *RYR1* in the nematode *C. elegans*.

## 2.1 Allelic imbalance in *IARS* (*c.19826G>A)*

We present three siblings with a mild congenital muscular dystrophy. Notably, both parents are unaffected suggesting an autosomal recessive inheritance. Using a combination of RNA-seq and WES we have also uncovered a missense mutation on the maternal allele (c.19826G>A) and identified a muscle-specific allelic imbalance in the gene *IARS.* This imbalance results in a loss of the paternal allele and limited expression of the mutant allele carrying the missense variant (maternal allele). Our objective is to validate pathogenicity of this knock-down in *C. elegans* before proceeding with further studies in more complex and expensive models. Therefore, we hypothesize that RNA interference mediated knock-down of the worm homolog *iars-1* in the muscle of *C. elegans* will result in a muscle-related phenotype.

## 2.2 RYR1 variant (c.526G>A)

We present a cohort of 21 individuals who all carry the same rare novel heterozygous, missense variant c.526G>A in the gene *RYR1*, yet display extremely diverse muscle phenotypes. This cohort is ideal to study the molecular basis of clinical heterogeneity and so our goal is to first confirm pathogenicity of the variant. Assessing the pathogenicity of this variant is extremely important considering the highly polymorphic nature of *RYR1.* Therefore, we

hypothesize that a CRISPR/Cas9 mediated knock-in of this variant in the worm homolog *unc-68* in *C. elegans* will result in a muscle-related phenotype.

### 2.3 RYR1 variant (c.12083C>T)

We present a forty-seven year old woman and her son who suffer from a static myopathy with proximal and distal weakness since a young age. Gene panel results were negative, and so our objective is to identify a candidate disease-causing variant and validate pathogenicity. Using RNA-seq we identified a pattern of differential expression that was suggestive of a fiber-type disproportion. We also discovered an *RYR1* variant (c.12083C>T) that went unreported by the gene panel despite the gene having already been associated with a vast variety of myopathies. Therefore, we hypothesize that this *RYR1* variant is causative of the patients' phenotypes and that a CRISPR/Cas9 mediated knock-in of this variant in *unc-68* in *C. elegans* will result in a muscle-related phenotype.

# Chapter 3: Functional characterization of *IARS* knock-down in *C. elegans* and cellular models.

## 3.1 Abstract

Congenital muscular dystrophy (CMD) is an umbrella term for many heritably degenerative diseases leading to the progressive break down of skeletal muscle tissue starting from infancy. Histological signatures on muscle biopsy often include dystrophic lesions (127) and affected individuals may experience muscle weakness, delay of motor development, and joint or spinal rigidity (128). There are currently at least 12 known genes associated with the various subtypes of CMD (129). However, there exists still a subset of patients who are clinically and histologically diagnosed with CMD, but do not carry mutations in these genes. Three siblings, for whom exome sequencing were inconclusive, presented with a mild congenital muscular dystrophy (3/3) and scoliosis (2/3). Observation of muscle biopsy tissue revealed atrophic polygonal fibers. RNA-sequencing was carried out on muscle tissue from one affected, and when combined with data from whole-exome sequencing a missense mutation on the maternal allele (c.19826G>A) and a muscle-specific allelic imbalance in the gene *IARS* was revealed. This imbalance results in a loss of the paternal allele. The pathogenicity of this variant was assessed using a *C. elegans* model with muscle-specific vulnerability to RNA interference. Thus, small interfering RNA encoding a complementary sequence to *IARS* was used to introduce a muscle-specific knock-down which lead to the observation of a progressive disorganization of the body wall muscle tissue compared to a control. Otherwise, lifespan, paralysis rate, and motility appeared unaffected.

## 3.2 Introduction

Protein synthesis is the coordinated process by which a strand of mRNA is translated into a polypeptide chain that gains mechanical and chemical functionality upon taking up a very specific three-dimensional conformation. During translation mRNA moves from the nucleus into the cytosol where it binds to a ribosome composed of two major subunits, coined the small and large subunit, connected by intersubunit bridges (130). The small subunit is responsible for decoding the mRNA by matching codons on mRNA with the correct anticodons found on

aminoacylated transfer RNA (tRNA), while the large subunit catalyzes the bond between peptides (131). A polypeptide chain samples many configurations while in a random coil until many chaperones and folding proteins assist in achieving its final conformation. All proteins demonstrate incredible specificity, being able to bind only a few ligands out of the thousands of possibilities, owing to the kinetics and thermodynamics of distinct patterns of amino acids along the polypeptide chain that are brought together to build ligand binding sites after folding. The affinity of a ligand to a binding site is dependent on the formation of many weak non-covalent bonds which in turn is based on the conformational fit of the two molecules. Disulphide bonds are also important for added stability and tolerance to environmental conditions in many structural proteins (132). Thus, single amino acid perturbations in any part of the polypeptide chain can have dire effects on the structure of the protein potentially jeopardizing its function.

### 3.2.1 The role of cytosolic isoleucyl-tRNA synthetase

Because proteins play such a critical role in all processes and reactions that occur within living cells, the fidelity of protein translation from mRNA is invaluable to survival. Aminoacyl-tRNA synthetases (aaRSs) are a group of cytosolic ligating enzymes whose main function is to translate the universal genetic code by catalyzing the aminoacylation of tRNAs by their cognate amino acid. In the first step of this reaction an aaRS will bind both the cognate amino acid and ATP in the active site, expelling pyrophosphate and forming the intermediate aminoacyl-adenylate (aa-AMP) through nucleophilic attack (133). In the second step, tRNA binds to the anticodon binding domain of the aaRS to receive the amino acid and release AMP in order to form aminoacyl-tRNA which is now ready to participate in protein synthesis. With the exception of lysine, exactly one aaRS exists for each of the 20 human tRNAs (133). Notably, both mitochondria and chloroplasts typically encode their own aaRS and tRNA pairs that resemble most those of bacteria(133). Thus the correct translation of nucleotide triplet to amino acid during protein synthesis is heavily reliant on the precise structures and interrelationships of the different sets of tRNA and aaRS (134).

aaRSs can be classified into two subgroups that are primarily distinguished by the structure of their catalytic domain. Class I aaRS have a catalytic domain composed of a

Rossmannoid fold of five parallel β-sheets (135) whereas that of Class II aaRs is composed of an seven antiparallel β-sheet fold flanked by alpha-helices (136). However, over time other differences between the two Classes have also been elucidated. By nature of the Rossmannoid fold, Class I synthetases share two signature motifs (HIGH and KMSK) (133) while Class II share three (137, 138). Furthermore, most Class I enzymes bind to the minor groove of the tRNA acceptor stem and aminoacylate the 2'-OH of the A76 terminal ribose rather than bind to the major groove and aminoacylate the 3'OH as observed in most Class II aaRSs (137-139). Moreover, ATP binds to Class I and II either in a straight or bent configuration, respectively (133, 137). The rate limiting steps also differ in that class I is limited by aminoacyl-tRNA release and Class II is limited by the activation of the amino acid (133). Finally, tRNAs serviced by Class I aaRSs generally bind large amino acids with low polarity while Class II enzymes generally service tRNA that bind smaller, more polar amino acids (140).

Isoleucyl-tRNA synthetase (IARS) functions to bind isoleucyl to its cognate tRNA and is one of the twenty enzymes belonging to Class I aaRSs. Both classes of aaRSs share a catalytic core domain (CCD) that activates the amino acid, and an anticodon-binding domain (ABD) that binds the anticodon of cognate tRNA (141). However, Lo *et al* found that in-frame splice variants found across human AaRS have been shown to disrupt or ablate the CCD and retain noncatalytic domains forming natural isoforms known as catalytic nulls (142). These null transcripts appear to be biologically functional meaning the role of AaRS may be more diverse than what is currently known. Indeed, a variety of cell-based assays revealed that of the variants containing catalytic nulls, 46% play a role in transcriptional regulation, 22% in cell differentiation, 10% in immunomodulation, 10% in cytoprotection, 4% in adipose and cholesterol transport, 4% in proliferation, and 4% in inflammation (142). It is worth noting that only one catalytic null was found for the *IARS* gene. IARS is also known to possess editing activities that allow it to hydrolyse mismatched tRNAs and prevent mistranslation(143). Specifically, IARS from various *E. coli* strains is able to form Ile-AMP, Val-AMP, and Leu-AMP intermediates; however, only Ile-AMP reacts with $tRNA^{Ile}$ to form Ile-tRNAIle, whereas the other aa-AMP structures are hydrolyzed. Eldred and Schimmel also confirmed that enzymatically

synthesized mischarged species of Val-tRNA$^{Ile}$ but not Phe-tRNA$^{Ile}$ are rapidly deacylated by IARS.

### 3.2.2 Disease phenotypes

Disease phenotypes associated with *IARS* have only recently been discovered and so not much is known about the pathomechanisms involved. In 2013, the homozygous missense mutation c.235G>C, p.V79L located in the CCD was segregated by homozygosity mapping and linkage analysis in Japanese Black cattle with perinatal weak calf syndrome (144). Many of the cattle displayed neonatal weakness and intrauterine growth retardation leading to increased mortality rates. Other clinical symptoms also included anemia, depression, variable body temperature, astasia, weakness, difficulty nursing and increased susceptibility to infection. Additionally, the aminoacylation activity of mutant IARS protein was reduced by 38%. Shortly after this study, ten compound heterozygous mutations in IARS were associated with a multisystemic disease characterized by growth retardation, impaired intellectual development, hypotonia and hepatopathy (GRIDHH MIM:617093) in five unrelated human patients across several different studies (14, 145-147). In these cases, mutation of the *IARS* gene resulted in one loss of function allele and one hypomorphic allele. Of the ten human mutations identified five were found in the CCD and one was found in the ABD (146).

Most recently, a homozygous missense mutation (c.290A>G) causing IARS deficiency has been identified as the cause of refractory early-onset inflammatory bowel disease in a young boy with neonatal liver failure, deranged clotting, transaminitis, cholestasis, and pancolitis (148). There also exists a rare chronic autoimmune disease called antisynthetase syndrome (anti-SS) caused by autoantibodies that mistakenly attack aaRS and damage healthy tissues. The primary phenotypes of this disorder include interstitial lung disease, arthritis and myositis, though certain autoantibodies are more likely to cause specific symptoms. Interestingly, within a Japanese cohort of patients with anti-SS, muscular and extramuscular symptoms between patients was relatively homogenous with the exception of those with the autoantibody targeting IARS (anti-OJ) where severe muscle weakness was much more prominent (149). The most common clinical presentation of cytosolic IARS deficiency are abnormalities of the central nervous system, facial dysmorphisms, dysmaturity, failure to thrive, poor feeding, and liver

failure among less common features such as abnormalities of the senses, kidneys, and skin (150). This diversity of phenotypes may be attributable to insufficient aminoacylation activity to meet translational demand resulting in reduced synthesis of related enzymes in specific organs or periods of life (150, 151).

### 3.2.3 Congenital muscular dystrophies

While IARS disorders are entirely multisystemic, the majority of cases demonstrate muscular symptoms in the form of either weakness or hypotonia. Nonetheless, never before has IARS been associated with a congenital muscular dystrophy. CMDs are a group of heritable degenerative diseases of the skeletal muscle characterized by progressive muscle wasting with variable severity depending on age of onset and the muscle groups involved. Affected individuals may experience muscle weakness, delay of motor development, and joint or spinal rigidity (128). Occasionally, the heart and diaphragm muscles are also involved during the advanced stages of the disease after the patient's mobility has been significantly reduced ultimately jeopardizing the safety of the patient (152). Diagnosis of CMD is typically based on clinical findings, brain and muscle MRI, muscle and skin biopsy histology and/or immunochemistry, and genetic testing (153).

Most CMD follow an autosomal recessive mode of inheritance and the associated genes generally encode proteins that provide support to the muscle structure. These genes can be separated into three groups (154). The first group encompasses genes encoding proteins of the basement membrane or extracellular matrix. For example, the genes associated with Bethlem and Ullrich CMD, *COL6A1*, *COL6A2*, and *COL6A3*, play an integral role in the formation of type VI collagen. This collagen makes up part of the basal lamina of the extracellular matrix that surrounds each myofiber to provide structural support (155). The second group contains genes encoding proteins involved in dystroglycan glycosylation. A clear example of group two is the causative gene of Fukuyama CMD, *FKTN*, which is involved in the glycosylation of alpha-dystroglycan (α-DG) by N- and O-linked glycans (156). Glycosylated α-DG and β-DG form a dystroglycan complex that binds the extracellular matrix to the actin cytoskeleton within the sarcolemma (156) providing mechanical stabilization as well as playing a role in mediating interactions between the cytoskeleton and extracellular matrix (157). Finally, the third group

incorporates CMD caused by defective selenoprotein 1 (*SEPN1*): a selenocystein-containing protein located on the endoplasmic reticulum known to regulate oxidative stress and may be involved in excitation contraction coupling of the muscle (158). Rigid spine muscular dystrophy, which is associated with symptoms like scoliosis, muscle weakness of the axial muscles and limbs, as well as early-onset respiratory insufficiency, is the only CMD in the third group (159). Because genes associated with CMD typically affect structural proteins, the normally highly organized structure of muscle tissue is often disrupted. Indeed pathology of muscle tissue from CMD patients may involve split, whorled, or hypercontracted muscle fibers, increased internal nuclei, cores or minicores (153).

We present three siblings with a mild congenital muscular dystrophy in which a muscle-specific allelic imbalance was discovered in the gene *IARS* along with a maternally inherited mutation (c.19826G>A) located in the ABD. This imbalance results in a loss of the paternal allele. Because both parents are unaffected, we assumed recessive inheritance meaning that a second variant, resulting in the degradation of the paternal allele, remains elusive. This mode of inheritance agrees with what has been previously observed in *IARS*-associated disorders. However, unlike the multisystemic disorders reported in the literature our patients display clinical symptoms isolated to the muscle and spine likely owing to the tissue-specificity of the imbalance. Functional impact of IARS knock-down was assessed in *C. elegans* models in which progressive disorganization of the muscle tissue was observed. However, no significant difference in lifespan, rate of paralysis, or motility was observed between mutants and wild type (WT) worms. Therefore, the c.19826G>A seems to have functional consequences for muscle tissue structure in *C. elegans*. However, further investigation is required to validate its pathogenicity in humans.

## 3.3 Methods

### Patients and genetic testing

All three affected individuals and their parents were evaluated by experienced neurologists. Two siblings (II:1 and II:2) had a muscle biopsy has part of her clinical workup. The study was approved by the McGill University Health Centre Research and the CRCHUM Ethics

Boards. All participants have signed an inform consent authorizing genetic analysis in a research setting.

**Whole Exome Sequencing**

Genomic DNA from each of the three patients and their parents was extracted from peripheral blood lymphocytes using standard procedures and sent to the Genome Quebec Innovation Center (Montreal, Canada) for whole exome sequencing as previously described (160). DNA enrichment was performed using the Agilent SureSelect Human all Exon 50 Mb capture library (v4) and sequenced was performed on an Illumina HiSeq2000 sequencer using 90bp paired-end reads. FastQ files were aligned to the reference genome (Hg19) using BWA (161) and variant calling was performed with GATK (162). Variants were filtered according to minor allele frequencies, leading to exclusion of variants with a MAF >1% in dbSNP (163), The 1000 Genomes Project (48), Exome Variant Server (164), and internal exome databases. Data processing and computational analysis of WES data was performed by Martine Tétreault.

**RNA-Sequencing**

RNA was extracted from snap-frozen muscle tissue from patient II:1 using trizol (Invitrogen). Library generation was performed using TruSeq stranded mRNA library preparation kit (Illumina, San Diego, CA USA), and sequencing was performed on an Illumina HiSeq 2500 using 125bp paired-end reads and 0,25 lanes of data per sample.

Data analysis was performed using an in-house pipeline combining publicly available tools and custom scripts. FASTQ files were aligned to the reference genome (Hg19) using STAR (165) and variant calling was performed with GATK and annotated using ANNOVAR as well as custom scripts (46, 162). This pipeline allowed for the detection of multinucleotide variants (MNV), indels, and single nucleotide variants including synonymous, non-synonymous and splice junction variants. Variants were filtered according to minor allele frequencies, excluding variants with a MAF >5% in either The 1000 Genomes Project (48) or The GnomAD (49). The data was further filtered to keep protein-damaging variants (nonsense, missense, frameshift, indels and splice variants).

Finally, Sanger sequencing was performed on RNA extracted from lymphoblasts of patient II:1 as well as from fibroblasts of II:2, II:3 and the father to confirm the presence of the exonic *IARS* variant in the affected and to provide evidence for muscle-specificity of the allelic imbalance. The following forward and reverse primers were used: 5'-GGCTTGTCCTGGCAAGTGAT-3' and 5'-ACAGGATCCACCGGTCTGTA-3', respectively.

***C. elegans* strains**

Nematodes were cultured and handled as per standard methods (97). The strains used include N2, NR350 *rde-1*(*ne219*); *kzIs20*, DM8005 *raIs5*[*myo-3p::GFP::myo-3 + rol-6*(*su1006*)], and CB307 *unc-47*(*e308)* which were obtained from the *Caenorhabditis* Genetics Center (CGC). All experiments were performed in triplicates, with the exception of the microscopy experiment, which was performed in quadruplicate, at room temperature (22°C) using age-synchronized adult worms. Missing or bagging worms were censored from all experiments. *C. elegans* lost during experiments for reasons other than senescence (i.e., missing, bagging, or ruptured vulva) were censored.

**RNAi strains**

The RNAi *E. coli* strain HT115 carrying *iars-1, unc-89,* or an empty vector (EV) within the L4440 plasmid was a gift from Dr. Alex Parker. In order to confirm sequence complementarity or noncomplementarity, *iars-1* RNAi was purified using a QIAprep Spin Miniprep Kit and sent for Sanger sequencing at Genome Quebec using the standard L4440 reverse primer (5'-AGCGAGTCAGTGAGCGAG-3') (Sup. Fig. 1). Standard NGM plates were prepared as previously described (97) with the addition of 10ml of nystatin and 1ml of streptomycin. RNAi NGM plates were prepared similar to standard NGM plates but with 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 1mM ampicillin instead of streptomycin.

**Synchronization**

Ten to fifteen gravid adult worms were placed on a fresh plate seeded with the appropriate RNAi or empty vector, left overnight to lay eggs, and removed the following day. These progenies underwent the same process until two consecutive generations of worms born on RNAi was obtained. Thereafter, a pool of L4 animals was placed on fresh plates. The

following day, day 1 adult worms were picked from this pool to start new experiments. Exceptionally, DM8005 worms cultured on *iars-1* RNAi are first-generation animals due to the fact that global knockdown of *iars-1* results in sterility and embryonic lethality (166, 167).

**Paralysis assay**

35 day 1 adult hermaphrodite NR350 worms were transferred to three *iars-1* RNAi or three EV plates and scored daily for paralysis in a binary fashion. Worms were considered paralyzed if, after prodding by a platinum wire, they displayed restricted movement isolated to the head area. 153 and 117 events were counted for *iars-1* RNAi and EV respectively.

**Lifespan assay**

35 day 1 adult hermaphrodite NR350 worms were transferred to three *iars-1* RNAi or three EV plates and scored in a binary fashion. Worms were prodded with a platinum wire and considered dead if they failed to move. Lifespan is defined as the period from day 1 of adulthood until death. 173 and 119 events were counted for *iars-1* RNAi and EV respectively.

**Microscopy**

25 adult hermaphrodite DM8005 worms cultured on *iars-1* RNAi, EV, *unc-89* RNAi or OP50 *E. coli* were immobilized with 0.05% sodium azide and mounted onto slides with 2% agarose pads. Muscle integrity in each worm was scored as having low, medium, or high levels of disorganization on days 1, 3, and 5 of adulthood. All assays were carried out using a Zeiss Axio Observer inverted microscope.

**Motility assays**

35 day 1 adult hermaphrodite NR350 worms cultured on *iars-1* RNAi or EV were transferred to 3 unseeded RNAi plates and video recorded crawling for 30 seconds using a digital stereomicroscope on the lowest magnification (S9i; Leica, Heerbrugg, Switzerland; equipped with 10x/23 eyepieces and an integrated 10 M.P camera with 1080HD video display). Worms were subsequently suspended in 150µL of M9 buffer and video recorded swimming for another 30 seconds before being transferred back to plates seeded plates with the appropriate media. This process was repeated with the same worms for days 1, 5, 7 and 9 of adulthood.

132-252 worms grown on *iars-1* RNAi and 103-258 N2 worms grown on EV were filmed in the crawling assays. Similarly, 139-244 worms grown on *iars-1* RNAi and 106-238 worms grown on EV were filmed in the swimming assays. Notably, the number of worms recorded on each day decreased as worms died over time. Videos were uploaded onto WormLab software (Update 2020.1.1, MBF Bioscience, VT, USA) and tracked using parameters described in Sup. Table 1 in order to generate a track summary of standard metrics described in the WormLab User Guide Version 2020 (https://www.mbfbioscience.com/help/wormlab/Content/home.htm). Results were exported to Graphpad Prism 9 and analyzed using two-way ANOVA followed by Turkey multiple comparisons test.

**Statistics**

For paralysis, aldicarb, and lifespan assays Kaplan-Meier curves were generated using Graphpad Prism 9 and statistical significance was analysed using the Log-rank (Mantel-Cox) statistical test. For motility assays, video tracking results were exported to Graphpad Prism 9 and analysed using two-way ANOVA followed by Turkey multiple comparisons test.

## 3.4 Results

### 3.4.1 Clinical presentation

Our study includes three siblings diagnosed at the age of five with a mild congenital muscular dystrophy born of healthy non-consanguineous parents (Fig. 3a, Table 1). All three patients present with distal laxity, contractures, and affected gait although all have retained the ability to walk. Clinical workup in 2009 concluded that II:1, II:2, and II:3 each displays abnormal pulmonary function with a vital capacity (VC) of 53, 77, and 66%, respectively. Patients II:1 and II:3 also affected with scoliosis. Analysis of muscle tissue derived from patient II:1 conveyed the presence of atrophic polygonal fibers and rare normal size nuclei (Fig. 3b); however, whether this feature is present in the other two affected patients is unknown. Finally, none of the patients demonstrate involvement of the central nervous system.

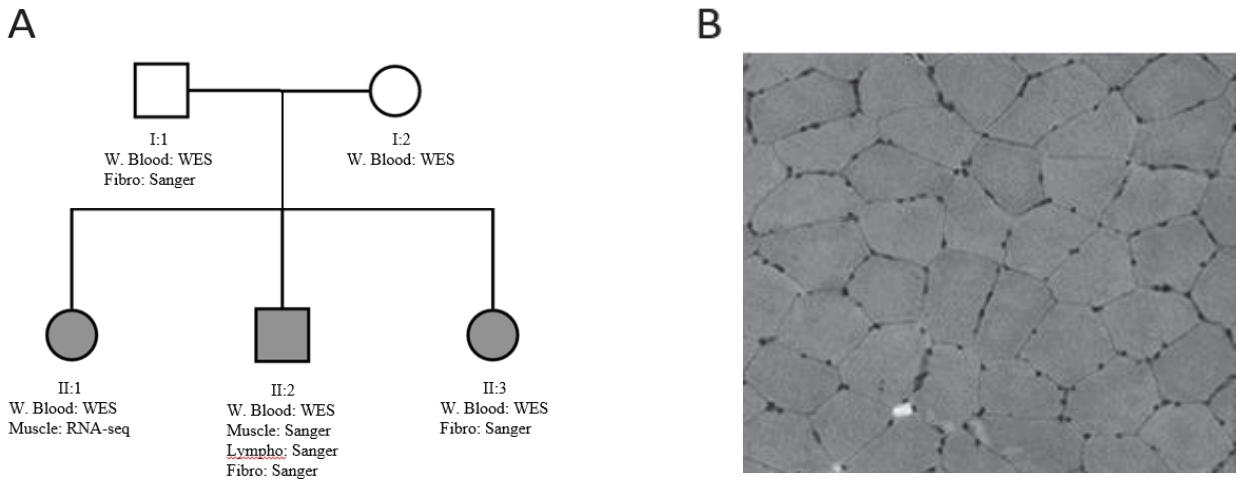



**Figure 3. Clinical evaluation of a family with congenital muscular dystrophy**

**(A)** Pedigree of three siblings affected with congenital muscular dystrophy. Below each individual is a description detailing their ID code and any genetic tests performed in respect to the biomaterial collected. W. Blood= Whole blood, Fibro= Fibroblast, Lympho=Lymphoblast. **(B)** Muscle biopsy from patient II:1 depicting atrophic polygonal fibers.

**Table 1. Clinical presentation of three siblings with congenital muscular dystrophy**

| Patient | **II:1** | **II:2** | **II:3** |
|---|---|---|---|
| Sex | female | male | female |
| Age at diagnosis | 5 | 5 | 5 |
| Age in 2009 | 28 | 21 | 23 |
| Distal laxity | + | + | + |
| Contractures | + | + | + |
| CNS involvement | - | - | - |
| Scoliosis | + | - | + |
| Ability to walk | + | + | + |
| Lost ability to walk | - | - | - |
| Abnormal pulm. f. | + | + | + |
| %VC at age | 53 | 77 | 66 |
| Immortalize lymphocyte* | Y | Y:Sanger | Y |
| Muscle biopsy available* | Y:RNA-seq | Y:Sanger | N |
| Skin biopsy available* | N | Y:Sanger | Y:Sanger |

* Y=yes or N=no followed by the molecular assay pursued using these materials if any. Ability to walk refers to whether or not an individual was able to learn to walk during childhood. Lost ability to walk refers to whether or not an individual has lost their ability to walk as a result of their condition.

### 3.4.2 Combinatorial use of RNA-seq and WES reveals a muscle-specific allelic imbalance in *IARS*

WES was performed on the three siblings and the parents failed to reveal any candidate variants of interest. Because both parents were asymptomatic, a recessive mode of inheritance was expected. However, no variants in accordance with an autosomal recessive mode of inheritance and segregating with the phenotype were identified. Therefore, we initiated RNA-sequencing derived from muscle biopsy from patient II:1 in order to probe for functional evidence of aberrant genetic conduct. Alone, RNA-seq also failed to reveal new candidate variants that abided by our expectation of recessive inheritance. However, by combining WES and RNA-seq datasets, we identified an undocumented allelic imbalance and a heterozygous non-synonymous point mutation (C.1982G>A, p.R661H) in the ABD of the gene *IARS*. This imbalance results in a loss of the paternal allele in an approximate 1:9 ratio. Indeed, WES data demonstrated that on a genomic level, the mother and all three children were heterozygous for the identified variant in *IARS* while the father was wild-type (Fig. 4a). Conversely, RNA-seq data suggested that at the level of the transcriptome, patient II:1 was homozygous for the identified variant (Fig. 4b). We validated these findings using Sanger sequencing of muscle and lymphocytes from patient II:2 and fibroblasts from patients II:2, II:3 and the father, which not only confirmed the allelic imbalance but was suggestive of muscle-specificity (Fig 4c). The near absent expression of the paternal wild-type allele mimics a recessive disease (Fig. 4d). Therefore, we hypothesized that investigation of the effects of *IARS* loss-of-function (LoF) could provide clues as to the genetic etiology of the muscle disorder observed in our patients.

### 3.4.3 Global knock-down of *iars-1* leads to progressive disorganization of the body wall muscle.

Approximately 83% of the *C. elegans* proteome contains genes homologous to humans (168). The human IARS protein (NP_001365503.1) shares 56% identity with the nematode equivalent iars-1 (NP_001366774.1) (169). Similar to how the p.R661H mutation is novel in humans, according to the Million Mutation Project only 26 non-synonymous mutations have been identified in *iars-1* across all available worm strains, none of them being the homologous equivalent to the human mutation (Sup. Fig. 2). This only attests to the rarity of the mutation and the high level of conservation of this position. Sanger sequencing of iars-1 RNAi revealed 99% identity with the iars-1 gene (Sup. Fig. 1), representing a high-level sequence

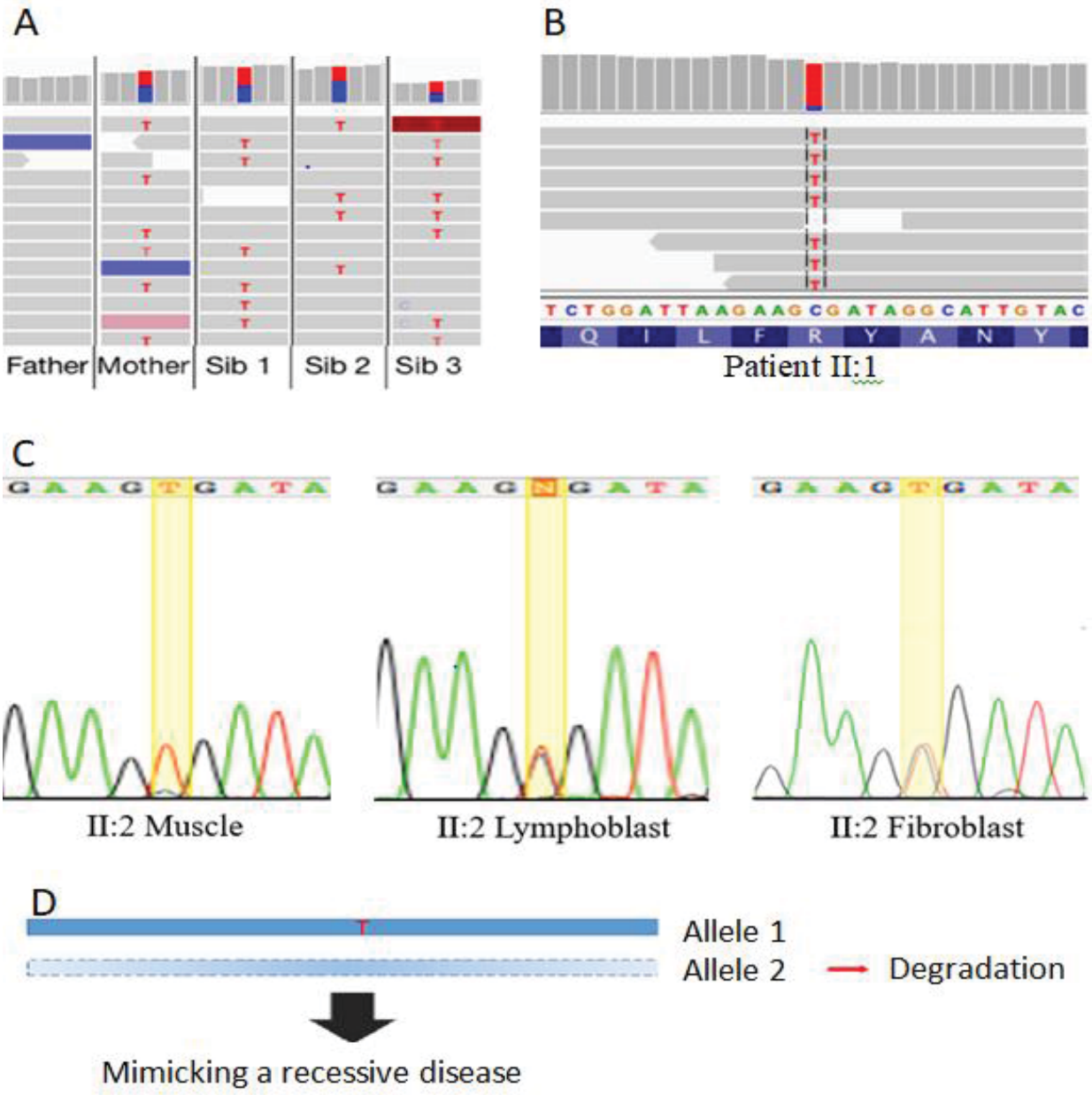



**Figure 4. WES and RNA-seq lead to the discovery of a muscle-specific allelic imbalance**
**(A)** Whole exome sequencing of whole blood from each family member indicates that the mother and all three children are heterozygous for the c.19826G>A variant at the level of the genome. **(B)** RNA-seq of muscle tissue from patient II:1 suggests homozygosity for the mutant allele at the level of the transcriptome. Conflicting zygosity between WES and RNA-seq data is suggestive of an allelic imbalance. **(C)** Sanger sequencing of muscle, lymphoblast, and fibroblast RNA from patient II:2 not only confirms the allelic imbalance but is also indicative of muscle-specificity. Sanger sequencing of fibroblast RNA from patient II:3 not shown but is like that of patient II:2. **(D)** A model for recessive inheritance. Degradation of the paternal wt allele leads to mono-allelic expression of the mutant allele mimicking a recessive mode of inheritance.

complementarity as required to induce a knock-down of iars-1 in C. elegans. These animals possess 95 striated body wall muscles with highly structurally, compositionally, and functionally conserved sarcomeres compared to vertebrate skeletal muscle (98). To investigate the biological consequences of IARS LoF on muscle tissue, we analyzed the integrity of body wall muscle in DM8005 worms cultured on iars-1 RNAi, empty L4440 vector, unc-89 RNAi or *OP50* E. coli*. The DM8005 strain* expresses a GFP-tagged MYO-3 protein localized to the myofilaments allowing for visualization of the body wall muscle (170). It is worth noting that this strain demonstrates the roller phenotype. The gene *unc-89*, homologous to human obscurin (*OBSCN*) and Striated Muscle Enriched Protein Kinase (*SPEG*). *OBSCN* is imperative to muscle cell architecture as it dictates the organization of the A-band in body wall muscle, and thus the sarcomere and sarcoplasmic reticulum (171, 172). Importantly, the *SPEG* gene also plays a role in cytoskeletal development in the muscle. Previously, null mutations in *unc-89* have been observed to disrupt the structure of myosin filaments (172); therefore, knock down of *unc-89* served as a positive control. Meanwhile, the empty L4440 vector and regular OP50 *E. coli* served as negative controls. Worms were scored on days 1, 3, and 5 of adulthood as having either low, medium, or high levels of muscular disorganization (Fig. 5). Loss of *iars-1* expression leads to a progressive disorganization of the muscle ultrastructure characterised by the loss of muscle fibers alignment. This phenotype was intermediate in severity relative to positive and negative controls. Thus, *iars*-1 must play a role in maintaining organized structure of the sarcomeres.

### 3.4.4 Lifespan and rate of paralysis are unaffected by *iars-1* knock-down

Throughout adulthood, *C. elegans* naturally undergo a progressive loss of motility that eventually leads to paralysis and ends in death. This paralysis is thought to be a result of neuromuscular incoordination as a consequence of a decline in muscle integrity (173). Because *iars-1* RNAi had a detrimental effect on muscle morphology, we sought to analyse the effects of muscle specific knock-down of *iars-1* on the rate of paralysis and longevity using the NR350 *C. elegans* strain. This strain expressed the *rde-1* gene only in the muscle rendering it susceptible to RNAi only in this tissue (112). As such, the effects of muscle-specific IARS knock-down can be analysed without interference from neighbouring tissues (112). Therefore, we scored adult

NR350 worms daily for either paralysis or death depending on the assay. *C. elegans* grown on lawns of *iars-1* RNAi displayed no difference in either paralysis or lifespan compared to EV controls. In fact, the median rate of paralysis for both conditions was 15 days (Fig. 6a), while the median survival was 18 and 17 days for worms grown on *iars-1* and EV (Fig. 6b), respectively. Thus, muscle-specific LoF of *iars-1* affects neither the rate of paralysis nor the lifespan of *C. elegans.*

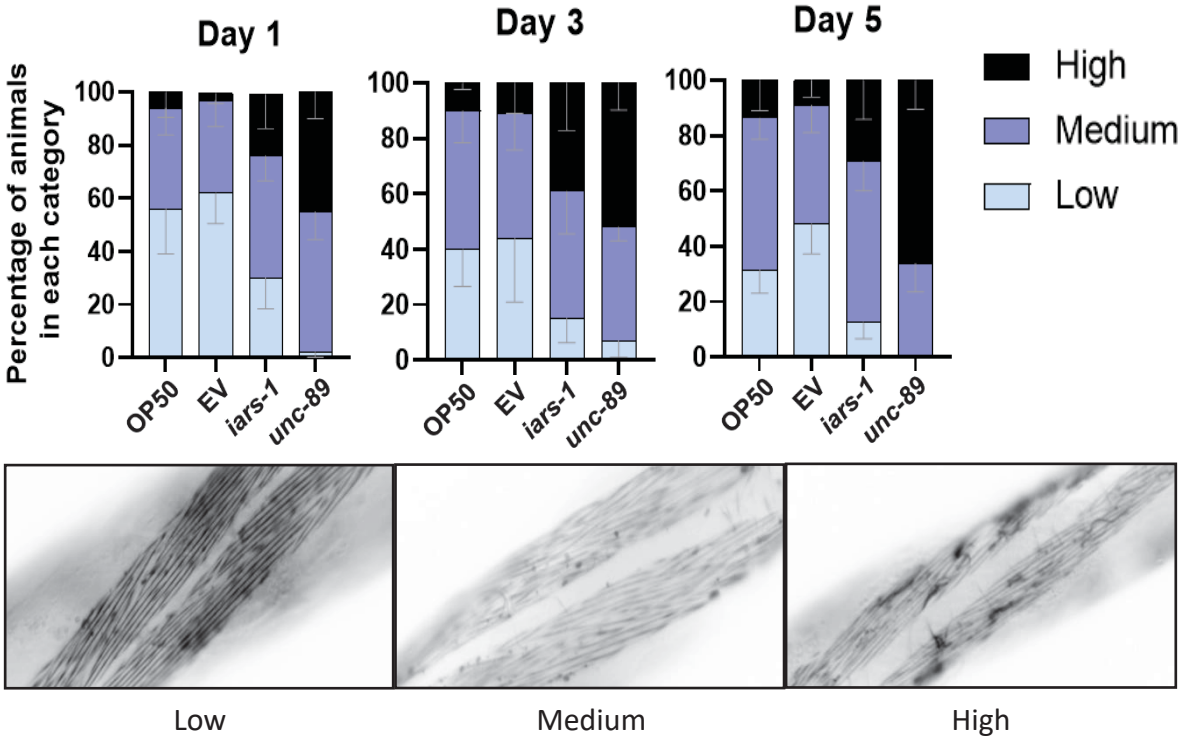



**Figure 5. Knock down of *iars-1* leads to disorganization of the muscle structure**
Fluorescent microscopy of myotubes from *C. elegans* cultured on OP50, EV, *iars-1 RNAi or unc-89* RNAi was performed to assess organization of the muscle structure. 100 worms per condition per day were categorized as having low, medium, and high levels of disorganization of the muscle structure on days 1, 3, or 5 of adulthood. Worms cultured on *iars-1* RNAi demonstrated an intermediate phenotype compared to two positive controls (OP50, EV) and one negative control (*unc-89)*.

### 3.4.5 Overall body morphology and motility are unaffected by *iars-1* knock-down

To deepen our understanding of the functional effects of muscle-specific *iars-1* knock-down, we screened for potential morphological and locomotive phenotypes using a combination of video microscopy and WormLab software from MBF Bioscience. Fortunately, the optical transparency of the body wall facilitates visualization of the internal structures by

microscopy (98). NR350 worms raised on either *iars-1* RNAi or EV were video recorded crawling on agarose gel or swimming in M9 buffer on days 1, 5, 9 and 12 of adulthood. Notably, a detailed description of all measured phenotypes is available in the WormLab User Guide Version 2020 (see Methods). Morphological features observed included body area, length, and width for which there were no significant differences relative to controls. During the crawling assays, we also analyzed speed, amplitude, wavelength, turn count, time spent in forward or reverse, and pirouette time among other parameter, and again no differences in locomotion were observed. Oftentimes, subtle phenotypes can be elucidated by submerging worms into an aqueous environment in which they are forced to swim. Nonetheless, no significant differences were detected in the swimming assays where swim speed, wave initiation frequency, curling, brush stroke, and activity are some of the many phenotypes taken into account. Therefore, it would seem that despite poor organization of the body wall muscle responsible for locomotion, muscle-specific knock-down of *iars-1* does not affect body morphology or locomotion.

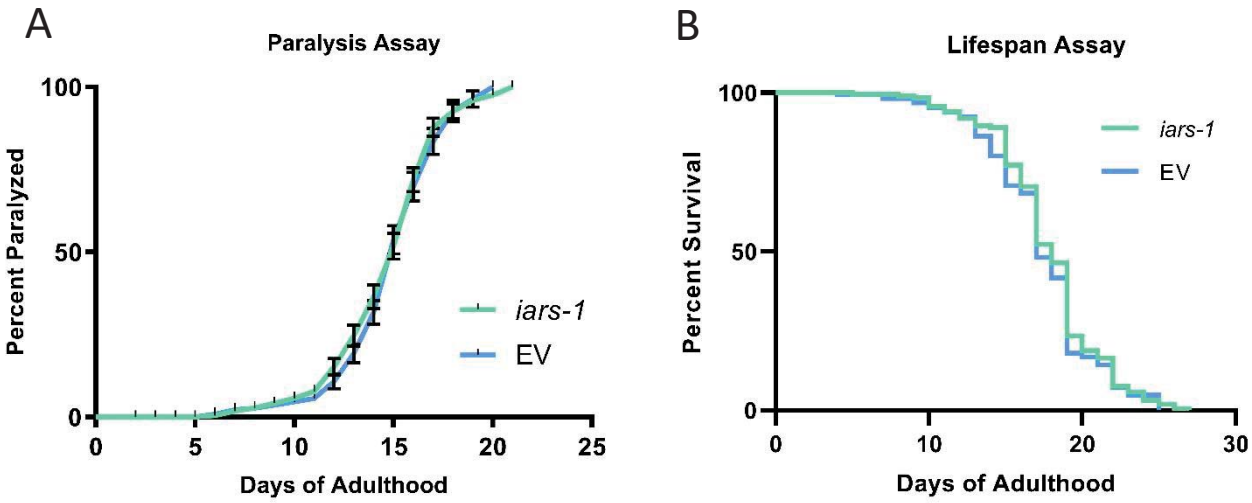



**Figure 6. Knock down of *iars-1* begets no change in paralysis rate or lifespan**
**(A)** Rate of natural-onset paralysis and **(B)** lifespan of *C. elegans* raised on *iars-1 RNAi* ($n_{paralysis}$=153, 162 censored; $n_{lifespan}$=173, censored=142) or EV ($n_{paralysis}$=117, censored=198; $n_{lifespan}$=119, censored=196). No significant differences were detected. Error bars are expressed as a percentage ±S.E.

## 3.5 Discussion

CMDs are a group of rare heterogeneous disorders that arise at birth or early infancy with symptoms such as delayed motor development, progressive weakness, and hypotonia or

"floppy infant syndrome". Disease severity, while dependent on age of onset and the affected muscle groups, is highly variable: some individuals demonstrate impaired ability to walk while others are unable to carry out normal daily activities such as getting out of bed, getting dressed or showering (174). A qualitative study on the impacts of living with a CMD found that deterrents of quality of life included aspects that affected physical, mental, and social health (174). Specifically, caregivers and affected individuals mentioned suffering from a loss of independence, the ability to participate in sports and daily activities, reduced mobility and ambulation, and emotional distress. In general, achieving early diagnosis has lead to better and long-lasting health outcomes (175), improved quality of life (176), reduced in-hospital mortality (2), and increased survival (177). Furthermore, retrospectively caregivers and parents have experienced less negative feelings and increased confidence in medical professionals with early diagnosis (178). Due to our limited understanding of disease-causing genes surrounding CMDs, the current diagnostic yield of gene panel leaves room for improvement. As such, many individuals linger within the diagnostic odyssey much to the detriment of their quality of life. Identification of novel disease-causing genes contributes to better overall diagnostic yield of CMDs affording more patients the opportunity for early diagnosis and the benefits it encompasses. Furthermore, a genetic diagnosis can inform prognosis and helps medical professionals monitor health issues and potential comorbidities (7). To date, there is no cure for CMD; therefore, treatments are mainly focused on managing individual symptoms. As gene-specific therapies are developed, a genetic diagnosis will be increasingly important for treatment of CMDs (7).

In this study, we present three siblings with an undiagnosed CMD despite exhaustive investigation by WES. To further investigations into the etiology of this disease, we implemented RNA-seq derived from muscle tissue. By combining data from WES and RNA-seq, we discovered a muscle-specific allelic imbalance along with a heterozygous missense mutation in the gene *IARS* that would have otherwise never been captured. This demonstrates the value of RNA-seq as a complementary diagnostic tool where DNA-based techniques alone are insufficient. The imbalance is characterized by the nearly complete knock-down of *IARS* in skeletal muscle tissue except for low levels of expression of the mutant allele. Therefore, we

pursued further functional analyses using an RNAi knock-down model of *iars-1* in *C. elegans* which revealed that this gene is important to the organization of the body wall muscle. Indeed, *C. elegans* cultured on *iars-1* RNAi demonstrated progressive dishevelment of myofibers; however, these worms did not display abnormalities in lifespan, rate of paralysis, body size or locomotive ability.

### 3.5.1 Possible explanations for the lack of a locomotive phenotype in *iars-1* knock-down worms

Worms cultured on *unc-89* RNAi, which served as a positive control in our microscopy experiment, move essentially as well as wt worms despite severe disorganization of thick and thin filaments (172, 179, 180). This puzzling phenomenon is similar to what was observed in *iars*-1 knock-down worms. The human homolog of *unc-89*, *OBSCN*, plays an important role in myofibril assembly and maintenance, muscle protein depredation and intracellular signalling (181). To-date, there is very little evidence of mutations in *OBSCN* causing disease. Mutations in the M10 domain of *TTN*, which interacts with OBSCN at the M-band, has been associated with three myopathic or dystrophic diseases (181). Similarly, the *SPEG* gene is important to myocyte cytoskeletal development has previously been associated with centronuclear myopathy (CNM5 MIM:615959) (182). Though knowledge of *OBSCN* and *SPEG* related disorders is extremely limited, there seems to be potential for some level of phenotypic heterogeneity, and so it would be unsurprising if not all of these phenotypes precipitated into *C. elegans* models. A previous study showed that a nonsynonymous mutation in *OBSCN* (c.13330C>T) along with a frameshift mutation in the gene *FLNC* segregated in two French patients, mother and son, but not in healthy relatives (183). Similar to *unc-89* knock-down worms, one of these patients presented specifically with a progressive distal dystrophy, muscle wasting and myofibrillar disorganization while the other had comparable symptoms with earlier age-of-onset and increased severity. It is somewhat surprising that the consequences of gene knock-down in *C. elegans* are less severe than a missense mutation in human patients; however, translatability to humans will always be the biggest limitation of using animal models to study disease. Nonetheless, the disturbance in muscle pathology observed in *iars-1* knock-down worms is suggestive of a role for *IARS* in congenital muscular dystrophy.

Tissue-specific RNAi strains of *C. elegans* are typically based on rescue of RNAi function in target tissues in an otherwise RNAi deficient genetic background (184). The muscle-specific NR350 strain *rde-1*(*ne219*); *kzIs20* was originally derived from a parental strain containing the *rde-1*(*ne219*) variation known to cause RNAi immunity (185). However, some suggest that the *rde-1*(*ne219*) strain may still be vulnerable to RNAi to some capacity (184). In the NR250 strain, a the transgene *kzIs20* was introduced to give it its unique properties (112). Specifically, this transgene allows *rde-1* to be expressed under the control of the promoter of the muscle-specific *h1h-1* gene (112). Importantly, *rde-1* operates in a cell autonomous manner (112), and so rescue of muscle-specific expression of *rde-1* is not propagated to other cell types and tissues. Even so, it is thought that tissue-specific promoters may still display low levels of expression in non-target tissues (184). The *kzIs20* transgene also contains the GFP marker pTG96 for NLS-GFP expression in muscle cells. RNAi-induced transcriptional gene silencing (RNAi-TGS) is the silencing of transgenes due to sequence homology between the transgene and the L4440 vector commonly used in RNAi experiments (186). Originally, it was thought that RNAi-TGS only affected LacZ-containing transgenes (186); however evidence is emerging for at least partial RNAi-TGS susceptibility in a wide range of reporter vectors including GFP-expressing transgenes (187). Therefore, it is possible that there is some level of suppression of the GFP-containing *kzIs20* transgene in the NR350 strain could possibly impact GFP fluorescence or rescue of *rde-1*. Ultimately, the suitability of the NR350 strain for muscle-specific susceptibility to RNAi could be affected by incomplete RNAi deficiency of the *rde-1*(*ne219*) variation, non-specificity of transgene promoters, or RNAi-TGS. This offers a possible explanation for the lack of a locomotive phenotype in the presence of disorganized sarcomeres observed in *C. elegans* exposed to *iars-1* RNAi. It is worth noting that the *kzIs20* genotype described in this section differs from that of the literature (112), but is correct as per the original author Hiroshi Qadota and the Caenorhabditis Genetics Center (CGC).

Another explanation could simply be that RNAi activity differed between assays. RNAi activity is known to be affected by RNAi concentration as well as environmental factors such as temperature (188). Throughout our experiments we ensured each plate was seeded with an overabundance of *E. coli* carrying the RNAi vector rather than controlling RNAi dosage. Because

no locomotive defects were observed across any of the conditions, access to food (and hence access to RNAi) was unlikely a factor of feeding success (and hence ingestion of RNAi). However, we cannot account for other factors such as pharyngeal pumping or epigenetic mechanisms across strains or promotor strength or copy number of dsRNA across RNAi vectors (189). Concerning environmental factors, our worms were kept at 20°C between experiments but may have experienced slight differences in temperature and handling during experimental manipulations which may have influenced RNAi intake or activity. The use of a hypersensitive RNAi strain vulnerable in all tissues would provide reassurance of RNAi activity.

Another reason a locomotive phenotype may not have been observed is that the 30 second time frame during which the worms were observed was too short to precipitate very subtle motility defects. That is to say that subtle differences may be more easily detected as the animal exhausts itself over a long period of time. To-date, several worm tracking softwares such as WormTracker 2.0 and CeleST (190) are able to capture and interpret the complex three-dimensional movements of one or more swimming *C. elegans* over long durations of time. Ultimately, if a locomotive phenotype exists in worms exposed to *iars-1* RNAi, observing the RNAi-induced depletion of *iars-1* over a prolonged period of time and during strenuous conditions may be the key to evoke the perceptible motility phenotype.

A final explanation for the lack of a motility phenotype is the incomplete representation of our patients' genotype. Because expression of the paternal allele was nearly absent in our patients, we chose to study the knock-down of *iars-1* in *C. elegans*. However, it is entirely possible that the lowly expressed mutant allele plays a significant role in disease pathogenesis. Furthermore, through Whole Genome Sequencing we discovered an intronic variant (hg38 chr9:92249527) in the gene *IARS* in II:1 that had been previously overlooked due to the nature of WES and RNA-seq techniques and later confirmed by Sanger sequencing in the father, II:2 and II:3. Regions of open chromatin often contain epigenomic modification sites or transcription factor binding sites that are able to bind various proteins in order to regulate DNA replication, nuclear organization, and transcription (189). If the intronic variant was within a position of open chromatin that was muscle-specific, it could help explain the observed muscle-specific loss of *IARS* expression. However, according to CATLAS (191) the chromatin at this loci

is open in muscle tissue among other tissues types, meaning it may have an impact on *IARS* expression but is not likely the cause for the tissue specific phenotype. However, a methylation site is observed in close proximity to our variant and seems to be specific to muscle tissues (Fig. 7a). Although further study is required to confirm our hypothesis, we believe the paternally inherited variant is not pathogenic by itself but could trigger a change in epigenomic patterns, such as methylation, resulting in the lost of the paternal allele in a tissue-specific manner. Thus, the apparent "hemizygote" phenotype in muscle (*e.g.,* solely expressing the maternal allele carrying the missense variant) mimics a recessive disease. This phenomenon would support the discrepancy between our patients' phenotype, solely affecting muscle tissues, and patients reported in the literature presenting a multi-systemic disease.

### 3.5.2 Future steps: In vitro cell models

Cellular models, especially those derived from the affected patients themselves, are the closest genetic representation one can achieve for human disease. One limitation of these models is that the *in vitro* environment vastly differs from that of a living organism which can potentially mask phenotypes or introduce new unrelated ones. We aim to assess the pathogenicity of *IARS* knock-down *in vitro* using two cellular models. The first model consists of a CRISPR/Cas9 derived knock-out of IARS in immortalized myoblasts. CRISPR/Cas9 was originally discovered as a bacterial defense mechanism against bacteriophages; however, it is now widely regarded as a highly precise, versatile, and affordable gene editing tool (192). For this model, we will exploit the error-prone NHEJ repair pathway in order to introduce a deleterious frameshift mutation in the *IARS* gene. In general, it is important to consider the various gene transcripts when creating a gRNA: for complete knock-out of *IARS* the gRNA must target exons that are shared by all transcripts (193).

Originally, we had hoped to study the allelic imbalance in fibroblasts derived from II:2, II:3 and the healthy father (I:1). From these tissues, we performed Sanger sequencing at two distinct loci on the *IARS* gene in order to 1) verify the presence of a paternally-inherited intronic variant in the affected patients and, 2) inspect whether or not patient-derived fibroblasts demonstrate the allelic imbalance observed in muscle tissue. Indeed, the patients were carriers of the intronic mutation attesting to its likely involvement in disease pathogenesis (Fig. 7b).

However, the patients appeared to be heterozygous for the exonic variant at the level of the fibroblast transcriptome suggesting the absence of an allelic imbalance. While this supports the muscle-specificity of the allelic imbalance, it also renders patient fibroblasts an inadequate model for its study.

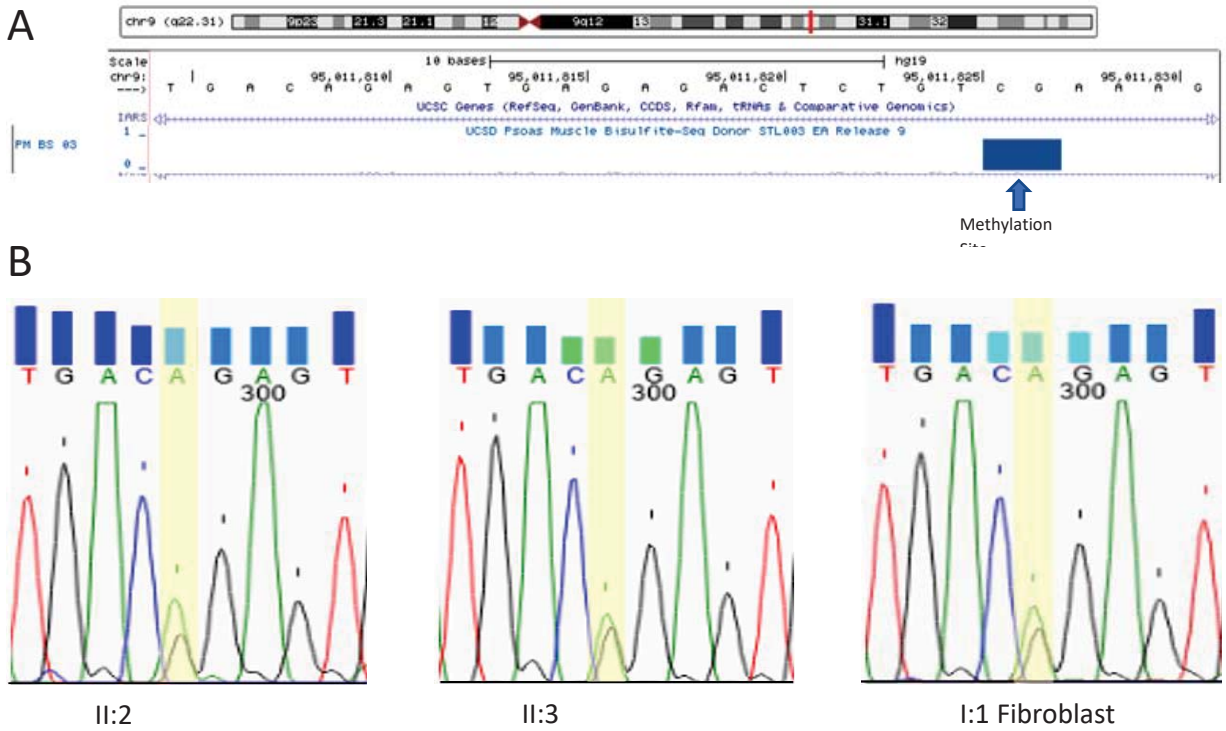



**Figure 7. Evidence for a paternally inherited intronic variant**

A potentially muscle-specific methylation site appears near the intronic variant identified in the father using the University of California Santa Cruz (UCSC) Genome Browser. **(B)** Sanger sequencing of fibroblast RNA from patient II:2, II:3, and the father confirms that the intronic variant is also present in these two children in the heterozygous state. Variant highlighted in yellow.

Transdifferentiation of patient fibroblasts into myoblasts offers a unique solution to enable capture of a complete picture of the patients' genotype including the allelic imbalance, despite the unavailability of donor muscle tissue. This approach involves the conversion of one cell type into another without passing through a state of pluripotency. Thus, transdifferentiation runs a low risk of teratomagenesis compared to reprogramming techniques (194). *MYOD1* is an important master regulator transcription factor that plays a role in muscle cell differentiation. When expressed ectopically, it can be used to transduce fibroblasts into myotubes, though inefficiently in human cells (195). Indeed, a major hurdle for successful

reprogramming is the sufficient activation of gene networks to induce the change as cells incompletely reprogrammed retain epigenetic memory from their former cell type (196). In order to solve this problem, new protocols are emerging to boost transdifferentiation efficacy by manipulating relevant signaling pathways (195) or genetically enhancing *MYOD1* (196). Because skin biopsies are relatively easy to obtain and less invasive than muscle biopsies, researchers are increasingly turning towards transdifferentiation for the study of pathomechanisms involved in neuromuscular disorders and their potential therapeutic targets (197). With these two models, we will assess amino-acylation activity using a spectrometric assay of pyrophosphate (PPi) release.

### 3.5.3 Concluding statement

Although next-generation sequencing has greatly expedited the search for novel disease-causing genes and variants, our biggest challenge today is overcoming the interpretation bottleneck caused by the incredible number of VUSs identified. In the past, other studies have demonstrated how RNA-seq can be used to guide variant analysis of genomic datasets through identification of aberrant splicing or expression (31-34). In this study, the application of RNA-seq in complementation to WES provided the necessary information to identify a muscle-specific allelic imbalance in the gene *IARS* as well as the potentially causative variant in three siblings with CMD. This work highlights the diagnostic value of a multi-omic approach for identifying genetic phenomena that are often overlooked by traditional DNA-based approaches.

# Chapter 4: Identification and functional analysis of two novel RYR1 variants

## 4.1. Introduction

Ease of movement is something we often take for granted as we navigate our daily lives. Unbeknownst to us, hundreds of small processes and reactions involving thousands of key proteins, enzymes, and channels take place at each moment of our lives to allow for basic bodily functions such as movement to occur. Muscle contraction relies on a phenomenon known excitation-contraction (EC) coupling which is the relationship between electrical stimulation of the muscle sarcolemma and the release of $Ca^{2+}$ from the sarcoplasmic reticulum in muscle cells (198). EC coupling begins when an action potential (AP) is propagated to the muscle fiber plasma membrane, known as the sarcolemma, from an innervating motor neuron. This leads to sodium-mediated radial depolarization of transverse invaginations in the sarcolemma known as T-tubules as well as longitudinal depolarization of muscle fibers (198). Located on the T-tubule membrane are DHPR-receptors responsible for detecting changes in the membrane potential. Upon stimulation, these receptors open ryanodine receptors (RyR) to which they are mechanically linked causing an efflux of $Ca^{2+}$ from the sarcoplasmic reticulum. Myoplasmic $Ca^{2+}$ concentrations are transiently increased to ten times their original value at which point RyR channels are maximally open (199). These $Ca^{2+}$ ions play an important role in exposing actin binding sites on myosin required for actin-myosin cross-bridging and muscle contraction. Finally, muscle contraction is terminated when $Ca^{2+}$ is actively reuptaken into the sarcoplasmic reticulum by sarcoendoplasmic reticulum calcium ATPase (SERCA) leading to the inaccessibility of actin binding sites and preventing cross bridging (200).

### 4.1.1 Structure and function of ryanodine receptor 1 (RyR1)

Many human myopathies are a result of damage to the integrity of any of the components involved in this complex network of interactions that allow for contractile function. The most common cause of congenital myopathies, which affect up to 1:90,000 people in the United States, are mutations in the gene *RYR1* mapping to chromosome 19q13.2 (201, 202). As opposed to *RYR2* and *RYR3* with expression in cardiac and nervous tissue

respectively, *RYR1* encodes a calcium channel with expression exclusive to the junctional terminal cisternae of the sarcoplasmic reticulum of skeletal muscle (203). Human *RYR1* is composed of 106 exons translating into a tertiary protein structure of about 5,038 amino acids in size (202). The functional homotetramer is one of the largest known ion channels with a mass of about 2250 kDA (202). Proteins of the sarco(endo)plasmic reticulum membrane initiate translation at the N-terminus while translocation to the membrane occurs concomitantly (203). RyR1 is thought to contain six to eight transmembrane helices and four transmembrane helical hairpin loops per monomer (203).

In rabbit RyR1, nine distinct domains in the cytoplasmic regions of each monomer are described in detail by Yan *et al* (204): The N-terminal domain (NTD), Channel domain and Central domain make up the central tower or pore of RyR1 responsible for allosteric regulation of channel opening and closing. The Handle domain and Helical Domain form a corona surrounding the central tower and provide a binding platform for RyR accessory proteins. The P1 and P2 Domains as well as three SPRY domains make up the periphery that attaches to the corona, acts as phosphorylation hotspots, and interacts with adjacent RyR1 monomers. Ultimately, characterization of the structure and function of the various regions in RyR1 provides a framework for interpretation of the biological impact of genetic mutations found in human disease.

Alternative splicing of mRNA is a mechanism to enable functional diversity of proteins encoded by a single gene. The mouse *RYR1* gene contains two alternative splice sites in modulatory regions that dictate the inclusion or exclusion of exon 70 (15bp) or exon 83 (18bp) to generate isoforms with tissue and temporal specificity (202, 205). During the embryonic period isoforms lacking exon 70 constitute 100% of expressed isoforms in the skeletal muscle; however, this number gradually decreases to about 30% of expressed transcripts in adulthood as the proportion of exon 70-positive isoforms increased gradually with development (205, 206). Similarly, exon 70-positive isoforms are predominately found in developed human muscle, while exon 70-negative isoforms were predominantly found in cultured myotubes (206). It is possible that the adult and neonatal isoforms demonstrate heterogeneity in channel activity,

protein interactions, and localization in response to a change in demand exhibited by the different developmental stages (205).

### 4.1.2 Disease phenotypes

Currently, over 700 variants in *RYR1* have been identified (207) which likely represents but a small subset considering the enormity of the gene. “RYR1-related diseases” is a blanket term for an array of disorders that affect the neuromuscular system each of which fall into one of three subgroups based on muscle histopathology, clinical phenotype, or pharmacogenomics (208). Those classified by muscle biopsy findings include central core disease, multiminicore disease, centronuclear myopathy, and congenital fiber type disproportion. Under those classified by symptoms there is king-denborough syndrome, rhabdomyolysis-myalgia syndrome, late-onset axial myopathy, and atypical periodic paralysis. Finally, under those classified by drug-gene interactions there is malignant hypothermia susceptibility and statin-induced myopathy. This list is by no means complete but provides a foundation for understanding the breadth of issues that arise from *RYR1* mutations. The National Organization for Rare Disorders (NORD) provides an excellently detailed account of most of these diseases (208). Here we will go into detail on a select two.

Some variants in *RYR1* are linked to a dominantly inherited pharmacogenetic disorder of the skeletal muscle that renders carriers susceptible to malignant hyperthermia (MH). MH susceptibility is characterized by a potentially lethal hypermetabolic response to potent volatile anesthetics such as halothane, sevoflurane, desflurane and the skeletal muscle relaxant succinylcholine (209). After exposure to a trigger, patients experience an uncontrollable rise in calcium in the myoplasm which puts $Ca^{2+}$ pumps into overdrive in an attempt to reuptake or dissipate the ions. The depletion of muscle ATP produces heat and disrupts the integrity of the muscle membrane resulting in hyperkalemia and rhabdomyolysis(209). Other clinical symptoms emerge as elevation of expired carbon dioxide, muscle rigidity, tachycardia, hyperthermia, and respiratory and metabolic acidosis (209). In a related disorder known as exertional rhabdomyolysis or exertional heat-illness, certain stressors such as intense exercise and heat have also been observed to trigger MH-like symptoms in swine (210), mice (211) and humans

(212). The incidence of MH episodes during anesthesia ranges from 1 in 5000 to 1 in 100,000 anesthesias; however, it is thought that the prevalence of genetic abnormalities associated with MH may be as high as 1:3000 individuals (209). This difference could be explained by the fact that MH typically triggers on the third anesthesia. Furthermore, carriers of MH-associated variants may never experience a triggering event and thus go undetected. Nevertheless, MH susceptibility presents a life threatening risk to those who are susceptible and yet preventative diagnostic tests have not yet become routine. This is because the standard diagnostic protocol is the highly sensitive (97%-98) and specific (78-94%) yet expensive and invasive *in vitro* contracture test involving the observation of muscle fiber contracture in the presence of halothane or caffeine (209). To date, there are 35 functionally-validated variants associated with MHS (207), the majority of which localize to the MHS/CCD region 1 (amino acids (aa) 35-614) and MHS/CCD region 2 (aa 2163-2458) in the N-terminal and central domain, respectively, of RYR1 (213). Fortunately, as more and more variants are discovered through application of next-generation sequencing techniques, genetic testing from blood is becoming an increasingly viable option.

Human skeletal muscle is made up of type I (slow twitch) and type II (fast twitch) fibers that specialize in endurance and burst movements, respectively (214). Under normal circumstances, these fibers are the same size. However, congenital fiber type disproportion (CFTD), though histologically heterogeneous, is characterized by the hypotrophy, atrophy and/or reductive size of slow twitch fibers by at least 12% (215) and up to 50% (216). Because this pattern appears with many other neurological disorders (215), the utility of CFTD as its own diagnostic entity has long been debated. However, a comprehensive literature review by Clarke *et al*. supports the diagnosis of CFTD by exclusion of other potential disorders and provides a strict set of inclusion and exclusion criteria (215). If rods, central nuclei or cores appear alongside fiber size disproportion on muscle biopsy, CFTD is to be overruled. Clinical symptoms of CFTD include muscle weakness most commonly in the limb girdle and proximal muscles, hypotonia, reduced reflexes, joint contractures, feeding difficulties, scoliosis and opthalmaparesis (11, 215). Very rarely do patients present with symptoms involving the cardiovascular system or intellectual ability. There are currently ten known disease-causing

genes for CFTD, the most common being *TPM3* (25-40%), *RYR1* (20%), and *ACTA1* (5%) (4, 11). These genes make up 50-70% of causative variations; therefore, exome sequencing or NGS congenital myopathy sequencing panel represent viable diagnostic options (11). Furthermore, genetic diagnosis of certain genes can inform prognosis and disease surveillance as is the case with variants found in *MYH7* or *TMP2* which have been associated with risks to cardiac health (11).

Ultimately, *RYR1* variations make up a significant portion of disease-causing variants in a plethora of different disorders of the skeletal muscle. Nonetheless, many variants in RYR1 and other genes remain to be identified. Thus, variant identification and functional analyses in known disease-causing genes still play an important role in increasing the diagnostic yield of human myopathies and related disorders.

## 4.2 Functional analysis of an RYR1 variant underlying a myopathy with variable expressivity

**Status**: This article is ready to be submitted to the European Journal of Human Genetics

**Authors :** Jennifer Hauteclocque[1,2], Jean-Denis Brisson[3,4], Adrien Rihoux[1,5], Alex Parker[1,2], Cam-Tu E. Nguyen[2,6*,] Martine Tetreault[1,2*]

1. Centre de recherche du centre hospitalier de l'Université de Montréal (CRCHUM), Montréal, Québec, Canada
2. Département de Neurosciences, Faculté de médecine, Université de Montréal, Montréal, Québec, Canada
3. Clinique des maladies neuromusculaires, Centre intégré universitaire de santé et services sociaux du Saguenay-Lac-Saint-Jean, Québec, Canada
4. Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada
5. Département of Microbiologie et Immunologie, Faculté de médecine, Université de Montréal, Montréal, Québec, Canada
6. Centre Hospitalier Universitaire – Sainte-Justine (CHSJ), Montréal, Québec, Canada

### Contributions

**JH:** Performed *C. elegans* analysis of *RYR1* variant. Wrote the manuscript. **JDB:** Performed clinical and genetic characterization of the patients. Revised the manuscript. **AR:** Participated in the *C. elegans* analysis. Revised the manuscript. **AP:** Participated in the *C. elegans* study design and data interpretation. Revised the manuscript. **CEN**: Participated in the study design, and clinical and genetic characterization of the patients. Revised the manuscript. **MT:** Study design, participated in the data analysis and interpretation. Participated in writing the manuscript.

**Key Words**: Clinical heterogeneity, *RYR1*-related myopathies, muscle, functional analysis, *C. elegans*

## Abstract

Neuromuscular diseases are known to be highly heterogeneous diseases where a single mutation is often associated with a wide range of phenotypes. Although the nature and location of the variant can explain the variability between different mutations, differences in clinical presentation amongst carriers of the same variant remain elusive. We present a human cohort composed of individuals carrying the same uncharacterized *RYR1* heterozygous missense variant (c.526G>A; p.E176K), but who demonstrate extremely diverse phenotypes. These phenotypes range from exertional rhabdomyolysis to fixed proximal limb muscle weakness with ptosis and include asymptomatic individuals. Our goal was to investigate the pathogenicity of this variant using a CRISPR/Cas9 mediated knock-in model in *C. elegans*. These animals demonstrated early-onset paralysis due to a decrease in acetylcholine secretion as characterized by decrease sensitivity to aldicarb. Furthermore, mutant worms also demonstrated a decrease in median and maximum lifespan and a reduction in size compared to the wild type. Finally, video microscopy of animals while crawling and swimming revealed signs of hindered motility. Ultimately, our model demonstrates reductions in synaptic transmission, longevity, and locomotion compared to controls which confirms pathogenicity of the identified variant in *C. elegans* and supports its pathogenicity in humans. These findings will not only have a direct impact on the affected individuals and their families, but will also improve diagnostic yield for myopathies, inform disease progression and risk management and lead to better overall care management. Finally, this cohort will make an excellent model for further studies of clinical heterogeneity owing to a single variant.

## Introduction

Neuromuscular diseases are known to be highly heterogeneous where a single mutation is often associated with a diversity of phenotypes with variable age of onset and disease progression (1). While variable expressivity is common to many diseases, the mechanisms involved have yet to be fully understood. Indeed, further investigation is required for more accurate prognosis of clinical manifestations, effective disease management, and the development of viable treatments and their use by individual patients (215). The remarkable clinical and genetic diversity observed in RYR1-related myopathies (RYR1-RMs) constitutes an

ideal model to study heterogeneity (3). RYR1-RMs are a spectrum of congenital myopathy that have been classified into four groups according to histopathological hallmarks observed on muscle biopsies. The most common RYR1-RMs are dominantly inherited central core disease (CCD) and recessively inherited multi-minicore disease (MMD) and less common are congenital fiber type disproportion (CFTD) and centronuclear myopathy (CNM) (4). Moreover, an array of interrelated conditions such as malignant hypothermia susceptibility (MHS), rhabdomyolysis-myalgia syndrome, and King Denborough syndrome have also been associated with RYR1-RMs (5).

This variability can partially be explained by the nature and location of *RYR1* mutations. To date, over 700 variants have been identified in RYR1 (6) as well as three hotspots enriched for mutations for dominantly inherited disorders. Specifically, MHS mutations are often found in the MHS/CCD region 1 (amino acids (aa) 35-614) and MHS/CCD region 2 (aa 2163-2458) located at the N-terminal and central domain, respectively. Conversely, CCD mutations are often found in the MHS/CCD region 3 (aa 4550-4940) located at the C-terminal (7-9). Functional studies suggest that these mutations cause passive leakage of calcium (10) and a reduction in RYR1-mediated $Ca^{2+}$ release (11). Alternatively, recessively inherited CMD, CFTD and MMD mutations can be found spread across all of RYR1 with particular enrichment of MMD mutations outside of the three hotspot regions (4). Recessive mutations are typically compound heterozygous with one missense mutation in conjecture with another missense, nonsense, splice-site, or frameshift mutation (9). Recessive mutations are thought to cause a decrease in RYR1 protein expression (9, 12) and generally have more severe phenotypes than dominant RYR1-RMs (9). While these findings can explain the variability between different mutations, less obvious is the reason for different clinical presentation owing to the same variant. Indeed, individual RYR1 mutations display striking inter- and intra-familial variability (9, 12). In some families, patients are affected with only a MHS phenotype whereas siblings are diagnosed with a myopathy (13). Some individuals with pathogenic mutations even demonstrate histological phenotypes yet appear asymptomatic (14).

Clinical and genetic heterogeneity represent a great challenge and often preclude a clear molecular diagnosis, thus requiring functional evidence to support genotype-phenotype

association. Here we report a cohort of 28 individuals from six unrelated families originating from the Saguenay-Lac-Saint-Jean region in Quebec, well known for the founder effect (15). Despite clear phenotypic heterogeneity, all individuals are carrying the same *RYR1* variant. RYR1 encodes a homotetrameric calcium channel that localizes to the membrane of the sarcoplasmic reticulum of skeletal muscle. In response to the depolarization of transverse (T)-tubules, RYR1 channels release calcium ions from the sarcoplasmic reticulum to the muscle cytosol in order to initiate excitation-contraction coupling (16). Functional impact of this variant was confirmed using a *C. elegans* model, highlighting the usefulness of such models. We report that mutant worms display altered body morphology and have reduced lifespan, synaptic transmission, and locomotion during crawling and swimming gaits.

## Methods

### *C. elegans* Strains

*C. elegans* maintenance was performed as per standard methods (17). All strains were grown at 20˚C on nematode growth agar (NGM) plates seeded with *E. coli* OP50. The strain PHX1948 *unc-68*(*syb1948*) containing the *unc-68* missense mutation (p.E182K) was generated by SunyBiotech (http://www.sunybiotech.com/) using CRISPR-Cas9 mediated genome editing of N2 animals. Forward (CGATAACGCACTGTCGATGA) and reverse (CGAAACACTTGCTCCCATCC) PCR and sequencing primers were used to confirm the *syb1948* mutation by the company itself (Sup. Fig. 3). The other strains included in this study are N2 and CB307 *unc-47*(*e308*) were obtained from the Caenorhabditis Genetics Center (CGC). Experiments were performed in triplicate at room temperature (22°C) using age-synchronized adult worms. *C. elegans* lost for reasons other than senescence (i.e., missing, bagging, or ruptured vulva) were censored.

### Synchronization

Ten to fifteen gravid adult worms were placed on a fresh plate seeded with OP50, left overnight to lay eggs, and removed the following day. Animals hatched from the eggs were then left to grow until they had reached the L4 stage after which a pool of L4 animals were placed on another fresh plate seeded with OP50. The following day, day 1 adult worms were picked from this pool to start new experiments.

**Paralysis Assay**

105 day 1 adult hermaphrodite worms were transferred to 3 OP50 plates and scored daily for paralysis. Because movement of the head is lost last, paralysed worms typically generate a visible triangle of space around their mouth void of OP50. Thus, worms were prodded with a platinum wire and considered paralyzed if they displayed movement restricted only to the head or the aforementioned triangle. Assay was done in triplicate for a total of 271 *syb1948* worms and 242 N2 worms.

**Lifespan Assay**

105 day 1 adult hermaphrodite worms were transferred to 3 OP50 plates and scored daily during the egg laying period, and every other day thereafter. Worms were prodded with a platinum wire and considered dead if they failed to move. Lifespan is defined as the period from day 1 of adulthood until death. Assay was done in triplicate for a total of 251 *syb1948* worms and 272 N2 worms.

**Aldicarb Assay**

NGM plates were spiked with 1mM aldicarb, seeded with OP50, and allowed to dry under a fume hood the same day the experiment was to take place. 105 day 1 adult hermaphrodites were transferred to 3 aldicarb spiked plates and prodded with a platinum wire every 30 minutes. Worms were considered paralysed if they failed to move. Assay was done in triplicate for a total of 266 CB307 worms, 259 *syb1948* worms, and 254 N2 worms.

**Motility Assays**

105 day 1 adult hermaphrodites were transferred to 3 blank NGM plates and video recorded crawling for 30 seconds using a digital stereomicroscope on the lowest magnification (S9i; Leica, Heerbrugg, Switzerland; equipped with 10x/23 eyepieces and an integrated 10 M.P camera with 1080HD video display). Worms were subsequently suspended in 150µL of M9 buffer and video recorded swimming for another 30 seconds before being transferred to plates seeded with OP50. This process was repeated with the same worms for days 1, 5, 9, and 12 of adulthood. 114-216 *syb1948* worms and 169-223 N2 worms were filmed in the crawling assays,

and 129-208 *syb1948* worms and 177-198 N2 worms were filmed in the swimming assays. These numbers vary depending on the number worms within the field of view of the microscope and the number of worms that survive up to a given day. Videos were uploaded onto WormLab (Update 2020.1.1, MBF Bioscience, VT, USA) software and tracked using parameters described in Sup. Table 1 in order to generate a track summary of standard metrics described in the WormLab User Guide Version 2020 (https://www.mbfbioscience.com/help/WormLab/Content/home.htm). Of note, one of the nine N2 day 1 plates was removed from the swimming analysis due to poor tracking.

**Statistics**

For paralysis, aldicarb, and lifespan assays Kaplan-Meier curves were generated using Graphpad Prism 9 and statistical significance was analysed using the Log-rank (Mantel-Cox) statistical test. For motility assays, video tracking results were exported to Graphpad Prism 9 and analysed using two-way ANOVA followed by Turkey multiple comparisons test.

## Results

**Clinical Description**

Twenty-eight paediatric and adult individuals with RYR1 mutation c.526G>A were identified (Table 1). These individuals derive from six different families, all from the province of Quebec and more specifically from the Saguenay-Lac-St-Jean region. All eleven symptomatic patients were heterozygous but one (proband) which was homozygous. The 17 asymptomatic individuals were assumed to be at risk of malignant hypothermia susceptibility. Of the 11 symptomatic patients, the most common symptoms were myalgia (8 patients), ptosis (7 patients), and muscle fatigability (3 patients). Only one patient displayed weakness (proximal legs), two displayed slight dysphagia, one displayed rhabdomyolysis, one displayed malignant hyperthermia and one was positive in vitro test for susceptibility to malignant hyperthermia. Seven patients demonstrated elevated CK levels typically within the few hundreds range (maximum 10 000 U/L without rhabdomyolysis manifestation). Of the five patients that had a muscle biopsy, one suggested metabolic myopathy and one suggested chronic neurogenic atrophy while the others appeared normal. Fifteen patients had an EMG but none revealed any

myopathic findings. Furthermore, no patient demonstrated muscle atrophy or hypertrophy, myotonia, motor milestones delay, orthopedic, respiratory, or cardiac issues.

**Heterozygous missense variant in *RYR1***

Next-generation sequencing gene panel or Sanger sequencing revealed a missense variant in *RYR1* (NM_000540.2: c.526G>A; p.E176K) in 28 individuals with striking phenotypic variability. This variant affects a conserved amino acid residue in the N-terminus of *RYR1* and is predicted to be pathogenic by Polyphon, SIFT, and Mutation Taster. Corroborating the pathogenicity of the identified variant, pathogenic variants (c.529C>T and c.533A>T) associated with MHS have been reported in proximity (18). The amino acid change p.E176K has not been observed in our in-house exome database but has been reported in Nomad (MAF: 0.000017) and ClinVar (5 entries). Amongst the ClinVar entries, two of them report the variant in patients presenting RYR1-related disorder phenotypes. No phenotypic information is available for the other reports. In all cases the variant was reported as a variant of unknown significance (VUS) due to lack of functional evidence. The uncertain impact of protein function of this variant combined to the highly polymorphic nature of *RYR1* (19) has resulted in inconclusive genetic testing and precluded a clear molecular diagnosis in these patients.

**S*yb1948* mutant *C. elegans* display early-onset paralysis and reduced synaptic transmission.**

The p.E176K missense mutation in humans is equivalent to a p.E182K amino acid change in *C. elegans* UNC-68 (Sup. Fig. 4). The *C. elegans* strain *syb1948* encoding this mutation was generated through CRISPR/Cas9 gene editing by SunyBiotech (https://www.sunybiotech.com/). We predict that swapping the negatively charged glutamine (E) residue for a positively charged lysine (K) residue will impair protein structure and have quantifiable effects on the health of the animal. After the egg-laying phase, *C. elegans* naturally undergo a progressive loss of motility that eventually ends in paralysis. This paralysis is thought to be a result of a decline in muscle integrity leading to deficits in neuromuscular coordination (20). In fact, the extent of a worm's motility, whether it is highly mobile and organized, stagnant unless prodded, or limited to the head and tail regions, is suggested to be a better indicator of life expectancy than chronological

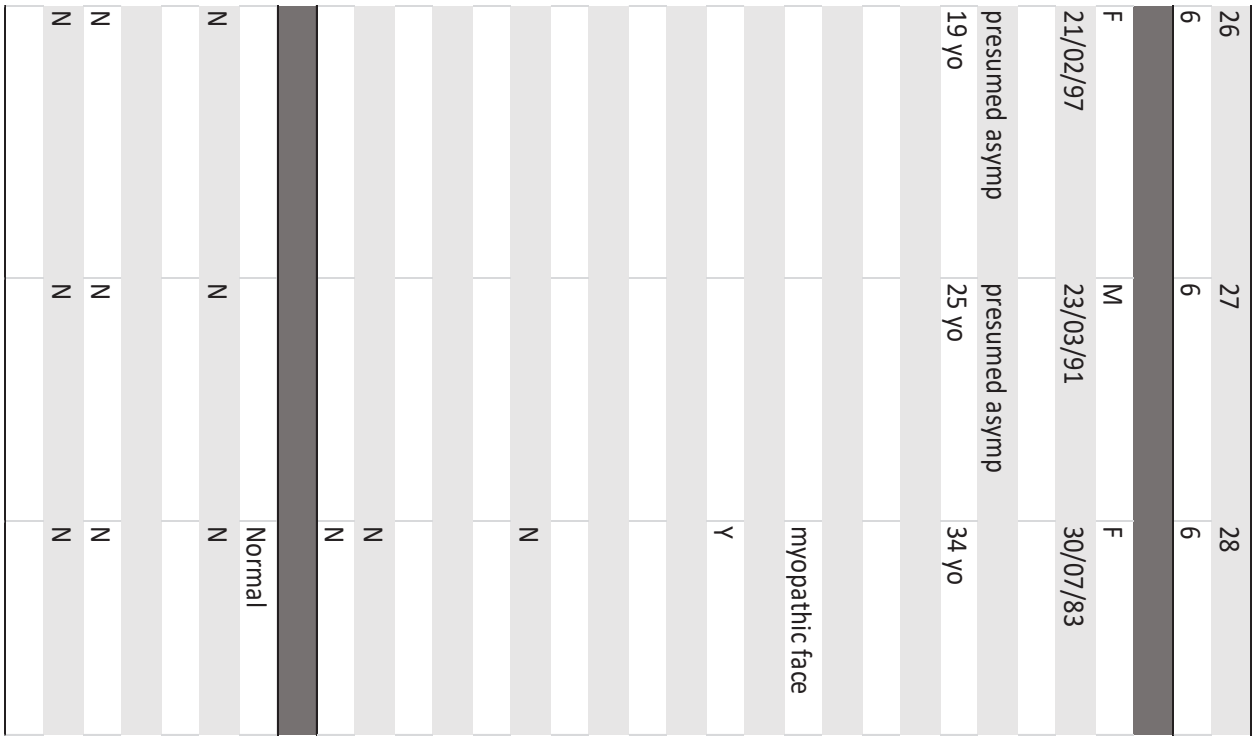
**Table 2. Clinical presentation of patient cohort with a shared *RYR1* mutation**

| Patient # | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Family # | 1 | 2 | 3 | 4 | 4 | 5 | 5 |
| **Clinical Data** | | | | | | | |
| Sex | M | M | M | F | M | M | M |
| Date of birth | 27/05/70 | 17/03/04 | 14/03/62 | 21/10/95 | 20/03/04 | 18/06/54 | 17/04/51 |
| Job | Janitor | | Deliverer | | | welder | |
| Age at first symptoms | young age | asymp | ~40 yo | asymp | childhood | | asymp |
| Age at diagnosis | 46 yo | 13 yo | 53 yo | 21 yo | 11 yo | 61 yo | 64 yo |
| Time between first symptoms and diagnosis | | | | | | | |
| Personal medical issues | | | N | | N | N | |
| First symptom | Myalgia | | myalgia | asymp | fatigability | myalgia | |
| Main symptom | Myalgia | | myalgia | | fatigability | myalgia | |
| Weakness | N | | N | N | N | | N |
| Facial weakness | N | | N | N | | | |
| Ptosis | N | | Y (slight) | N | Y | | N |
| Ophtalmoparesia | N | | N | N | | | |
| Bulbar symptoms | N | | | | | Y (slight dysphagia) | |
| Fatigability | | | | | Y | | |
| Malignant hyperthermia | Y | N | N | N | N | N | N |
| In vitro test for MH | | | | | | | |
| Heat intolerance | | | | | | N | |
| Cramps | | | | | | | |
| Myalgia | Y | | Y | | | Y | N |
| Rhabdomyolysis | | | N | N | | N | N |
| **Paraclinical Data** | | | | | | | |
| CK | 372-10 000 | | 81-1120 | | Normal | 140-578 | Normal |
| Muscle biopsy | N | N | N | N | Y | N | N |
| Biopsy suggesting congenital myopathy | | | | | N | | |
| Biopsy suggesting non-specific myopathy | | | | | N | | |
| MRI | N | N | N | N | N | | N |
| EMG | Y | N | Y | N | Y | Y | Y |
| EMG w myopathic changes | N | | N | | N | N | N |

asymp=asymptomatic

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 |
| | | | | | | | | |
| F | F | F | F | F | F | F | F | F |
| 17/01/45 | 11/05/77 | 10/11/08 | 02/01/50 | 27/01/53 | 03/11/63 | 24/10/66 | 25/08/59 | 12/04/76 |
| | | | | | | | | |
| 25 y | asymp | not evaluated | | | asymp | asymp | asymp | asymp |
| 70 yo | 28 yo | 8 yo | 65 yo | 62 yo | 50 yo | 46 yo | 53 yo | 39 yo |
| 45 yo | | | | | | | | |
| fibromyalgia | | | | fibromyalgia | | | | |
| myalgia | | | | myalgia | | | | |
| proximal leg weakness | | | | myalgia | | | | |
| Y (proximal leg) | | | N | N | | | | |
| | | | | | | | | |
| Y (slight) | | | N | slight on the left | | | | |
| | | | N | | | | | |
| | | | | | | | | |
| Y | | | | | | | | |
| N | | | N | N | | | | |
| | | | | | | | | |
| N | | | Y | | | | | |
| Y (stiffness) | | | Y stiffness | | | | | |
| Y | | | Y | Y | | | | |
| N | | | N | N | | | | |
| | | | | | | | | |
| Normal | Normal | Normal | Normal | Normal | Normal | Normal | | |
| Y | N | N | N | N | N | N | N | N |
| N | | | | | | | | |
| Chronic neurogenic atrophy | | | | | | | | |
| N | | N | | N | N | N | N | N |
| Y | N | N | N | N | N | N | Y | Y |
| N | | | | | | | N | N |

| 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|
| 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| | | | | | | | | |
| F | M | M | M | M | M | M | M | M |
| 19/02/08 | 17/10/90 | 28/09/97 | 22/11/92 | 02/03/62 | 25/10/39 | 22/02/93 | 16/07/63 | 06/04/61 |
| | | | | | | invalidity | | |
| asymp | asymp | asymp | 21 y | asymp | young age | 20 y | asymp | asymtomatic |
| 9 yo | 23 yo | 16 yo | | 55 yo | 75 yo | 22 yo | 52 yo | 54 yo |
| | | | | | | 2 y | | |
| | | | | | | myasthenia | | Lumbar surgery |
| | | | rhabdomyolysis | | | fatigability | | |
| | | | rhabdomyolysis | | | fatigability | | |
| | | | N | | N | | | N |
| | | | | | | | | |
| | | | | | Y (bilateral surgery) | Y fluctuating | | Y, very slight |
| | | | | | | Resolved diplopia | | |
| | | | | | slight dysphagia | | | |
| | | | | | | Y | | |
| | | | N | | N | N | | |
| | | | Y | | | Y | | |
| | | | | | | | | |
| | | | | | Y | | | |
| | | | Y | | Y | | | |
| | | | Y | | | N | | |
| | | | | | | | | |
| Normal | Normal | Normal | 344-170 000 | | 467-1800 | 254-1500 | | 257 |
| N | N | N | Y | N | Y | Y | N | N |
| | | | N | | N | N | | |
| | | | N | | Metabolic myopathy | N | | |
| N | N | N | N | N | N | N | | N |
| N | Y | Y | Y | N | Y | Y | Y | Y |
| | N | N | N | | N | N | N | N |




age (21). Therefore, we decided to analyse the effect of the mutant on the rate of paralysis to assess neuromuscular coordination. *syb1948* mutants and wild type (wt) N2 worms were observed daily from day 1 of adulthood until paralyzed. Mutant worms had a median paralysis of 13 days of adulthood which is slightly but significantly sooner than wt worms (14 days) (Fig. 1a). Thus, *syb1948* mutant worms clearly demonstrate an early-onset paralysis phenotype.

It has been demonstrated that arecoline, a muscarinic agonist able to stimulate acetylcholine (ACh) release in motor neurons, partially rescues locomotory deficits in aged *C. elegans* (20). Interestingly, UNC-68 has been suggested to play a presynaptic role in the generation of post-synaptic currents at the neuromuscular junction (NMJ) (22). To further characterise the early-onset paralysis of *syb1948* mutant worms we sought evidence of defective acetylcholine release. In *C. elegans* a muscle contraction is initiated by the release of ACh from cholinergic motor neurons in the presynaptic terminal of the neuromuscular junction (23). ACh is then taken up by receptors on the muscle membrane triggering $Ca^{2+}$ release by UNC-68 channels and initiating muscle contraction. Normally, acetylcholinesterase is released to degrade ACh and allow for the termination of the contraction (23). However, in the presence of aldicarb, an acetylcholinesterase inhibitor, ACh begins to accumulate in the NMJ (23). This

accumulation prevents muscle relaxation and eventually leads to paralysis (23). Thus, the rate of aldicarb-induced paralysis is a direct function of ACh release and is an effective method of assessing alterations in synaptic transmission (23). We observed *syb1948*, N2, and CB307 worms on plates spiked with 1mM of aldicarb for a number of hours until acute paralysis occurred. CB307, a strain carrying mutant *unc-47*, is known to be hypersensitive to aldicarb and was used as a positive control (24). *Syb1948* mutants display striking hyposensitivity to aldicarb demonstrated by a median paralysis time of 2.5 hours compared to the N2 wt and CB307 strains with median paralysis times of 2 hours and 1 hour respectively (Fig. 1b). Ultimately, this demonstrates slower ACh accumulation in *syb1948* mutants leading to delayed paralysis in the presence of aldicarb. We therefore conclude that the UNC-68(p.E182K) mutation alters synaptic transmission in *C. elegans*.

***Syb1948* mutant *C. elegans* display a reduction in lifespan.**

To dissect the effect of p.E182K on *C. elegans* longevity and thus the overall health of the worm, we analysed the lifespan of *syb1948* mutant worms compared to N2 wt worms. Animals were observed on a daily basis and the number of dead worms observed each day was recorded. A log-rank test on the survival curves revealed that the lifespan of mutant worms was significantly shorter than that of wt worms (Fig. 1c). Indeed, the median and maximum lifespan of the *syb1948* strain was 14 and 22 days, respectively. Comparatively wt worms had a mean lifespan of 15 days, which fell within the standard range (25), and a maximum lifespan of 25 days. Because these animals paralyse faster than the wt it is tempting to consider that perhaps their decreased lifespan is a result of their inability to move and find food. However, many studies suggest that a restriction or deprivation of food after the reproductive phase of adulthood leads to an increase in lifespan and enhanced resistance to stressors (26, 27). Unfortunately, the exact food intake of *syb1948* worms at any stage of life during our experiments and its effects on lifespan is unknown. Nonetheless, the truncated lifespan of mutant worms suggests that the UNC-68 (p.E182K) mutation has a detrimental impact on the health and longevity of *C. elegans.*

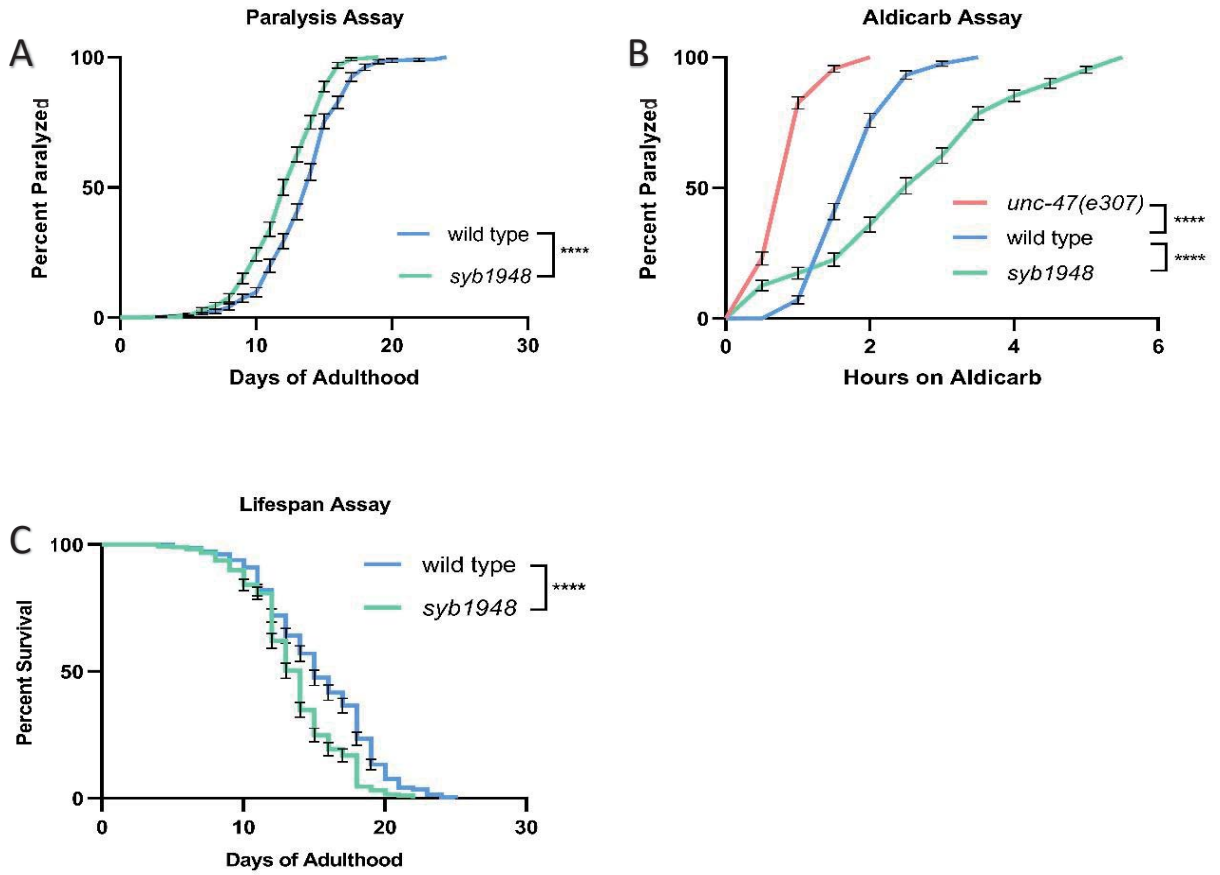



**Figure 1. The effect of the p.E182K variant on synaptic transmission and overall health.**
**(A)** Rate of natural-onset paralysis of *syb1948* mutants (n=271, 44 censored) and N2 wild type (n=242, 73) *C. elegans*. **(B)** Rate of paralysis of *syb1948* mutants (n=259, 11), N2 wild type (n=242, 16), and positive control *unc-47(e307)* (n=266, 4). *C. elegans* on NGM agar plates spiked with 1mM aldicarb. **(C)** Lifespan of *syb1948* mutants (n=251, 64 censored) and N2 wild type (n=272, 43) *C. elegans*. All graphs are expressed as a percentage ±S.E., ns=non-significant; *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

### *Syb1948* mutant *C. elegans* have a reduced body size.

To deepen our understanding of the functional role of the p.E182K variant, we characterized the overall morphology and locomotion of mutant *C. elegans* over several days of adulthood using video microscopy and WormLab software from MBF Bioscience. We found that mutant worms have a 5-18% smaller body size, measured by the area of the worm, than wt worms (Fig. 2a). Furthermore, mutant worms were also 4-17% shorter in length than their parental strain throughout adulthood (Fig. 2b) in a manner that correlated well with the total area. Finally, no significant changes in width were detected (not shown). Previous studies have demonstrated that *C. elegans* diameter, but not length, strongly correlated with muscle

strength (28). Therefore, if *syb1948* mutant worms demonstrate limitations in muscle strength in subsequent experiments we can eliminate the contributions of size-based effects and instead attribute these deficiencies to defects in muscle function.

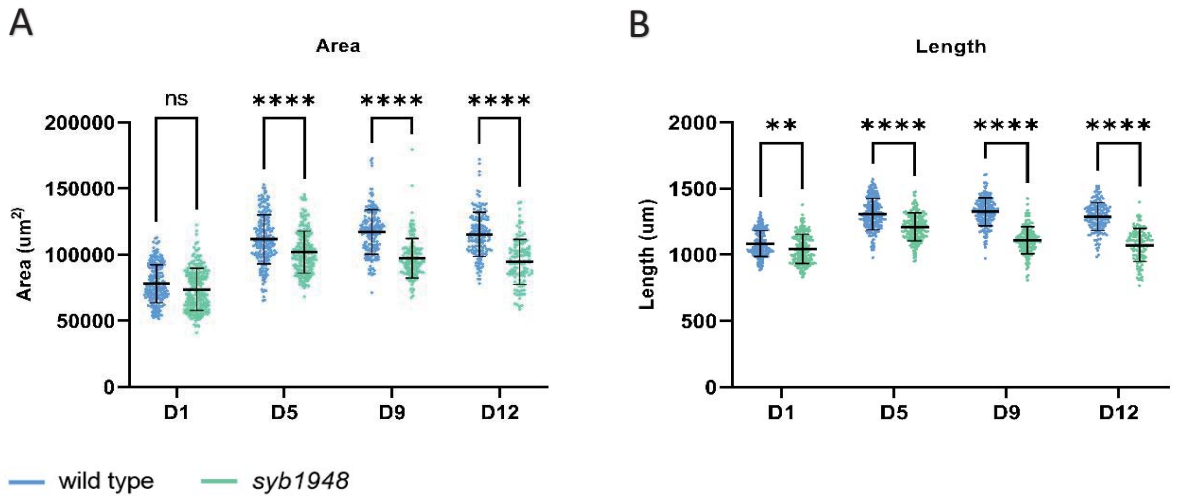



**Figure 2. Morphological changes observed during video microscopy**
**(A)** Body area and **(B)** length of *syb1948* (n=216,197,155, 114 (D1, D5, D9, D12)) and N2 wild type (n=223, 221, 192, 169) *C. elegans*. All graphs are expressed as mean ±S.D., ns=non-significant; *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

***Syb1948* mutant *C. elegans* demonstrate defects in motility within two locomotory gaits.**

*C. elegans* propel themselves forward on solid substances using undulatory locomotion that is initiated head to tail on a dorsal-ventral plane (29). Aging *C. elegans* display signs of progressive muscle degradation similar to human sarcopenia (21), and several *unc-68* variants have been shown to induce faster muscle aging (30). We sought evidence of impaired motility as a potential indicator of muscle degradation by observing crawling *syb1948* mutant worms on agarose gel using video microscopy. Mutant worms demonstrated a progressive reduction in crawling speed at later stages of adulthood being 18, 41, and 44% slower than wt worms on days 5, 9 and 12 respectively (Fig. 3a). Furthermore, mutant worms displayed a progressively shorter wavelength in the sinusoidal pattern of their trajectory that culminating to a 20% reduction by day 12 of adulthood (Fig. 3b). Notably, no significant difference was detected for amplitude or number of body bends per minute (not shown). While the observed reduction in

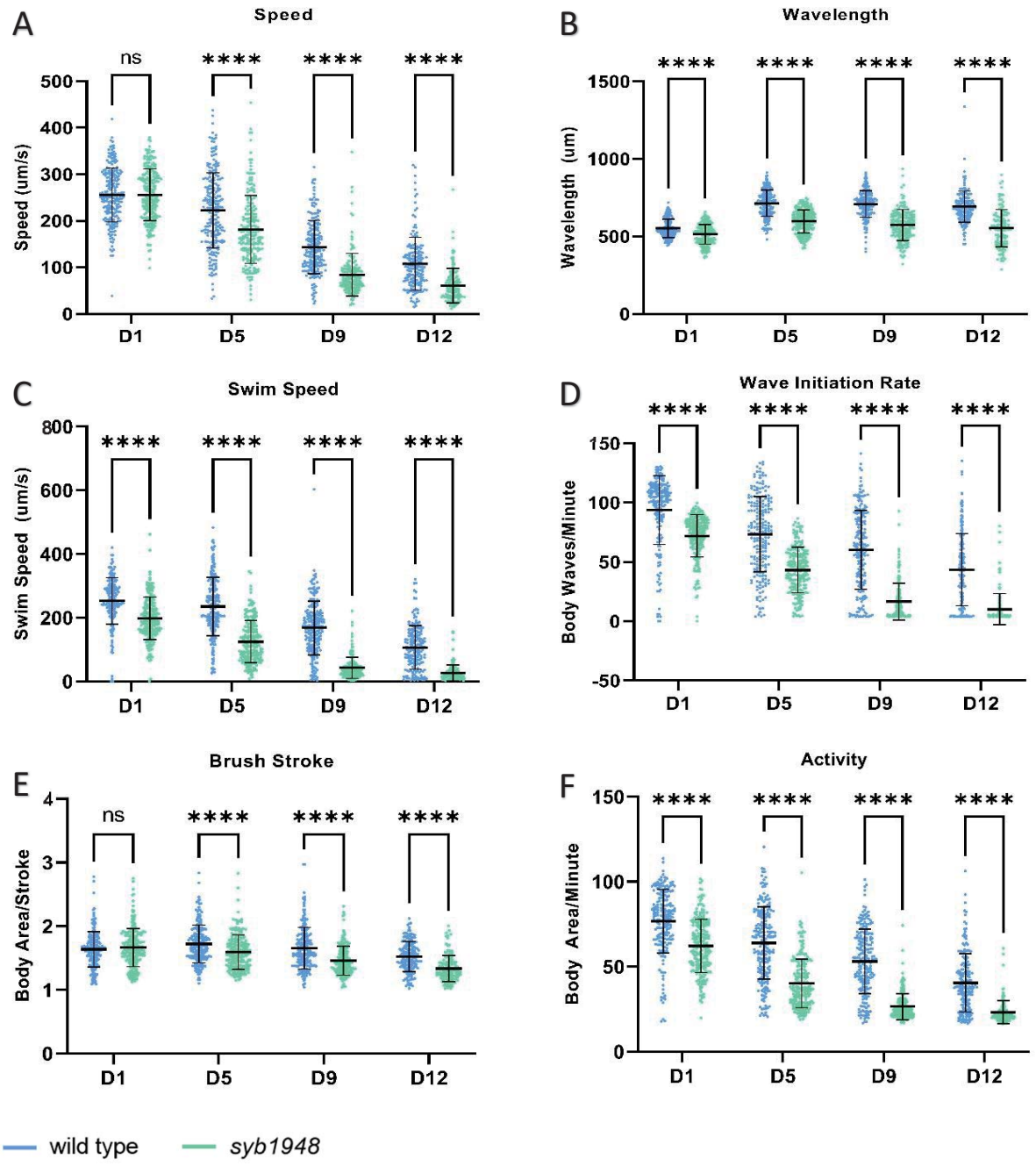



**Figure 3. Motility Differences in *syb1948* and N2 *C. elegans* while Crawling or Swimming**
**(A)** Speed and **(B)** wavelength of *syb1948* (n=216, 197, 155, 114 (D1, D5, D9, D12)) and N2 wild type (n=223, 221, 192, 169) *C. elegans* while crawling across agar gel. **(C)** Speed, **(D)** wave initiation rate, **(E)** brush stroke, and **(F)** activity of *syb1948* (n=208, 209, 194, 129 (D1, D5, D9, D12)) and N2 wild type (n=198, 209, 204, 177) *C. elegans* while suspended in M9 buffer. All graphs are expressed as mean ±S.D., ns=non-significant; *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

speed can partially be explained by the smaller body size of the *syb1948* worms, with this information alone it is unclear to what extent a muscle defect may play.

When observing worms crawling on agar gel, it becomes difficult to distinguish whether phenotypes arise from mechanical or behavioural deficits. For example, a lack of motility may be due to a disruption in excitation-contraction coupling. However, another explanation could be the animal's disinclination to move. *C. elegans* are adept swimmers and when suspended in solution they are much more inclined to act. Thus, we observed worms swimming in M9 buffer using video microscopy in order to more accurately compare motility differences owing to physical differences. The wave initiation rate or thrashing frequency of worms in liquid environments is a well-established indirect measure of muscle function and is often used to study the effect of drugs, chemicals and mutations on motility (31). On all tested days of adulthood, *syb1948* mutant worms demonstrated a significant reduction in both thrashing frequency (-23%, -41%, -73%, and -77%) and swim speed (-22%, -47%, -74% and -75%) compared to the wild type (Fig. 3c and d). Logically, the progression by which swim speed and thrashing frequency deteriorated correlated very well. As one might expect, mutant worms also displayed up to a 13% reduction in brush stroke (Fig. 3e), defined by the area "painted" by the body of the worm while thrashing, which culminated to a 50% decline when normalized to the time it takes to perform two brush strokes (Fig. 3f). This dramatic decline of motility is consistent with our observations in the crawling assays and is suggestive of a physical deficit in *syb1948* mutant worms caused by the p.E182K mutation.

## Discussion

RYR-RMs are the most common cause of congenital myopathy affecting up to 1:90,000 people in the United States (32). While over 700 variants have been confirmed (6), this is likely but a small subset considering the enormity of the gene. Human RYR1 is composed of 106 exons and is thus one of the largest known genes. Each protein monomer is made up of approximately 5,038 amino acids that join to form a tetramer with a mass of 2,250,000 Da making it the largest known ion channel (33, 34). The advent of next-generation sequencing has greatly accelerated the identification of new pathogenic variants for inherited diseases. However, a major limitation of its application is our ability to distinguish between pathogenic

and benign variants. In fact, our ability to identify genetic variants far exceeds our ability to interpret their functional effects leading to an overwhelmingly large number of variants of unknown significance (VUS). This is particularly true for genes like RYR1 that, due to their large size, are prone to a proportionally large number of polymorphisms (19). Functional validation of these VUSs is therefore still a crucial step in establishing pathogenicity despite the known association between RYR1 and inherited myopathies. In fact, the identification of novel pathogenic variants improves diagnostic yield, provides insight on disease progression, and leads to better care management.

The human RYR1 protein (NP_000531.2) and the *C. elegans* homolog UNC-68 (NP_001256074.1) share approximately 45 percent identity (35). Due to the high level of genetic conservation (36), the preservation of the structure, function, and composition of the sarcomere (37), and the rapidity of genome editing (34) *C. elegans* represent a powerful genetic model for RYR related disorders (34). In this study, we analyse the effects of the human RYR1 p.E176K variant on CRISPR/Cas9 gene edited *C. elegans*. We find that this model displays an increased sensitivity to aldicarb, a decrease in total body size and length, and several defects in locomotion. Another study using CRISPR/Cas9 gene edited *C. elegans* to model RYR1 variants linked to human myopathies also reported an increased sensitivity to aldicarb compared to wt in six of the eight strains tested while the other two showed no difference (38). Furthermore, they also showed that most of the mutant strains demonstrated a reduction in length compared to the wt and that all eight strains had smaller amplitude: length and wavelength: length ratios when crawling. This reduction in length is possibly a result of the perturbation of UNC-68's role in either pharyngeal pumping or synaptic transmission, both of which could diminish the ability to feed (38). Finally, they also reported a reduction in baseline thrashing frequency in all mutant strains at the larval L1 stage of life, and in one of the eight strains at adulthood. This data reveals that many RYR-RM models of *C. elegans* display defects in synaptic transmission, morphology, and locomotion consistent with our findings in this study. Based on this data, we propose that the p.E182K variant is pathogenic in *C. elegans* and that the homologous p.E176K variant is likely to have consequences for human carriers.

In contrast to mammals who encode three separate RYR genes (*RYR1*, *RYR2*, *RYR3*) that perform various functions in skeletal, cardiac, and nervous tissue respectively, *C. elegans* encode but a single *unc-68* gene to carry out all their functions (34). Marques and collaborators demonstrate that alternative promoters and alternative splicing within the Divergent Region 2 (DR2) of *unc-68* lead to six tissue-specific isoforms with mutually exclusive first exons (exon 1.1 or exon 1.2) (34). This allows the single *unc-68* gene to carry out the diversity of functions carried out by RYR genes. Isoforms containing exon 1.1 were shown to make up 92-100% of isoforms expressed in body wall muscle, which were otherwise only found at low levels (12%) in neurons. Alternatively, neurons were found to be enriched (87%) in exon 1.2-containing isoforms. This group also demonstrated that mutant lines lacking the 1.1 exon, but not those lacking the 1.2 exon, demonstrated a reduction in swimming motility measured by thrashing rate suggesting that isoforms containing exon 1.1 but not 1.2 are important for swimming. Therefore, the swimming defects exhibited by *syb1948* mutant worms may be a result of dysfunctional exon 1.1-containing isoforms that are primarily characteristic of muscle. Our approach could benefit from a more direct measurement of muscle functionality such as a force-based assay (28) or observation of muscle tissue morphology to support this hypothesis.

*Syb1948* mutant worms also display early-onset paralysis and decrease sensitivity to aldicarb suggesting a reduction in Ach release at the NMJ. One study suggested that release of $Ca^{2+}$ by the *C. elegans* ryanodine receptors is necessary for normal quantal size at the NMJ (22), proposing several mechanisms by which ryanodine receptor may be involved in neurotransmitter release by the presynaptic neuron: 1) increasing neurotransmitter vesicle size, 2) increasing the number of docked vesicles, 3) promoting vesicle loading with neurotransmitter, 4) stimulating intervascular fusion, or 5) increasing the size or duration of the fusion pore during kiss-and-run synaptic release. We propose that the p.E182K mutation disrupts this function leading to a reduction in Ach release, altered synaptic function, and ultimately paralysis. Marques *et al* found that *C. elegans* neurons express both 1.1 and 1.2-containing isoforms in an approximate 1:9 ratio (34). Therefore, our mutation may also affect 1.2-containing *unc-68* isoforms. The almost exclusive expression of vertebrate RYR1 in skeletal muscle may explain the lack of paralytic phenotypes within our human cohort and among RYR-

RMs in general. While the translatability to humans is a limitation of using animal models, these data still provide insight into gene interactions and pathways.

Curiously, *unc-68* mutant worms demonstrate a reduction in body size. As previously mentioned, small body size may be a consequence of food deprivation. Nonetheless, an alternative explanation should be considered. The calcineurin pathway is one of three mechanisms typically associated with body size in *C. elegans* (39, 40); however, to our understanding the *unc-68* has not yet been assigned to the calcineurin pathway. Coimmunoprecipitation in mice skeletal muscle cells (C2C12) has revealed a $Ca^{2+}$-dependent association between calcineurin and FK506-binding protein (FKBP12): an accessory subunit of RyR (41). Furthermore, a calcineurin, FKBP12, RyR trimer was suggested to play an important role in the regulation of RyR channel activity, $Ca^{2+}$ signalling, muscle contraction and muscle relaxation (41). Finally, the calcineurin gene shares a high level of homology between *C. elegans* and mammals and its biochemical function seems to be conserved (40). Thus, it is realistic to suggest that the interaction between calcineurin and the ryanodine channel may also be conserved. Furthermore, our variant is found in the N-terminal region of the *RYR1* gene which forms a cytosolic structure that regulates channel opening through interactions with other proteins and molecules (34). Thus, it is possible that the p.E182K mutation in *C. elegans* may disrupt the calcineurin pathway leading to the small body size phenotype.

This work highlights the importance of validating the pathogenicity of suspected gene mutants especially in the context of very large polymorphic genes. However, the cause of phenotypic heterogeneity surrounding this variant remains a mystery. As previously mentioned, the clinical heterogeneity typical of RYR-RMs can partially be explained by the nature and location of the *RYR1* mutations. This is exemplified by the regional hotspots identified for MHS and CCD (7-9) and the difference in pathomechanisms involved between recessively and dominantly inherited disease phenotypes (9-12, 42). However, an explanation for the different clinical presentations represented in carriers of the same variant, such is the case in our human cohort, remains to be elucidated. Recent studies have demonstrated that variability in gene expression or methylation patterns can be associated with variable phenotype (43). The presence of variants in promoters, enhancer or repressor regions can have

an important impact on gene expression (44). Similarly, DNA methylation is known to be a common epigenetic modification and a regulator of gene expression. Methylation levels can differ between individuals since they are influenced by both genetic variations and environmental factors (45). In the case of RYR1-related disorders a study defined an expression and methylation profile associated with recessive *RYR1* mutations (46). In the transcriptome, they identified decreased expression in muscle-specific microRNAs, increase expression of class II histone deacetylase and increase expression of DNA cytosine-5-methyltransferase 1 and 2 (*DNMT1* and *DNMT2*). Differential methylation analysis identified an overall hypermethylation in recessive *RYR1* patients. In fact, hypermethylation cause muscle-specific monoallelic RYR1 expression in up to 55% of recessive cases as a result of gene silencing (47). However, these findings were not consistent in patients affected with a dominant RYR1 disorder. Defining an omic signature associated with RYR1 severity in our cohort of patients would give insight on pathological mechanisms involved in disease variability and ultimately contribute to better risk assessment and evaluation of treatment efficacy.

**Acknowledgements**

The authors would like to thank the patients who partake in this study. We would also like to highlight the collaboration of Sarah Duhaime, Ericka Guitard, Gilles Tossing and Audrey Labarre for their great support with the *C. elegans* experiments. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). The project was funded by the Fondation Courtois, the Rare Disease Foundation and Muscular Dystrophy Canada. JH received bursaries for the Canadian Institute of Health Research, Neuroscience Department of Université de Montréal and the CRCHUM. AJP and M.T. received salary award from Fond de recherche du Québec – Santé

**Conflicts of Interest**

The authors declare that they have no conflict of interest.

References

1. Heydemann, A., K.R. Doherty, and E.M. McNally, *Genetic modifiers of muscular dystrophy: implications for therapy.* Biochim Biophys Acta, 2007. **1772**(215): p. 216-28.

2. McClellan, J. and M.C. King, *Genetic heterogeneity in human disease.* Cell, 2010. **141**(215): p. 210-7.
3. McCarthy, T.V., K.A. Quane, and P.J. Lynch, *Ryanodine receptor mutations in malignant hyperthermia and central core disease.* Hum Mutat, 2000. **15**(5): p. 410-7.
4. Witherspoon, J.W. and K.G. Meilleur, *Review of RyR1 pathway and associated pathomechanisms.* Acta Neuropathol Commun, 2016. **4**(1): p. 121.
5. Knuiman, G.J., et al., *The histopathological spectrum of malignant hyperthermia and rhabdomyolysis due to RYR1 mutations.* J Neurol, 2019. **266**(4): p. 876-887.
6. Litman, R.S., et al., *Malignant Hyperthermia Susceptibility and Related Diseases.* Anesthesiology, 2018. **128**(1): p. 159-167.
7. Lawal, T.A., J.J. Todd, and K.G. Meilleur, *Ryanodine Receptor 1-Related Myopathies: Diagnostic and Therapeutic Approaches.* Neurotherapeutics, 2018. **15**(4): p. 885-899.
8. Zhou, H., et al., *Molecular mechanisms and phenotypic variation in RYR1-related congenital myopathies.* Brain, 2007. **130**(Pt 8): p. 2024-36.
9. Klein, A., et al., *Clinical and genetic findings in a large cohort of patients with ryanodine receptor 1 gene-associated myopathies.* Hum Mutat, 2012. **33**(6): p. 981-8.
10. Yang, T., et al., *Elevated resting [Ca(2+)](i) in myotubes expressing malignant hyperthermia RyR1 cDNAs is partially restored by modulation of passive calcium leak from the SR.* Am J Physiol Cell Physiol, 2007. **292**(5): p. C1591-8.
11. Loy, R.E., et al., *Muscle weakness in Ryr1I4895T/WT knock-in mice as a result of reduced ryanodine receptor Ca2+ ion permeation and release from the sarcoplasmic reticulum.* J Gen Physiol, 2011. **137**(1): p. 43-57.
12. Attali, R., et al., *Variable myopathic presentation in a single family with novel skeletal RYR1 mutation.* PLoS One, 2013. **8**(7): p. e69296.
13. Jungbluth, H., *Multi-minicore Disease.* Orphanet J Rare Dis, 2007. **2**: p. 31.
14. Davis, M.R., et al., *Principal mutation hotspot for central core disease and related myopathies in the C-terminal transmembrane region of the RYR1 gene.* Neuromuscul Disord, 2003. **13**(215): p. 151-7.
15. Gauvin, H., et al., *Genome-wide patterns of identity-by-descent sharing in the French Canadian founder population.* Eur J Hum Genet, 2014. **22**(6): p. 814-21.
16. Franzini-Armstrong, C., F. Protasi, and V. Ramesh, *Shape, size, and distribution of Ca(2+) release units and couplons in skeletal and cardiac muscles.* Biophys J, 1999. **77**(3): p. 1528-39.
17. Brenner, S., *The genetics of Caenorhabditis elegans.* Genetics, 1974. **77**(1): p. 71-94.
18. Monnier, N., et al., *Correlations between genotype and pharmacological, histological, functional, and clinical phenotypes in malignant hyperthermia susceptibility.* Hum Mutat, 2005. **26**(5): p. 413-25.
19. North, K.N., et al., *Approach to the diagnosis of congenital myopathies.* Neuromuscul Disord, 2014. **24**(215): p. 97-116.
20. Glenn, C.F., et al., *Behavioral deficits during early stages of aging in Caenorhabditis elegans result from locomotory deficits possibly linked to muscle frailty.* J Gerontol A Biol Sci Med Sci, 2004. **59**(12): p. 1251-60.
21. Herndon, L.A., et al., *Stochastic and genetic factors influence tissue-specific decline in ageing C. elegans.* Nature, 2002. **419**(6909): p. 808-14.

22. Liu, Q., et al., *Presynaptic ryanodine receptors are required for normal quantal size at the Caenorhabditis elegans neuromuscular junction.* J Neurosci, 2005. **25**(29): p. 6745-54.
23. Oh, K.H. and H. Kim, *Aldicarb-induced Paralysis Assay to Determine Defects in Synaptic Transmission in Caenorhabditis elegans.* Bio Protoc, 2017. **7**(14).
24. Vashlishan, A.B., et al., *An RNAi screen identifies genes that regulate GABA synapses.* Neuron, 2008. **58**(3): p. 346-61.
25. Gems, D. and D.L. Riddle, *Defining wild-type life span in Caenorhabditis elegans.* J Gerontol A Biol Sci Med Sci, 2000. **55**(5): p. B215-9.
26. Lee, G.D., et al., *Dietary deprivation extends lifespan in Caenorhabditis elegans.* Aging Cell, 2006. **5**(6): p. 515-24.
27. Savage-Dunn, C., et al., *Genetic screen for small body size mutants in C. elegans reveals many TGFbeta pathway components.* Genesis, 2003. **35**(4): p. 239-47.
28. Rahman, M., et al., *NemaFlex: a microfluidics-based technology for standardized measurement of muscular strength of C. elegans.* Lab Chip, 2018. **18**(15): p. 2187-2201.
29. Von Stetina, S.E., M. Treinin, and D.M. Miller, 3rd, *The motor circuit.* Int Rev Neurobiol, 2006. **69**: p. 125-67.
30. Nicoll Baines, K., et al., *Aging Effects of Caenorhabditis elegans Ryanodine Receptor Variants Corresponding to Human Myopathic Mutations.* G3 (Bethesda), 2017. **7**(5): p. 1451-1461.
31. Buckingham, S.D. and D.B. Sattelle, *Fast, automated measurement of nematode swimming (thrashing) without morphometry.* BMC Neurosci, 2009. **10**: p. 84.
32. Amburgey, K., et al., *Prevalence of congenital myopathies in a representative pediatric united states population.* Ann Neurol, 2011. **70**(4): p. 662-5.
33. Robinson, R., et al., *Mutations in RYR1 in malignant hyperthermia and central core disease.* Hum Mutat, 2006. **27**(10): p. 977-89.
34. Marques, F., et al., *Tissue-specific isoforms of the single C. elegans Ryanodine receptor gene unc-68 control specific functions.* PLoS Genet, 2020. **16**(10): p. e1009102.
35. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
36. Gieseler, K., H. Qadota, and G.M. Benian, *Development, structure, and maintenance of C. elegans body wall muscle.* WormBook, 2017. **2017**: p. 1-59.
37. Duronio, R.J., et al., *Sophisticated lessons from simple organisms: appreciating the value of curiosity-driven research.* Dis Model Mech, 2017. **10**(12): p. 1381-1389.
38. Graham, B., M.A. Shaw, and I.A. Hope, *Single Amino Acid Changes in the Ryanodine Receptor in the Human Population Have Effects In Vivo on Caenorhabditis elegans Neuro-Muscular Function.* Front Genet, 2020. **11**: p. 37.
39. Morck, C. and M. Pilon, *C. elegans feeding defective mutants have shorter body lengths and increased autophagy.* BMC Dev Biol, 2006. **6**: p. 39.
40. Bandyopadhyay, J., et al., *Calcineurin, a calcium/calmodulin-dependent protein phosphatase, is involved in movement, fertility, egg laying, and growth in Caenorhabditis elegans.* Mol Biol Cell, 2002. **13**(9): p. 3281-93.
41. Shin, D.W., et al., *Ca(2+)-dependent interaction between FKBP12 and calcineurin regulates activity of the Ca(2+) release channel in skeletal muscle.* Biophys J, 2002. **83**(5): p. 2539-49.

42. Treves, S., et al., *Congenital muscle disorders with cores: the ryanodine receptor calcium channel paradigm.* Curr Opin Pharmacol, 2008. **8**(3): p. 319-26.
43. Pai, A.A., J.K. Pritchard, and Y. Gilad, *The genetic and mechanistic basis for variation in gene regulation.* PLoS Genet, 2015. **11**(1): p. e1004857.
44. Mora, A., et al., *In the loop: promoter-enhancer interactions and bioinformatics.* Brief Bioinform, 2016. **17**(6): p. 980-995.
45. He, Z., et al., *Role of genetic and environmental factors in DNA methylation of lipid metabolism.* Genes Dis, 2018. **5**(1): p. 9-15.
46. Rokach, O., et al., *Epigenetic changes as a common trigger of muscle weakness in congenital myopathies.* Hum Mol Genet, 2015. **24**(16): p. 4636-47.
47. Zhou, H., et al., *Epigenetic allele silencing unveils recessive RYR1 mutations in core myopathies.* Am J Hum Genet, 2006. **79**(5): p. 859-68.

## 4.3 Case study: A novel RYR1 variant associated with type I fiber predominance identified using RNA-sequencing

**Status**: This article is ready to be submitted to Genomic Medicine

**Authors:** Jennifer Hauteclocque[1,2], Adrien Rihoux[1,3], Alex Parker[1,2], Erin K. O'Ferrall[4,5], and Martine Tétreault[1,2]

1. Centre de recherche du centre hospitalier de l'Université de Montréal (CRCHUM), Montréal, Québec. Canada
2. Département de Neurosciences, Faculté de médecine, Université de Montréal, Montréal, Québec, Canada
3. Département de Microbiologie et Immunologie, Faculté de médecine, Université de Montréal, Montréal, Québec, Canada
4. Department of Pathology, McGill University Health Centre, Montreal Neurological Institute Hospital, Montreal, QCH3A 2B4, Canada.
5. Rare Neurological Diseases Group, Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QCH3A 2B4, Canada.

**Contributions**

**JH**: Performed RNA-sequencing, Sanger validation, and *C. elegans* analysis. Wrote the manuscript. **AR:** Participated in the *C. elegans* analysis. Revised the manuscript. **AP**: Participated in *C. elegans* study design and data interpretation. Revised the manuscript. **EKO**: Clinical characterization. Revised the manuscript. **MT**: Study design, participated in data analysis and interpretation. Participated in writing the manuscript.

**Key Words**: Case study, *de novo*, gene expression, muscle, functional analysis, *C. elegans*

## Abstract

Congenital myopathies (CM) are pediatric disorders characterized by muscle weakness and hypotonia that affect up to 1:25,000 people. Patients often present with contractures, scoliosis, dysmorphic features while involvement of the cardiovascular or respiratory system is rare. The advent of next generation sequencing (NGS) techniques is a rapidly evolving field that has greatly accelerated the genetic diagnosis of CMs. In fact, gene panels for CM have an impressive hit rating of >50%. Nonetheless, a large proportion of patients do not carry any of the known disease-causing genes and lack genetic resolution. Traditionally, DNA-based NGS techniques have been deployed to probe the genome for novel causative variants but are impeded by the staggering number of variants identified whose functional effects, if any, are unknown. Thus, RNA-sequencing, in its ability to provide insight into gene expression, allele-specific expression, gene splicing, and fusions, represents an invaluable tool to narrow the search for candidate variants by providing clues to the biological meaning of novel variants. We present a female patient with a static myopathy manifesting as proximal and distal weakness since the age of two, for whom gene panel was inconclusive. Other clinical features include long face, facial weakness, high arched palate, delayed motor milestones and poor sport performance. RNA-sequencing on muscle tissue revealed an expression profile suggestive of type I fiber predominance leading to the identification of the pathogenic variant in the gene *RYR1* (c.12083C>T, p.S4028L). Interestingly, this variant was not flagged by gene panel. Further functional analyses pursued in a CRISPR/Cas9 mediated knock-in model in *C. elegans* revealed early-onset paralysis characterized by hyposensitivity to aldicarb, reduced longevity, decreased body size, and limited motility. The homologous c.12083C>T variant is clearly pathogenic in *C. elegans* and thus likely has consequences for human carriers as well. On top of demonstrating the diagnostic utility of RNA-seq, these findings will have a direct impact on the affected individuals and their families, improve diagnostic yield for CMs, inform disease progression and risk management, and lead to better overall care.

## Introduction

Congenital myopathies (CMs) are a spectrum of neuromuscular disorders that are typically characterized by hypotonia, weakness, distinctive histopathological findings on muscle biopsy,

and elevated creatine phosphokinase levels (1). While adult-onset cases have been reported (215), these diseases primarily affect the pediatric population with an incidence rate of 1:25,000 (3) to 1:90,000 (4). In the past, clinical presentation and histopathological features had been used to categorize CMs into four major subgroups: 1) core myopathies; 2) nemaline myopathies; 3) centronuclear myopathies; 4) congenital myopathies with fiber type disproportion. While this classification remains relevant for diagnosis (5, 6), the advent of genetic testing has revealed significant clinical overlap among the typical subclasses of CMs thereby challenging their definitions (7). Meanwhile accurate subgrouping is needed in order to optimize anticipatory guidance and to address symptomatic issues as no FDA approved treatments currently exist (7).

Modern genetic tests greatly hasten diagnosis and often bypass the need for more costly or invasive tests such as muscle biopsy (7, 8). Next-generation sequencing (NGS) targeted gene panels are designed to investigate a select set of genes or gene regions with known or suspected associations with a given phenotype or disease. Many studies have aimed to demonstrate the clinical utility of these panels and report estimates of diagnostic yield at 20-49% for neuromuscular diseases (8-11) and 26% for CMs specifically (11). While highlighting the incredible progress that has already been made in the field of gene discovery, these outputs also emphasize the many variants that remain to be discovered or classified. However, it is rare that these tests reveal no interesting candidate variants at all. Instead, inconclusive diagnostic gene panels often divulge variants of unknown significance (VUS), a term for genetic changes that have an unknown, if any, effect on health. Therefore, a pressing limitation of genetic testing is our ability to interpret the biological impact of the overwhelming number of VUSs identified by these tests. Variant evidence for pathogenicity is built upon population, functional, segregation, de novo, allelic, computational, and predictive data (12). Classification of novel variants helps to improve diagnostic yield of gene panel for congenital myopathy and contributes to our understanding of the clinical and histopathological similarities between subgroups.

We present a molecularly undiagnosed patient and her son who both displayed proximal and distal weakness from a young age. Gene panel derived from the mother was inconclusive

but flagged VUSs in three large polymorphic genes: *SYNE1*, *SYNE2*, and *TTN*. In this study, we used NGS RNA-sequencing (RNA-seq) to identify and assess functional impact of candidate variants in these patients. This technology uniquely sequences the active transcriptome of the target tissue with up to 70% sensitivity (13). The advantage of RNA-seq is that it also provides insight unto gene expression, allele-specific expression, gene splicing, and fusions (14, 15) which can provide clues to help interpret the significance of VUSs. While RNA-seq of muscle biopsy tissue from the mother did not reveal evidence of pathogenicity for the variants flagged by gene panel, it did reveal an expression profile consistent with type I fiber predominance.

The most common cause of CMs are variants in ryanodine receptor 1 (*RYR1*). RYR1 is a hemotetrameric 2.5 MDa calcium channel located on the sarcoplasmic reticulum of muscle fibers that plays an intimate role in calcium signaling during excitation-contraction coupling (16). Being one of the largest human genes consisting of 5,038 amino acids, *RYR1* is particularly prone to polymorphisms and is commonly flagged for VUSs during genetic testing (17). Indeed, up to 700 *RYR1* variants have been discovered, though many remain to be functionally validated (18). Guided by the expression analysis, we also identified a candidate variant in *RYR1* that was seemingly overlooked by gene panel despite its known association with CMs and type I fiber predominance. We subsequently performed functional analyses of *C. elegans* carrying the variant in the homologous gene *unc-68* in order to validate pathogenicity. This model demonstrated early onset paralysis, altered synaptic transmission, reduced longevity, smaller body size and a decreased motility which clearly confirms pathogenicity of the variant in *C. elegans*. Therefore, we conclude that this mutation likely has consequences for human carriers as well.

## Methods

### Patients and genetic testing

The proband was evaluated by experienced neurologists and had a muscle biopsy has part of her clinical workup. The study was approved by the McGill University Health Centre Research Ethics Board. The proband has signed an inform consent authorizing genetic analysis

in a research setting. Prior to enrollment, the proband undergone a gene panel looking at 176 myopathy/muscular dystrophy genes (MNG Laboratories, Atlanta, GA, USA)

**RNA-Sequencing**

RNA was extracted from snap-frozen muscle tissue using trizol (Invitrogen). Library generation was performed using TruSeq stranded mRNA library preparation kit (Illumina, San Diego, CA USA), and sequencing was performed on an Illumina HiSeq 2500 using 125bp paired-end reads and 0,25 lanes of data per sample.

Data analysis was performed using an in-house pipeline combining publicly available tools and custom scripts. FASTQ files were aligned to the reference genome (Hg19) using STAR (19). Variant calling was performed with GATK and annotated using ANNOVAR as well as custom scripts (20, 21). This pipeline allowed for the detection of multinucleotide variants (MNV), indels, and single nucleotide variants including synonymous, non-synonymous and splice junction variants. Variants were filtered according to minor allele frequencies, excluding variants with a MAF >5% in either the 1000 Genomes Project or GnomAD. The data was further filtered to keep protein-damaging variants (nonsense, missense, frameshift, indels and splice variants). Read counts were obtained using featureCounts (22), and differential expression analysis was carried out against three healthy controls using DESeq2 with p-value and fold change thresholds of 0.05 and 1.5 respectively (23).

***C. elegans* Strains**

*C. elegans* maintenance was performed as per standard methods (24). All strains were grown at 20˚C on nematode growth agar (NGM) plates seeded with *E. coli* OP50. The strain PHX2444 *unc-68*(*syb2444*) containing the *unc-68* missense mutation *unc-68* (p.S4273L) was generated by SunyBiotech (http://www.sunybiotech.com/) using CRISPR-Cas9 mediated genome editing of N2 animals. Forward (CGGTTATGGACTTCTACTGG) and reverse (CCATCTTGGTGTCACTATCC) PCR and sequencing primers were used to confirm the *syb2444* mutation by the company itself (Sup. Fig. 5). The other strains included in this study are N2 and CB307 *unc-47*(*e308*) were obtained from the Caenorhabditis Genetics Center (CGC). All

experiments were performed in triplicate at room temperature (22°C) using age-synchronized N2 and *syb2444* adult worms. Only for the aldicarb assay was the CB307 strain used in addition. *C. elegans* lost to bagging or ruptured vulva were censored.

**Synchronization**

Ten to fifteen gravid adult worms were placed on a fresh plate seeded with OP50, left overnight to lay eggs, and removed the following day. Animals hatched from the eggs were then left to grow until the L4 stage was reached after which a pool of L4 animals was placed on a fresh plate seeded with OP50. The following day, day 1 adult worms were picked from this pool to start new experiments.

**Paralysis Assay**

105 day 1 adult hermaphrodite worms were transferred to three OP50 plates and scored daily for paralysis in a binary fashion. Worms were considered paralyzed if, after prodding by a platinum wire, they displayed restricted movement isolated to the head area. Worm strains observed included *syb2444* and N2 from which 274 and 262 events were counted, respectively.

**Lifespan Assay**

105 day 1 adult hermaphrodite worms were transferred to 3 OP50 plates and scored binarily daily. Worms were prodded with a platinum wire and considered dead if they failed to move. Lifespan is defined as the period from day 1 of adulthood until death. 270 and 257 events were observed in *syb2444* and N2 worms, respectively.

**Aldicarb Assay**

NGM plates were spiked with 1mM aldicarb, seeded with OP50, and allowed to dry under a fume hood on the day the experiment was to take place. 105 day 1 adult hermaphrodites were transferred to three aldicarb spiked plates, prodded with a platinum wire and scored binarily for paralysis every 30 minutes. 301, 301, and 308 events were observed for CB307, PHX2444 and N2 strains, respectively.

**Motility Assays**

105 day 1 adult hermaphrodites were transferred to 3 blank NGM plates and video recorded crawling for 30 seconds using a digital stereomicroscope on the lowest magnification (S9i; Leica, Heerbrugg, Switzerland; equipped with 10x/23 eyepieces and an integrated 10 M.P camera with 1080HD video display). Worms were subsequently suspended in 150µL of M9 buffer and video recorded swimming for another 30 seconds before being transferred to plates seeded with OP50. This process was repeated with the same worms for days 1, 5, and 7 of adulthood. 117-243 *syb2444* worms and 179-258 N2 worms were filmed in the crawling assays, and 128-279 *syb2444* worms and 194-242 N2 worms were filmed in the swimming assays. Notably, the number of worms recorded on each day decreased as worms died over time. Videos were uploaded onto WormLab software (Update 2020.1.1, MBF Bioscience, VT, USA) and tracked using parameters described in Sup. Table. 1 to generate a track summary of standard metrics described in the WormLab User Guide Version 2020 (https://www.mbfbioscience.com/help/wormlab/Content/home.htm). Results were exported to Graphpad Prism 9 and analyzed using two-way ANOVA followed by Turkey multiple comparisons test.

**Statistics**

For paralysis, aldicarb, and lifespan assays Kaplan-Meier curves were generated using Graphpad Prism 9 and statistical significance was analyzed using the Log-rank (Mantel-Cox) statistical test. For motility assays, video tracking results were exported to Graphpad Prism 9 and analyzed using two-way ANOVA followed by Turkey multiple comparisons test.

## Results

**Clinical characteristics and genetic testing**

The proband is a female patient, currently an adult, affected with a static myopathy since the age of 2 with proximal and distal weakness. During the pregnancy, a decrease in fetal movement in utero was reported. The patient also presented with a long face, facial weakness, and high arched palate. Delayed motor milestones and poor sport performance (unable to run normally and never able to skate or bike) were also observed. Although important muscle

weakness was present, creatine kinase levels were in the normal range. A modified trichrome-stained section of the muscle biopsy demonstrated the presence of rods (Fig. 1a). The electromyography demonstrated small polyphasic units but no fibrillations, positive sharp waves or myotonic discharges; thus, consistent with a myopathy diagnosis. Her affected son presented with similar symptoms leading us to suspect a dominant mode of inheritance (Fig. 1b).

DNA-based gene panel was conducted and resulted in an inconclusive diagnosis with VUSs in *SYNE1*, *SYNE2*, and *TTN*. The different clinical and pathological phenotypes associated with these genes could not clearly be associated with our patient's clinical presentation. Furthermore, all three genes contain many exons and are known to be highly polymorphic. To determine if any of these VUSs have a functional impact we conducted RNA-sequencing analysis.

**RNA-seq reveals a differential gene expression profile consistent with type I fiber predominance**

Skeletal muscle can be classified into two types of fibers: type I (slow twitch fibers) and type II (fast twitch fibers) (25). Under normal circumstances, these fibers are the same size. We performed a differential gene expression analysis from RNA of muscle biopsy from the proband which revealed a striking expression profile suggestive of type I fiber predominance. Specifically, we observed a clear over expression of slow twitch muscle fiber genes and under expression of fast twitch muscle fiber genes compared to three healthy controls (Fig. 1c). Several light and heavy myosin chains (*MYL2*, *MYL6B*, *MYH7*, and *MYH7B*) and all three troponin genes (*TNNC1*, *TNNT1*, *TNNI1*) specific to slow twitch muscle fibers were over expressed. Similarly, *MYL1*, *TNNC2*, *TNNT2*, and *TNNI2* which are specific to fast twitch muscle fibers were under expressed with a log2 fold change of at least -2.2 compared to controls. These genes play an important role in muscle contraction in both type I and type II fibers. Muscle contraction, which occurs through actin myosin filament sliding, is a process where actin and myosin work together in a cycle of bound and unbound states in order to shorten the muscle filament towards the M line of the sarcomere (26). Muscle myosin II consists of two heavy chain subunits, two alkali light chain subunits and two regulatory light chain subunits. In

a relaxed state where $Ca^{2+}$ concentrations are low, the conformation of troponin I prevents actin-myosin filament sliding and muscle shortening. However, during muscle contraction, RYR1 channels release calcium from the sarcoplasmic reticulum to rapidly increase cytoplasmic $Ca^{2+}$ concentrations from 10nM to 100nM during which the channel is maximally open (27). $Ca^{2+}$ ions then bind to troponin C causing a shift in the position of the troponin complex and revealing the actin binding sites required for actin-myosin cross-bridging. Meanwhile, troponin T operates as an anchor for the troponin complex to the muscle fibers (26). Subsequent analysis of the muscle biopsy confirmed a striking predominance of type I fibers (Fig. 1d) attesting to the huge potential of RNA-seq in diagnosis.

**RNA-seq uncovers a heterozygous missense variant in *RYR1***

Gene panel of the proband's DNA revealed VUSs in several large polymorphic genes including *SYNE1*, *SYNE2*, and *TTN* but was otherwise inconclusive. Further investigation of these genes through RNA-sequencing did not detect evidence of functional impact, including differential expression or alternative splicing. However, analysis of variant calling data guided by the findings from our differential gene expression analysis yielded a *de novo* heterozygous missense mutation in the gene *RYR1* (NM_000540.2, c.12083C>T, p.S4028L) that was not reported by gene panel despite the gene's known importance to CMs and associations with type I fiber predominance. This mutation affects a conserved amino acid residue in the cytoplasmic topological domain of RYR1 (28) and is predicted to be pathogenic by Polyphen (29), CADD (30) and Mutation Taster (31). While this variant is not accounted for in the gnomAD database, it has been reported in the LOVD database as pathogenic but with no supporting evidence (32). It has further been reported in ClinVar (33) (4 submissions) as a variant with conflicting interpretations of pathogenicity. One source described an 8-year-old boy with a CM from the age of 2 years, normal creatine kinase levels (109U/L), and no characteristic features on muscle biopsy. The *RYR1* c.12083C>T variant was identified in this patient using microarray-based NGS and verified by qPCR or Sanger sequencing (34). Another source suspected the c.12083C>T variant of pathogenicity in a female with a congenital myopathy who underwent targeted NGS sequencing but not muscle biopsy (35). Due to the highly polymorphic nature of *RYR1*, more evidence is needed to confirm pathogenicity of this variant.

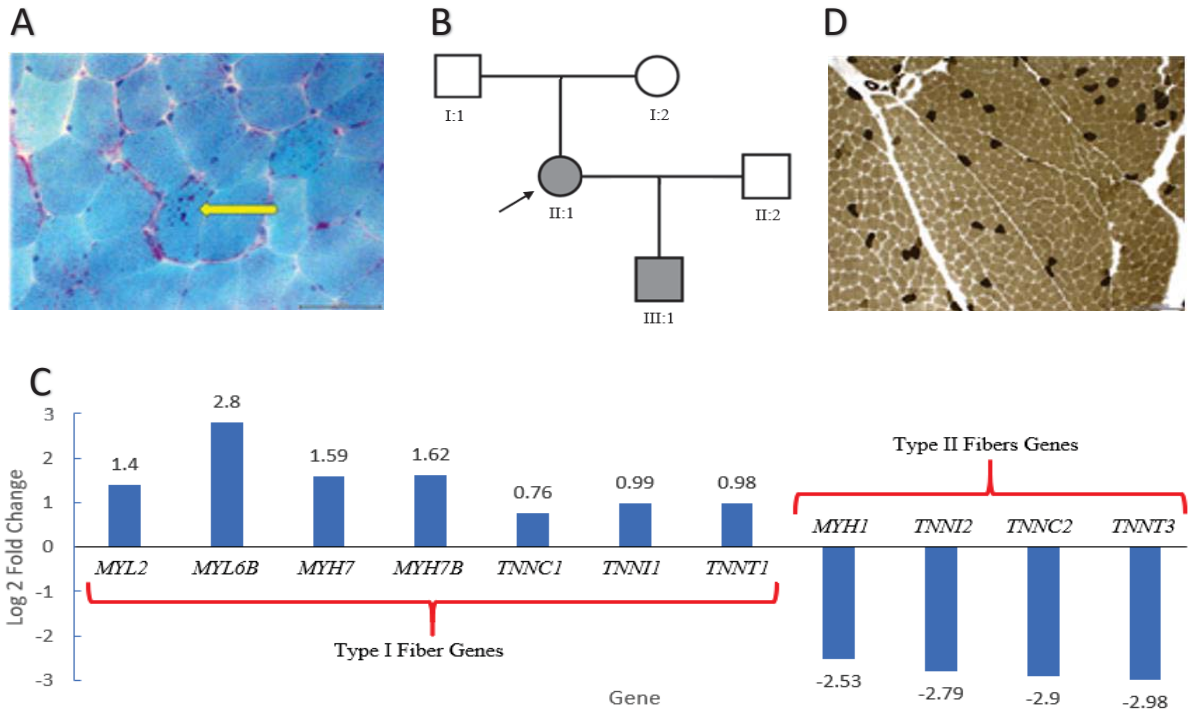



**Figure 1. Patient displays striking fiber type I predominance with rods**
**(A)** Modified gomori trichome-stained section of muscle biopsy demonstrates rods (yellow arrow). **(B)** Pedigree of the proband and her affected son. **(C)** RNA expression profile depicting over expression of type I fiber genes and under expression of type II fiber genes in the proband versus three healthy controls. This expression profile is suggestive of type I fiber predominance. **(D)** ATPase reacted section at low pH 4.3 showing a striking type I fiber predominance.

### *Syb2444* mutant *C. elegans* demonstrate a reduction in lifespan and overall health

*C. elegans* have been widely used as models of neuromuscular disease to study muscle function, development, and organization due to the high level of structural and functional conservation of the sarcomere they share with vertebrates (36). This likeness in combination with a complete and publicly available genome sequence and the rapidity of genome editing makes *C. elegans* a powerful genetic model for RYR related disorders (37). The p.S4028L mutation in human *RYR1* is equivalent to a p.S4273L mutation in the homologous gene *unc-68* of *C. elegans* (Sup. Fig. 6). We suspect the change from polar serine (S) to non-polar aliphatic leucine (L) will have damaging effects to the structural conformation of the RYR1 and Unc-68 proteins resulting in palpable phenotypes. The *C. elegans* strain PHX2444 *unc-68*(*syb2444*) containing this mutation was generated through CRISPR/Cas9 gene editing by SunyBiotech

(https://www.sunybiotech.com). On a superficial level, *syb2444* worms appeared small and sluggish, but with no noticeable developmental problems. In order to dissect the effect of the variant on the overall health of the mutant worms, we conducted a lifespan assay consisting of daily observations from day 1 of adulthood until each worm was no longer animate. Mutant worms demonstrated a significant reduction in lifespan having a median survival of 10 days of adulthood (Fig. 2a). Alternatively, N2 wild type (wt) worms displayed a median lifespan of 15 days of adulthood which is similar to what has been previously observed (38). Thus, the p.S4273L mutation clearly has a detrimental effect on longevity in *C. elegans*.

**_Syb2444_ mutant _C. elegans_ demonstrate early-onset paralysis and altered synaptic transmission**

*C. elegans* are thought to undergo a natural decline in muscle integrity and neuromuscular coordination as they age resulting in a gradual loss of motility and eventual paralysis (39). In fact, rate of paralysis has been proposed to be better indicator of life expectancy than chronological age (40). We measured the rate of paralysis to better understand the decline in longevity of mutant worms in relationship to the overall health of the neuromuscular system (Fig. 2b). A staggering difference was observed where *syb2444* mutants lost locomotive ability almost twice as fast as the wt having a median paralysis rate of 8 and 14 days of adulthood, respectively. Therefore, we suspect that the p.S4273L mutation has adverse effects on the neuromuscular system resulting in severe early-onset paralysis.

It has been shown that *unc-68* in *C. elegans* may play a presynaptic role involving neurotransmitter release at the neuromuscular junction (NMJ) (41). We hypothesized that the early-onset paralysis phenotype observed in *syb2444* worms could be associated with the neuronal functions of *unc-68*. Previous studies have demonstrated that arecoline, a muscarinic agonist, rescues locomotory defects in *C. elegans* through stimulation of acetylcholine (ACh) in motor neurons (39). Thus, we endeavored to elucidate the effects of the p.S4273L mutation on the efficacy of synaptic transmission. A muscle contraction is initiated by the secretion of ACh from the presynaptic terminal of motor neurons at the NMJ (42). Through a chain of downstream effects, this causes UNC-68 channels on the sarcoplasmic reticulum to release

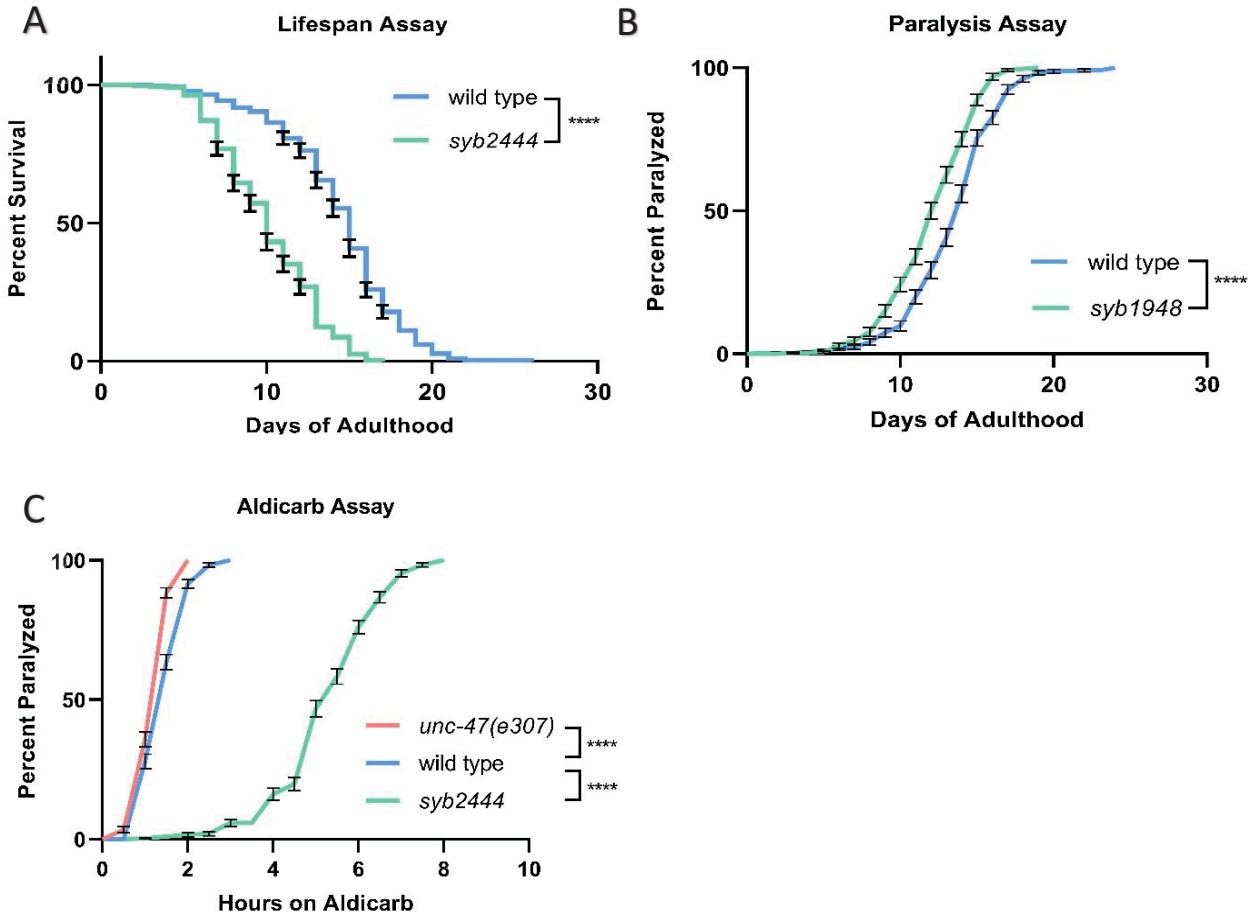



**Figure 2. The effect of *syb2444* on mortality and synaptic transmission**
**(A)** Lifespan and **(B)** paralysis of *syb2444* mutants ($n_{lifespan}$=270, censored=45; $n_{paralysis}$=274, 41) and N2 wild type ($n_{lifespan}$=257,58; $n_{paralysis}$=262, 53) *C. elegans*. **(C)** Rate of paralysis of *syb2444* mutants (n=301, 14), N2 wild type (n=308, 7), and positive control *unc-47(e307)* (n=301, 14) *C. elegans* on NGM agar plates spiked with 1mM aldicarb. All graphs are expressed as a percentage ±S.E., ns=non-significant; *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

calcium into the muscle cytosol and initiation excitation-contraction coupling (43). Normally, ACh is degraded by acetylcholinesterase in order to terminate muscle contraction (42). However, in the presence of the acetylcholinesterase inhibitor aldicarb, ACh begins to accumulate in the NMJ leading to muscle rigidity and eventual paralysis (42). Comparison of the time-course of aldicarb- induced paralysis provides insight into the relative efficiency of synaptic transmission of mutant and wt worms based on the correlation between neurotransmitter release and rate of paralysis (42).

*Syb2444*, wild type, and *unc-47(e307)* worms were observed on plates spiked with 1mM of aldicarb for a number of hours until paralysis occurred (Fig. 2c). The CB307 strain carrying mutant *unc-47* was used as a positive control due to its known hypersensitivity to aldicarb (44). While the median paralysis of *unc-47(e307)* and wild type worms occurred at about 1.5 hours on aldicarb, *syb2444* mutants demonstrated a salient hyposensitivity to aldicarb with a median survival occurring at 5.5 hours. Thus, the slower accumulation of ACh in the synaptic cleft of these animals compared to the wt and positive control is suggestive an effect of the p.S4273L variant on synaptic transmission. However, without further investigation such as an electrophysiological analysis or a levamisole assay, the cause of this hyposensitivity whether presynaptic or postsynaptic remains unknown (42, 45, 46).

**_Syb2444_ mutant _C. elegans_ exhibit a reduced body size**

We endeavored to characterize the morphology and locomotion of *syb2444* animals compared to the wt by recording *C. elegans* crawling on agar gel for 30 seconds at days 1,5, and 7 of adulthood using video microscopy and Wormlab software (MBF Bioscience). When viewed casually *syb2444* worms clearly demonstrate smaller body morphology than the wt. Indeed, video microscopy confirmed a 32%, 56%, and 62% reduction in the total body area of mutant worms compared to the wt on all days observed (Fig. 3a). In correlation with body area, mutant worms demonstrated both a stunted body length (20-51%) and width (11-22%) compared to their parental strain (Fig. 3b, c). Likewise, several mutations in *unc-68* have led to smaller body morphology (46) potentially due to disruption of one of the three gene pathways known to regulate body size in *C. elegans* (47). A recent study demonstrated that changes in the diameter but not length of *C. elegans* strongly correlate with muscle strength (48). Seeing how the p.S4273L mutation affects the width of mutant worms, we believed investigation of locomotive ability may help elucidate a muscle strength or muscle functionality phenotype.

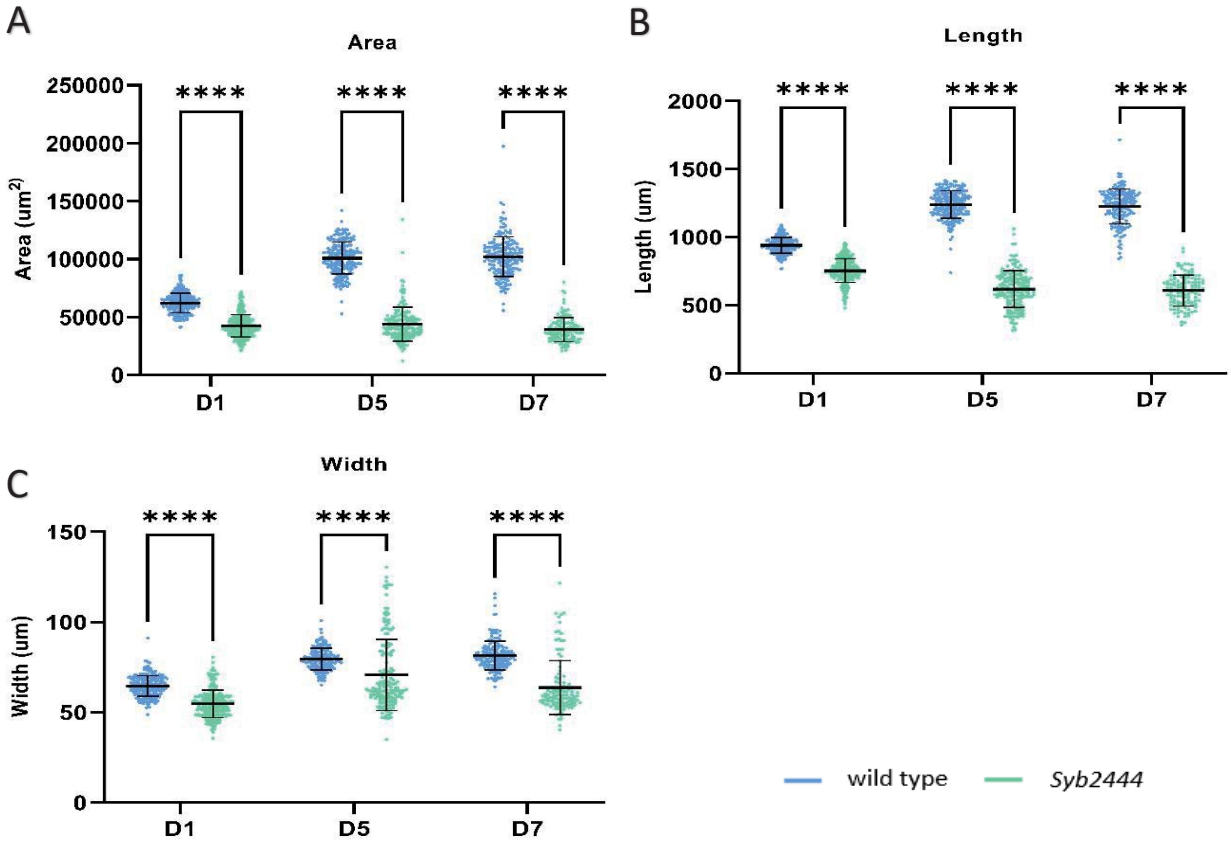



**Figure 3. The effect of *syb2444* on mortality and body size**
**(A)** Body area, **(B)** length, and **(C)** width of *syb2444* mutant (n=117-243) and N2 wild type (n=179-258) *C. elegans* on days 1, 5, and 7 of adulthood.

***Syb2444* mutant *C. elegans* demonstrate a reduction in motility while crawling on agar gel.**

*C. elegans* propel themselves forward on semi-solid substances by alternatively contracting striated muscle cells along the dorsal and ventral body wall thereby producing undulations (49). As a consequence of senescence, *C. elegans* undergo progressive muscle degradation similar to human sarcopenia (40). Age-dependent muscle weakness is thought to be a result of the oxidization of UNC-68 and the depletion of its binding protein FKB-2. These events cause the channel to become leaky draining the sarcoplasmic reticulum of its calcium stores (50). Moreover, expedited muscle aging, limited muscle strength, and decreases in muscle contraction rapidity have been observed in *C. elegans* with various *unc-68* mutations (51, 52). Therefore, using video microscopy, we explored the possibility of a muscle phenotype

through the characterization of locomotion of mutant *C. elegans* crawling on agar gel for 30 seconds at days 1,5, and 7 of adulthood.

*Syb2444* mutant worms demonstrated a notable reduction in speed compared to the wt beginning on day 1 of adulthood and continuing across all time points observed (-49%, -72%, -72%) (Fig. 4a). As would be expected of smaller animals, these worms had a consistently shorter average wavelength in their sinusoidal pattern by up to 55% compared to controls (Fig. 4b). Uncharacteristic of smaller body size however, mutant worms also demonstrated a larger mean and maximum wave amplitude compared to the wt strain by up to 47% and 60%, respectively (Fig. 4c, d). This change could instead explain the observed reduction in speed as the efficiency of forward movement would diminish as more time and energy is expended on movement perpendicular to the direction of motion. By way of explanation, one study predicts that increases in maximum amplitude or "hyper-bending" during backward motion may be a result of either increased contraction of the muscles within the bend of the worm or increased relaxation of the muscles opposite to the bend (53). However, with only the cumulative amplitude of both forward and backward movements given by our chosen assay, more precise analyses of *syb2444* worms would be required in order to ponder this explanation. *Syb2444* mutants also sustained backward movements for 36-48% of the time they were recorded which was 78%, 105%, and 62% more often as the N2 wt strain (Fig. 4e). This behavior is thought to be an avoidance mechanism and occurs both spontaneously and in response to environmental or intrinsic factors such as humidity, gravidity, sex, chemostimulation and mechanostimulation (54). Finally, *syb2444* animals also display a significant decrease (21-45%) in the number of body bends initiated per minute (Fig. 4f). Measuring body bends is a standard assay used to quantify performance during non-strenuous, non-maximal locomotion (46, 55-57). A decrease in the rate of body bends can have behavioral or physical explanations affecting either the worm's willingness or ability to move. Altogether, this data suggests that the p.S4273L mutation has a detrimental impact on locomotion that can only partially be explained by smaller body size. The difference in physical and behavioral contributions to this phenotype remains to be elucidated.

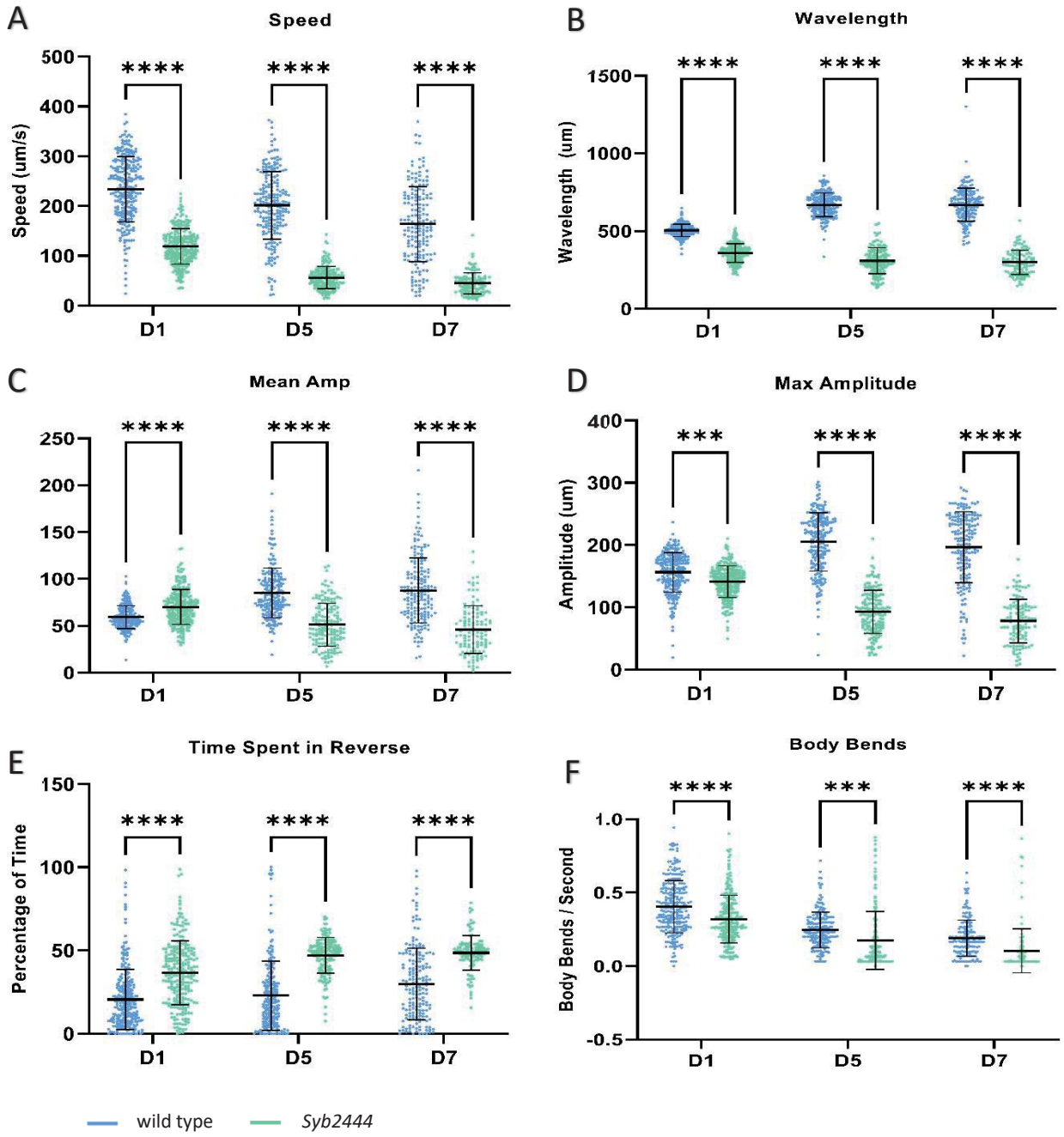



**Figure 4. Differences in motility between crawling PHX2444 and N2 *C. elegans***

Differences in **(A)** speed, **(B)** wavelength, **(C)** mean amplitude, **(D)** maximum amplitude, **(E)** reversal time, and **(F)** body bends of PHX24444 (n=117-243) and N2 (n=179-258) *C. elegans* crawling on agar gel on days 1, 5 and 7 of adulthood.

***Syb2444* mutant *C. elegans* display diminished locomotive ability while swimming**

The undulatory pattern of *C. elegans* movement as described by wavelength as a function of body length is dictated by the mechanical load of the medium through which the animal moves. As such, the S-shaped movements of worms crawling on agar gel and the C-shaped movements of worms thrashing in liquid media are thought to be extremities of a continuum of locomotory gaits (58). *C. elegans* are adept swimmers and when placed in viscous solutions the worms are inclined to thrash. Thus, study of locomotion in these media limits the influence of behavioral changes on movement. Swimming assays are especially important for the study of locomotion in strains carrying an *unc-68* mutation as they tend to travel shorter distances, undertake extended periods of immobility, and assume irregular extended or curled body postures on agar gel (59). S*yb2444* worms suspended in M9 buffer observed using video microscopy were much less motile than the wt. Not only were mutant worms dramatically slower (85-93%) than the wt (Fig. 5a), but the number of wave movements initiated or "thrashing" that occurred was 83-92% less frequent (Fig. 5b). This is consistent with other *unc-68 mutations*, including a null mutation, that also demonstrate a reduction in thrashing frequency compared to the wt (46, 59). Similar to body bend count on solid media, calculating thrashing frequency is a common assay for the indirect measurement of muscle function in swimming worms and is often used to study the effect of drugs, chemicals, and mutations on motility (60). This significant decline in motility of *syb2444* worms was consistent with our observations in the crawling assays and is suggestive of a physical rather than behavioral deficit.

Superficially, *syb2444* worms were also observed to pause and assume irregular curled positions resembling the letter "O" or the number "6" even in liquid solution. Video microscopy revealed that mutant worms spent up to 14% of the time recorded in a curled position (Fig. 5c). Comparatively, the highest percentage of time the N2 strain spent in a curled position was 0.11% on day 1. Frequent curling has been observed before in some *unc-68* mutant strains (59) and is known to occur in *C. elegans* exposed to most anthelmintic drugs such as pyrantel, levamisole, and tibendimidine (61). Unsurprisingly, the area "painted" by the body of the worm as it swims, known as brush stroke (Fig. 5d), was also reduced in mutant worms even when normalized to the time take to perform two complete swim strokes (Fig. 5e). These data

suggest that the p.S4273L mutation in *syb2444* worms has a detrimental effect on high-frequency locomotion due to physical deficits.

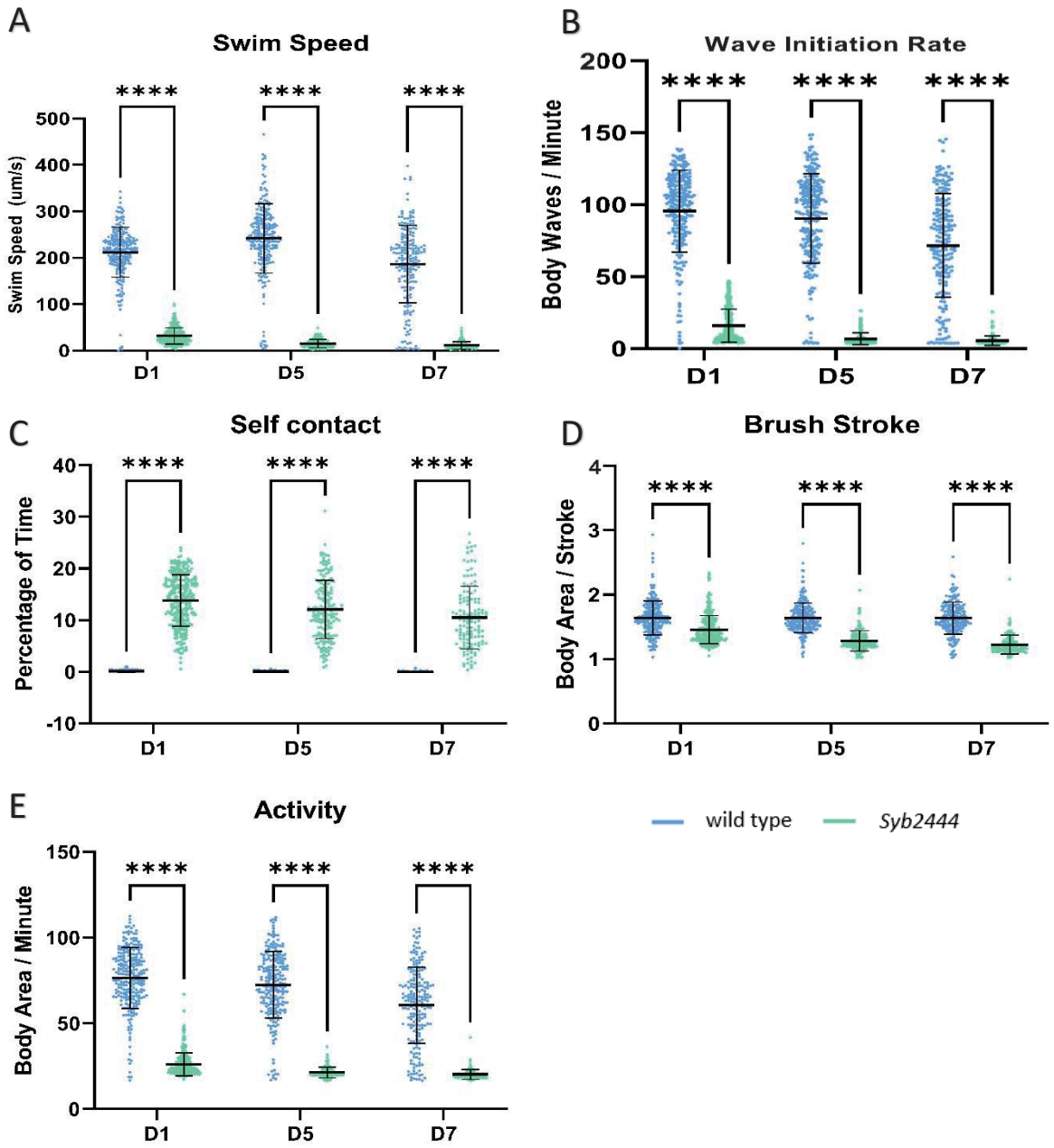



**Figure 5. Differences in motility between swimming PHX2444 and N2 *C. elegans***
Differences in **(A)** swim speed, **(B)** thrashing frequency, **(C)** self-contact time, **(D)** brush stroke, **(E)** and activity of PHX24444 (n=128-279) and N2 (n=194-242) *C. elegans* swimming in M9 buffer on days 1, 5 and 7 of adulthood.

## Discussion

CMs are a group of rare disorders with an estimated incidence rate of 1:25,000 (3) to 1:90,000 (4). Variants in *RYR1* are the most common cause of CM. According to Clinvar, up to 1691 variants in *RYR1* have been discovered, 10% of which have been classified as pathogenic or likely pathogenic. Alternatively, 79% of these variants are classified as VUS. The highly polymorphic nature of *RYR1* and lack of supporting evidence for unique variants precludes a clear molecular diagnosis in many patients. This work highlights the importance of validating the pathogenicity of suspected gene mutants using functional analyses in the context of large polymorphic disease-associated genes. Variant classification not only improves diagnostic yield and anticipatory guidance, but also contributes to our understanding of genetic, clinical, and histopathological features associated with type I fiber predominance.

The diagnosis of CMs is ever reliant on the study of muscle morphology, EMG, MRI, muscle ultrasound, serum creatine kinase, and nerve conduction (62). Genetic tests are increasingly used in a clinical setting and have greatly contributed to the discovery of 32 genetic causes for CM and an average diagnostic hit rating of >50% for gene panels (63). Many studies attempt to fill this diagnostic gap using WES or WGS for gene and variant discovery and are generally fruitful. The diagnostic shortfall originates in our fundamental lack of knowledge surrounding the functional impact of genetic variants leading to limited interpretability. While WGS provides an all-encompassing look into the genome, the large dataset generated is often overbearing. While WES focuses on the coding regions of the genome where most pathogenic variants are thought to lie, the intronic and non-coding regulatory elements such as promoters, enhancers, and microRNAs that could harbor up to 9-30% of pathogenic variants are overlooked (62, 64). Similarly, RNA-seq provides an even more targeted glance at the actively transcribed coding regions of the genome specific to the source material (65); moreover, the functional effect (i.e., alternative splicing) of variants hidden in non-coding regions can also be elucidated. While this necessitates use of the affected tissue which can be a limiting factor for certain diseases, it also helps to narrow the search for candidate variants and improves diagnostic yield by providing a wealth of information in a selective dataset (64). In this study, analysis of variant calling and differential gene expression outputs generated by the cost-

effective RNA-seq (65, 66) allowed for the build-up of evidence towards the type I fiber predominance phenotype and the causative *RYR1* variant observed in our patient. Interestingly, the variant c.12083C>T (p.S4028L) in *RYR1* was detected by gene panel but was filtered out of the reported list of interesting candidate VUSs. While the patient had not yet been diagnosed with type I fiber predominance but instead was described as having proximal and distal weakness at the time the gene panel was performed, this lapse is surprising considering the importance of *RYR1* variants to CMs and other myopathic disorders. To our knowledge, this is also the first-time type I fiber predominance has been observed on the gene expression level. These findings attest to the utility of RNA-seq in building evidence for variant interpretation. With further research, this expression profile could prove useful as a biomarker for type I fiber predominance and histologically related diseases like congenital myopathy with type I fiber predominance or congenital fiber type disproportion.

*C. elegans* serve as an excellent rapid, low-cost, first pass model for translational research in neuromuscular diseases. *C. elegans* and humans not only share a high level of genetic conservation (36), but the structure, function and composition of the sarcomere are also well preserved (67). In fact UNC-68 (NP_001256074.1) the human homolog Ryr1 share approximately 47 percent identity (68). In combination with the simplicity of *C. elegans* husbandry and rapidity of genome editing, *C. elegans* are valuable preclinical models for RYR related disorders (37). In this study, we find that *syb2444* mutant *C. elegans* demonstrate early onset paralysis, hyposensitivity to aldicarb, reduced lifespan, decreased size, impaired locomotive ability. Similarly, many other *C. elegans* strains carrying *unc-68* mutations corresponding to human myopathic variants demonstrate decreased lifespan, size, wavelength, and amplitude while crawling, altered thrashing frequency during swimming, hyposensitivity to aldicarb and other motility defects (37, 46, 51, 52, 59). These findings are relatively consistent with our own. Other phenotypes observed in *unc-68* mutant strains generally included halothane and caffeine hypersensitivity, levamisole hyposensitivity, disorganization of the muscle structure, and defective pharyngeal pumping. Based on the data collected from our *C. elegans* model, we propose that the p.S4273L variant in *unc-68* is pathogenic in *C. elegans* and

that the homologous p.S4028L variant in *RYR1* variant is likely the cause of our patients' phenotype.

While *C. elegans* encode only a single *unc-68* gene, mammals have three separate RYR genes (*RYR1*, *RYR2*, *RYR3*) that correspond to functions in skeletal, cardiac, and nervous tissue respectively (37). However, six tissue-specific isoforms with mutually exclusive first exons (exon 1.1 or exon 1.2) are thought to allow *C. elegans* to carry out the various functions performed by the RYR genes (37). Evidence suggests that exon 1.1 and exon 1.2 containing isoforms are predominantly expressed in body wall muscle and neurons, respectively. Interestingly, *C. elegans* mutants lacking exon 1.1 demonstrated reduced thrashing rate in liquid media while those lacking exon 1.2 containing isoforms performed similar to the wt suggesting the latter isoforms may be unnecessary for swimming. Because *syb2444* mutants also demonstrate reduced thrashing rate during swimming locomotion it is possible the p.S4273L variant influences exon 1.1 containing isoforms.

Alternatively, both exon 1.1 and 1.2 containing isoforms can be found in neuronal tissue in an approximate 1:9 ratio (37). *Unc-68* has been found to play both presynaptic and postsynaptic roles including regulation of normal quantal size at the NMJ (41). Specifically, *unc-68* was suggested to be involved in neurotransmitter release by means of several proposed mechanisms still under study: 1) increasing neurotransmitter vesicle size, 2) increasing the number of docked vesicles, 3) promoting vesicle loading with neurotransmitter, 4) stimulating intervesicular fusion, or 5) increasing the size or duration of the fusion pore during kiss-and-run synaptic release. In mice, presynaptic RYR receptors play a defined role in the regulation of neurotransmitter ACh quanta at the NMJ (69). *Syb2444* mutant worms display early-onset paralysis and decreased sensitivity to aldicarb suggesting a reduction in Ach release at the NMJ. Therefore, this mutation may also affect 1.2-containing *unc-68* isoforms. Human RYR1 is almost exclusively expressed in skeletal muscle, which could explain the lack of a paralytic phenotype in our human cohort and among other RYR1-associated CMs. Translatability is a limitation of all non-human models; however, the data generated by our *C. elegans* assays still relays important information on gene interactions and pathways. Nonetheless, our approach could benefit from

a force-based assay (48) or observation of muscle tissue organization (51) that would provide direct insight unto muscle strength and integrity.

*C. elegans* are highly amenable to genetic manipulation and with the advent of CRISPR/Cas9 gene editing, RNA interference and transgenics, more personalized genetic models based on *C. elegans* such as PHX2444 *unc-68*(*syb2444*) are coming into existence (51, 70). Consequently, *C. elegans* are emerging as tools for preclinical drug discovery. Their short generation time, small body size, large brood size, large repertoire of scorable phenotypes, and cellular and physiological relevance with humans makes them more suitable for large scale testing using compound libraries (71). Already, these invertebrates have been used for high-throughput drug screening for a plethora of diseases such as Parkinson's disease (72), Alzheimer's disease (73), and RYR1 related myopathies (74), some with encouraging reports of phenotypic rescue. The most common complaints with congenital myopathies are muscle weakness and fatigue which, to varying degrees depending on severity, may interrupt daily life, endanger an individual's independence, or even threaten their life. Therefore, effective therapies that address not only individual symptoms but also the disease as a whole are sorely needed. Application of high-throughput compound screening of personalized genetic models conferring human genetic variants may be the key to novel treatments for myopathic disorders.

**References**


1. Nance, J.R., et al., *Congenital myopathies: an update.* Curr Neurol Neurosci Rep, 2012. **12**(215): p. 165-74.
2. Nicolau, S., et al., *Congenital myopathies in the adult neuromuscular clinic: Diagnostic challenges and pitfalls.* Neurol Genet, 2019. **5**(4): p. e341.
3. Cassandrini, D., et al., *Congenital myopathies: clinical phenotypes and new diagnostic tools.* Ital J Pediatr, 2017. **43**(1): p. 101.
4. Amburgey, K., et al., *Prevalence of congenital myopathies in a representative pediatric united states population.* Ann Neurol, 2011. **70**(4): p. 662-5.
5. Maras Genc, H., A. Guven, and B. Talim, *Diagnostic yield of muscle biopsy in infants: Retrospective analysis of clinical and histopathologic findings.* Clin Neuropathol, 2021.
6. Grenho-Rodrigues, S., et al., *Diagnostic yield of muscle biopsies in pediatric population: a tertiary center experience.* Rev Neurol, 2021. **72**(8): p. 283-287.
7. Butterfield, R.J., *Congenital Muscular Dystrophy and Congenital Myopathy.* Continuum (Minneap Minn), 2019. **25**(6): p. 1640-1661.
8. Chae, J.H., et al., *Utility of next generation sequencing in genetic diagnosis of early onset neuromuscular disorders.* J Med Genet, 2015. **52**(3): p. 208-16.

9. Winder, T.L., et al., *Clinical utility of multigene analysis in over 25,000 patients with neuromuscular disorders.* Neurol Genet, 2020. **6**(215): p. e412.
10. Wu, L., et al., *Next-Generation Sequencing to Diagnose Muscular Dystrophy, Rhabdomyolysis, and HyperCKemia.* Can J Neurol Sci, 2018. **45**(3): p. 262-268.
11. Nishikawa, A., et al., *Targeted massively parallel sequencing and histological assessment of skeletal muscles for the molecular diagnosis of inherited muscle disorders.* J Med Genet, 2017. **54**(215): p. 104-110.
12. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.* Genet Med, 2015. **17**(5): p. 405-24.
13. Piskol, R., G. Ramaswami, and J.B. Li, *Reliable identification of genomic variants from RNA-seq data.* Am J Hum Genet, 2013. **93**(4): p. 641-51.
14. Cummings, B.B., et al., *Improving genetic diagnosis in Mendelian disease with transcriptome sequencing.* Sci Transl Med, 2017. **9**(386).
15. Kremer, L.S., et al., *Genetic diagnosis of Mendelian disorders via RNA sequencing.* Nat Commun, 2017. **8**: p. 15824.
16. Zalk, R., et al., *Structure of a mammalian ryanodine receptor.* Nature, 2015. **517**(7532): p. 44-9.
17. North, K.N., et al., *Approach to the diagnosis of congenital myopathies.* Neuromuscul Disord, 2014. **24**(215): p. 97-116.
18. Litman, R.S., et al., *Malignant Hyperthermia Susceptibility and Related Diseases.* Anesthesiology, 2018. **128**(1): p. 159-167.
19. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
20. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.
21. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
22. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.
23. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
24. Brenner, S., *The genetics of Caenorhabditis elegans.* Genetics, 1974. **77**(1): p. 71-94.
25. Engel, W.K., *The essentiality of histo- and cytochemical studies of skeletal muscle in the investigation of neuromuscular disease. 1962.* Neurology, 1998. **51**(3): p. 655 and 17 pages following.
26. Cooper, G.M., *The Cell: A Molecular Approach*. 2nd edition ed. 2000, Sunderland (MA): Sinauer Associates.
27. Bezprozvanny, I.B., et al., *Activation of the calcium release channel (ryanodine receptor) by heparin and other polyanions is calcium dependent.* Mol Biol Cell, 1993. **4**(3): p. 347-52.

28. UniProt, C., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.
39. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
30. Kircher, M., et al., *A general framework for estimating the relative pathogenicity of human genetic variants.* Nat Genet, 2014. **46**(3): p. 310-5.
31. Schwarz, J.M., et al., *MutationTaster2: mutation prediction for the deep-sequencing age.* Nat Methods, 2014. **11**(4): p. 361-2.
32. Fokkema, I.F., et al., *LOVD v.2.0: the next generation in gene variant databases.* Hum Mutat, 2011. **32**(5): p. 557-63.
33. Information, N.C.f.B., *ClinVar; [VCV000201152.7]*.
34. Dai, Y., et al., *A comprehensive genetic diagnosis of Chinese muscular dystrophy and congenital myopathy patients by targeted next-generation sequencing.* Neuromuscul Disord, 2015. **25**(8): p. 617-24.
35. Stehlikova, K., et al., *Muscular dystrophies and myopathies: the spectrum of mutated genes in the Czech Republic.* Clin Genet, 2017. **91**(3): p. 463-469.
36. Gieseler, K., H. Qadota, and G.M. Benian, *Development, structure, and maintenance of C. elegans body wall muscle.* WormBook, 2017. **2017**: p. 1-59.
37. Marques, F., et al., *Tissue-specific isoforms of the single C. elegans Ryanodine receptor gene unc-68 control specific functions.* PLoS Genet, 2020. **16**(10): p. e1009102.
38. Gems, D. and D.L. Riddle, *Defining wild-type life span in Caenorhabditis elegans.* J Gerontol A Biol Sci Med Sci, 2000. **55**(5): p. B215-9.
49. Glenn, C.F., et al., *Behavioral deficits during early stages of aging in Caenorhabditis elegans result from locomotory deficits possibly linked to muscle frailty.* J Gerontol A Biol Sci Med Sci, 2004. **59**(12): p. 1251-60.
40. Herndon, L.A., et al., *Stochastic and genetic factors influence tissue-specific decline in ageing C. elegans.* Nature, 2002. **419**(6909): p. 808-14.
41. Liu, Q., et al., *Presynaptic ryanodine receptors are required for normal quantal size at the Caenorhabditis elegans neuromuscular junction.* J Neurosci, 2005. **25**(29): p. 6745-54.
42. Oh, K.H. and H. Kim, *Aldicarb-induced Paralysis Assay to Determine Defects in Synaptic Transmission in Caenorhabditis elegans.* Bio Protoc, 2017. **7**(14).
43. Franzini-Armstrong, C., F. Protasi, and V. Ramesh, *Shape, size, and distribution of Ca(2+) release units and couplons in skeletal and cardiac muscles.* Biophys J, 1999. **77**(3): p. 1528-39.
44. Vashlishan, A.B., et al., *An RNAi screen identifies genes that regulate GABA synapses.* Neuron, 2008. **58**(3): p. 346-61.
45. Richmond, J.E., *Electrophysiological recordings from the neuromuscular junction of C. elegans.* WormBook, 2006: p. 1-8.
46. Graham, B., M.A. Shaw, and I.A. Hope, *Single Amino Acid Changes in the Ryanodine Receptor in the Human Population Have Effects In Vivo on Caenorhabditis elegans Neuro-Muscular Function.* Front Genet, 2020. **11**: p. 37.
47. Morck, C. and M. Pilon, *C. elegans feeding defective mutants have shorter body lengths and increased autophagy.* BMC Dev Biol, 2006. **6**: p. 39.

48. Rahman, M., et al., *NemaFlex: a microfluidics-based technology for standardized measurement of muscular strength of C. elegans.* Lab Chip, 2018. **18**(15): p. 2187-2201.
40. Liewald, J.F., et al., *Optogenetic analysis of synaptic function.* Nat Methods, 2008. **5**(10): p. 895-902.
50. Forrester, F., *The Role of Modified UNC-68 in Age-related Caenorhabditis elegans Muscle Function Loss*, in *Department of Philosophy*. 2018, Columbia University. p. 97.
51. Nicoll Baines, K., et al., *Aging Effects of Caenorhabditis elegans Ryanodine Receptor Variants Corresponding to Human Myopathic Mutations.* G3 (Bethesda), 2017. **7**(5): p. 1451-1461.
52. Sakube, Y., H. Ando, and H. Kagawa, *An abnormal ketamine response in mutants defective in the ryanodine receptor gene ryr-1 (unc-68) of Caenorhabditis elegans.* J Mol Biol, 1997. **267**(4): p. 849-64.
53. Nahabedian, J.F., et al., *Bending amplitude - a new quantitative assay of C. elegans locomotion: identification of phenotypes for mutants in genes encoding muscle focal adhesion components.* Methods, 2012. **56**(1): p. 95-102.
54. Zhao, B., et al., *Reversal frequency in Caenorhabditis elegans represents an integrated response to the state of the animal and its environment.* J Neurosci, 2003. **23**(12): p. 5319-28.
55. Wang, D. and X. Xing, *Assessment of locomotion behavioral defects induced by acute toxicity from heavy metal exposure in nematode Caenorhabditis elegans.* J Environ Sci (China), 2008. **20**(9): p. 1132-7.
56. Hart, A.C., *Wormbook* in *Behavior*, T.C.e.R. Community, Editor. 2006.
57. Hwang, H., et al., *Muscle contraction phenotypic analysis enabled by optogenetics reveals functional relationships of sarcomere components in Caenorhabditis elegans.* Sci Rep, 2016. **6**: p. 19900.
58. Gjorgjieva, J., D. Biron, and G. Haspel, *Neurobiology of Caenorhabditis elegans Locomotion: Where Do We Stand?* Bioscience, 2014. **64**(6): p. 476-486.
59. Maryon, E.B., R. Coronado, and P. Anderson, *unc-68 encodes a ryanodine receptor involved in regulating C. elegans body-wall muscle contraction.* J Cell Biol, 1996. **134**(4): p. 885-93.
60. Buckingham, S.D. and D.B. Sattelle, *Fast, automated measurement of nematode swimming (thrashing) without morphometry.* BMC Neurosci, 2009. **10**: p. 84.
61. Lycke, R., A. Parashar, and S. Pandey, *Microfluidics-enabled method to identify modes of Caenorhabditis elegans paralysis in four anthelmintics.* Biomicrofluidics, 2013. **7**(6): p. 64103.
62. Laura L. Smith, V.A.G., Alan H. Beggs, *Gene Discovery in Congenital Myopathy*. Stem Cell Biology and Regenerative Medicine, ed. K. Turksen. Vol. Regenerative Medicine for Degenerative Muscle Diseases. 2015, New York, NY: Human Press.
63. Gonorazky, H.D., C.G. Bonnemann, and J.J. Dowling, *The genetics of congenital myopathies.* Handb Clin Neurol, 2018. **148**: p. 549-564.
64. Gonorazky, H.D., et al., *Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease.* Am J Hum Genet, 2019. **104**(5): p. 1007.
65. Ku, C.S., et al., *Exome versus transcriptome sequencing in identifying coding region variants.* Expert Rev Mol Diagn, 2012. **12**(3): p. 241-51.

66. Djebali, S., et al., *Bioinformatics Pipeline for Transcriptome Sequencing Analysis.* Methods Mol Biol, 2017. **1468**: p. 201-19.
67. Duronio, R.J., et al., *Sophisticated lessons from simple organisms: appreciating the value of curiosity-driven research.* Dis Model Mech, 2017. **10**(12): p. 1381-1389.
68. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
69. Khuzakhmetova, V.F., Samigullin, D.V. & Bukharaeva, E.A., *The role of presynaptic ryanodine receptors in regulation of the kinetics of the acetylcholine quantal release in the mouse neuromuscular junction.* Biochemistry (Moscow) Supplement Series A: Membrane and Cell Biology, 2014. **8**: p. 144–152.
70. Sleigh, J., Sattelle DB, *C. elegans models of neuromuscular diseases expedite translational research.* Translational Neuroscience 2010. **1**: p. 214-227.
71. O'Reilly, L.P., et al., *C. elegans in high-throughput drug discovery.* Adv Drug Deliv Rev, 2014. **69-70**: p. 247-53.
72. Braungart, E., et al., *Caenorhabditis elegans MPP+ model of Parkinson's disease for high-throughput drug screenings.* Neurodegener Dis, 2004. **1**(4-5): p. 175-83.
73. EmelyneTeo, S.Y.J.L., Sheng Fong, Anis Larbi, Graham D.Wright, NicholasTolwinski, Jan Gruber, *A high throughput drug screening paradigm using transgenic Caenorhabditis elegans model of Alzheimer's disease.* Translational Medicine of Aging, 2020. **4**: p. 11-21.
74. Volpatti, J.R., et al., *Identification of drug modifiers for RYR1-related myopathy using a multi-species discovery pipeline.* Elife, 2020. **9**.

# Chapter 5: Discussion

## 5.1 The importance of variant identification and functional analyses

Despite recent advancements in next-generation DNA-sequencing technologies, many rare hereditary myopathies remain without a genetic diagnosis. Though individually rare, collectively these disorders affect tens of thousands of people worldwide representing a major burden to those affected as well as to the healthcare system. While genetic resolution is achieved in more than 50% of rare monogenic diseases, this number plummets to approximately 30% for those who do not receive a genetic diagnosis after an initial evaluation (218). The diagnostic success of the latter cohorts is largely dependent on the current level of understanding of disease-causing genes. That is to say that gene discovery acts in a positive feedback loop where the identification of novel pathogenic variants provides insight into gene-associated inheritance patterns and disease-mechanisms which informs the identification of subsequent variants. Despite having played a major role in the discovery of many disease-causing genes, modern approaches to gene discovery including WGS and WES are largely restricted by their limited ability to determine the functional effect of identified VUSs. Indeed, the number of variants identified by these techniques vastly exceeds our ability to interpret their impact, if any, on biological systems. Therefore, there is a pressing need for a more encompassing approach to variant identification and interpretation.

Our goal is to increase diagnostic yield of inherited myopathies and muscular dystrophies using a multi-omic approach for the identification and classification of novel variants in both known and new disease-causing genes. A successful molecular diagnosis is important as it ascertains access to the appropriate resources, informs prognosis and associated disease-risks, fosters genetic counseling, and confers psychosocial benefits to the patients and their families (218). Furthermore, access to targeted treatments and optimized disease management as a consequence of early diagnosis can have a significant impact on the development, progression and severity of symptoms (218). While the majority of myopathies lack disease-altering therapies, proper management based on a multidisciplinary approach can be an effective way of helping patients retain independence while limiting the secondary

effects of muscle weakness and surveilling possible comorbidities (219). However, because the optimal approach differs among and between diseases, a molecular diagnosis greatly facilitates the ability to tailor treatment to the individual needs of each patient.

## 5.2 Major findings

In this study, we underwent the identification and validation of potentially pathogenic variants among three cohorts of patients with myopathic, dystrophic or related presentations. Throughout these studies we demonstrate the utility of RNA-seq as a stand-alone tool and as a complementary diagnostic tool to traditional DNA-based sequencing techniques. For instance, in a cohort of three siblings with a congenital muscular dystrophy, we demonstrate how a combination of RNA-seq and WES allowed for the identification of a muscle specific allelic imbalance in the gene *IARS* along with a heterozygous missense mutation in the same gene where previous analysis by WES alone had been inconclusive. Similarly, gene panel had been unfruitful in a woman with a proximal and distal myopathy; however, analysis of RNA-seq data lead to the discovery of an expression profile suggestive of type I fiber disproportion which allowed for the detection of the causative *RYR1* variant. In these examples, the genetic abnormalities imparted by RNA-seq were used to guide variant analysis and ultimately pinpoint the causative variants. Therefore RNA-seq demonstrates much potential in the way of transgressing the interpretation bottleneck associated with DNA-based approaches to gene discovery.

In all cases, functional analyses of candidate variants was undertaken using *C. elegans* models made to be genetically representative of our patient cohorts using either RNAi or CRISPR/Cas9 gene editing. These models provided moderate to strong evidence for the pathogenicity of IARS knockdown and of two novel *RYR1* variants. Indeed, *iars-1* deficient *C. elegans* demonstrated histological abnormalities in the body wall muscle that progressively worsened with age. Although no motility defects were detected, this data confirms the presence of a muscle phenotype as a consequence of *iars-1* knockdown. Moreover, both *RYR1* variant models demonstrated defects in body morphology, lifespan, rate of paralysis, synaptic transmission, and motility. These phenotypes are highly suggestive of pathogenicity. As

demonstrated, *C. elegans* models offer an abundance of scorable phenotypes often with some degree of translatability to a human phenotype. Therefore, we advocate the use of *C. elegans* as an affordable, rapid, and coordinated first-pass evaluation of candidate variant pathogenicity for which homologous genes are available.

Each individual project brought new insights into novel disease-causing variants, gene mechanisms, or potential biomarkers for disease. Abnormalities in *IARS* have been associated with an array of multisystemic diseases including perinatal weak calf syndrome (144), GRIDHH (14, 145-147), refractory early-onset inflammatory bowel disease (148), and antisynthetase syndrome, most of which embody a skeletal muscle phenotype. While the role of *IARS* in protein synthesis is ubiquitous, whether it plays another role in muscle tissue is unknown. We observed that global knock-down of *iars-1* in *C. elegans* lead to progressive disorganization of myotubes. Therefore, *iars-1* may play an important role in the organization of the muscle structure. In the second and third study, we demonstrated the importance of functional validation of variants in known-disease causing genes, especially those that are highly polymorphic. We also proposed an explanation for the small body size observed in both our *RYR1* models and in several others (220). For the first time, *unc-68* was speculated to be involved in the calcineurin pathway which is known for its role in regulating body size in *C. elegans* (221, 222). Finally, we also discovered a gene expression profile suggestive of type I fiber predominance that may have utility as a potential biomarker for type I fiber predominance and associated diseases.

Identification of these novel variants provides closure for the participating patients and their families. Furthermore, these mutations and those found nearby are now more likely to be flagged on gene panels in future individuals suffering from muscle disorders. This is especially valuable in the case of IARS as it has not yet been associated with a muscle disorder. Finally, each of these studies provides a thorough outline of simple techniques used for the rapid classification of VUSs including animal models (functional data), variant prevalence (population data), and pathogenicity scores (predictive data) among others. While we focus on rare disorders, these techniques could easily be applied to the classification of VUSs in more common diseases.

## 5.3 Limitations of RNA-seq

We describe the utility of RNA-seq in detecting the functional impact of genetic variants in detail; however, a major challenge associated with this technique is its inability to detect intronic variants. Indeed, RNA-seq can only detect mutations in the coding regions representing approximately 1% of the genome (23), while 9-30% of pathogenic variants that affect RNA-expression or processing are found in the non-coding regions (223). We experienced this shortcoming first-hand when both RNA-seq and WES failed to detect a paternally inherited intronic variant (hg38 chr9:92249527) in the *IARS* gene suspected of having pathological relevance. In the ideal scenario, WGS and RNA-seq would be used in combination in order to overcome this constraint. However, this could prove to be an expensive endeavor for the depth of certain projects. Oftentimes it is sufficient to sequence the surrounding region of an identified transcriptomic anomaly using Sanger sequencing or targeted NGS in search of the causative variant.

In one particular study, we also demonstrate how RNA-seq provides a highly sensitive (56-58) and unbiased (58) picture of global gene expression. RNA-seq is thought to be a better alternative to traditional , hybridization-based approaches that are greatly restricted by pre-determined probe sets (78), non-specific hybridization, cross-hybridization and limited individual detection range (57). However, RNA-seq imparts its own limitations in regards to transcript quantification. First, small mRNAs may be difficult to count or missed entirely due to standard size selection during RNA-seq library preparation. Second, the accuracy of gene counts decreases the more and more genes share overlapping transcripts (224). Nonetheless, RNA-seq is at least as accurate as DNA microarrays, correlates very well with qPCR data (56), and remains the preferred approach for quantifying the transcriptome.

Perhaps the most important limitation of RNA-seq is the lack of standardization. Indeed, a surface-level evaluation of articles employing RNA-seq reveals a diversity of bioinformatics algorithms used for each phase of RNA-seq analysis (35, 224). Indeed, most laboratories, including our own, make use of custom-pipelines and scripts that are tailored to the research goals of a particular study. To understand the various tools available for each phase of RNA-seq

takes both time and computational knowledge (225), and while several standardized pipelines such as TRAPLINE (226) or ARPIR (225) exist to facilitate use by those without computer expertise, none are applied systematically. While lack of a standardized approach leads to complications of reproducibility and data reporting, no single pipeline can meet the need of all scientists. Rather, RNA-sequencing would benefit from a set of best practices to guide analysis and interpretation (35).

### 5.4 Limitations of *C. elegans* models

The simplicity of the *C. elegans* body plan is both an attractive and disadvantageous feature of this model. For example, their relatively simple nervous system is composed of only 302 neurons (103) compared to the billions found in the human body. This simplicity greatly facilitates the study of diseases or phenomena affecting this system. On the other hand, *C. elegans* lack many organs and such as a brain, blood, and eyes and organ systems such as a cardiovascular system that makes them unsuitable for the study of many human diseases (227). Meanwhile, differentiated muscle fiber types (slow and fast twitch) are absent in *C. elegans* models, yet we demonstrate that a variant associated with human type I fiber predominance modeled in these animals nonetheless elucidates palpable phenotypes.

Another double-edged sword is their small size. The mere 1mm length of *C. elegans* makes them exceptionally easy to store and inexpensive to feed. However, this feature also significantly complicates biochemical procedures, especially in a tissue-specific manner (227). Oftentimes, researchers must make use of whole worm extracts as isolation of a specific tissue is too complex. We experienced this complication when using RNAi to create a muscle specific knock-down of *iars-1*. While we would have liked to have confirmed this knock-down using RT-qPCR there was no simple or convenient method of isolating skeletal muscle tissue for RNA extraction. While it would have otherwise been possible to quantify *iars-1* using RT-PCR of pMYO-3::polyA-binding protein pull downs, this method is quite complex. Therefore, we had no option but to have confidence in the mechanism of RNAi and the 99% complementarity of our *iars*-1 RNAi to the target sequence.

Finally, to our advantage, human homologous genes have been found for up to 83% of the *C. elegans* proteome (168). While this coverage is incredible considering the evolutionary distance of these two species, there exist still some human genes for which no homolog exists. An example of this is the gene Nebulin Related Anchoring protein (*NRAP*), in which we identified a candidate variant in a cohort with adult-onset distal myopathy. This gene plays a structural role in the myofibrils and has been previously been associated with myofibrillar myopathy (226, 229). Because a homologous *NRAP* gene does not exist in *C. elegans*, we will instead be using a zebra fish model for the investigation of the identified variant.

## 5.5 Prospective endeavors

Accurate genetic knowledge permits the hypothesis of potential treatment options and their subsequent evaluation in animal models and *in vitro*. In these studies, we have established at least two *C. elegans* models with clearly perceptible phenotypes that are personalized representations of human disease. The next step would be high-throughput drug screening of these models in order to identify pharmaceuticals effective at alleviating symptoms for carriers of the variant under study. Already, these invertebrates have been used for high-throughput drug screening for a plethora of diseases such as Parkinson's disease (230), Alzheimer's disease (231), and RYR1 related myopathies (232), some with encouraging reports of phenotypic rescue. As no FDA approved treatments for myopathies currently exist, management of disease symptoms is key to preventing loss of independence and improving quality of life.

We will also pursue further investigations into the pathogenicity of the allelic imbalance of *IARS* identified in the first cohort. To determine if alterations in DNA-methylation could be mimicking a recessive disease, we will employ methyl-ATAC-seq in order to identify potential active regulatory sequences and their methylation state. Methylation profiles will be compared between patients and controls and between muscle and lymphocytes to detect muscle-specific modifications. We will also create an IARS knock-out model in myoblast cell lines using CRISPR/Cas9 genome editing as well as transduced myoblasts from patient fibroblasts. The impact of the *IARS* variants and depletion of synthetase activity will then be measured using a spectrometric assay of pyrophosphate release.

Finally, we will also be conducting further studies on the second cohort who present significant clinical heterogeneity despite sharing the genetic etiology. Phenotypic heterogeneity is not uncommon among RYR1-RMs and can partially be explained by the nature and location of *RYR1* mutations. However, differences in presentation among carriers of the same variant remain elusive in RYR1-RMs and other diseases. The homogeneity of this cohort, being carriers of the same mutation and originating form a founder population, represents a rare and ideal opportunity for study of clinical heterogeneity in relationship to a single variant. We will classify patients into distinct phenotypic subgroups according to their clinical presentation and severity. We will then investigate the presence of transcriptomic signatures associated with subgroups using RNA-sequencing in order to identify potential biomarkers that could be associated with phenotypes and disease severity. Likewise, we will also be investigating epigenomic signatures using whole genome bisulfite sequencing in order to identify differences in methylation profiles that segregate to different subgroups. Finally, we will integrate the methylation and RNA-sequencing data in order to evaluate the level of correlation between gene expression and methylation profile allowing us to determine if differentially expressed genes are a result of methylation modifications.

# Chapter 6: Conclusion

Throughout these studies, we have investigated three cohorts with undiagnosed myopathies or dystrophies despite exhaustive investigation. Identification of candidate variants was achieved through a combination of NGS techniques, while validation was pursued using *C. elegans* models. In three siblings with a CMD, we identified a muscle-specific imbalance along with a candidate causative variant in the gene *IARS* using a combination of WES and RNA-seq. While functional validation in *C. elegans* was inconclusive, we uncovered a potential role of *IARS* in myotube organization. In a patient with a proximal and distal myopathy and affected son, we uncovered a pathogenic heterozygous missense mutation in the gene *RYR1* using RNA-seq. This variant, when translated into *C. elegans* precipitated aberrant changes in lifespan, rate of paralysis, synaptic transmission, body morphology and motility. In fact, the latter two phenotypes were so severe they could be observed on a superficial level. Finally, in a cohort of patients with a range of clinical phenotypes, we identified a single common variant in the gene *RYR1* using a combination of WGS and gene panel. When this variant was modeled in *C. elegans* we again observed changes in lifespan, rate of paralysis, synaptic transmission, body morphology and motility. Therefore, we concluded that both *RYR1* variants are clearly pathogenic in *C. elegans* and thus likely have consequences for their human carriers as well.

We highlight the value of RNA-seq as a diagnostic tool for monogenic diseases. Twice we demonstrate how this approach allowed for the identification of genetic phenomena (i.e. allelic imbalance and atypical gene expression) that was used to guide candidate variant identification. We also emphasize the utility of *C. elegans* as a rapid, low-cost, first pass model for translational research in neuromuscular diseases. The information gained from these studies will contribute to a better diagnostic yield in patients with myopathic or dystrophic disorders, facilitate the discovery of more new pathogenic genes in the future, and have a direct impact on the participating individuals and their families whose diagnostic odyssey has come to an end.

# References


1. González-Jamett AM, B.J., María A, *Hereditary Myopathies*. 2018: IntechOpen.
2. Bodilsen, J., et al., *Early versus late diagnosis in community-acquired bacterial meningitis: a retrospective cohort study.* Clin Microbiol Infect, 2018. **24**(2): p. 166-170.
3. Cassandrini, D., et al., *Congenital myopathies: clinical phenotypes and new diagnostic tools.* Ital J Pediatr, 2017. **43**(1): p. 101.
4. Claeys, K.G., *Congenital myopathies: an update.* Dev Med Child Neurol, 2020. **62**(3): p. 297-302.
5. Cardamone, M., B.T. Darras, and M.M. Ryan, *Inherited myopathies and muscular dystrophies.* Semin Neurol, 2008. **28**(2): p. 250-9.
6. Wang, C.H., et al., *Consensus statement on standard of care for congenital myopathies.* J Child Neurol, 2012. **27**(3): p. 363-82.
7. Laura L. Smith, V.A.G., Alan H. Beggs, *Gene Discovery in Congenital Myopathy*. Stem Cell Biology and Regenerative Medicine, ed. K. Turksen. Vol. Regenerative Medicine for Degenerative Muscle Diseases. 2015, New York, NY: Human Press.
8. Gonorazky, H.D., C.G. Bonnemann, and J.J. Dowling, *The genetics of congenital myopathies.* Handb Clin Neurol, 2018. **148**: p. 549-564.
9. Ravenscroft, G., et al., *Recent advances in understanding congenital myopathies.* F1000Res, 2018. **7**.
10. Fraiman, Y.S. and M.H. Wojcik, *The influence of social determinants of health on the genetic diagnostic odyssey: who remains undiagnosed, why, and to what effect?* Pediatr Res, 2020.
11. Lawal, T.A., J.J. Todd, and K.G. Meilleur, *Ryanodine Receptor 1-Related Myopathies: Diagnostic and Therapeutic Approaches.* Neurotherapeutics, 2018. **15**(4): p. 885-899.
12. Jones, K., et al., *Interventions for promoting physical activity in people with neuromuscular disease.* Cochrane Database Syst Rev, 2021. **5**: p. CD013544.
13. Hedermann, G., et al., *Aerobic Training in Patients with Congenital Myopathy.* PLoS One, 2016. **11**(1): p. e0146036.
14. Smigiel, R., et al., *New evidence for association of recessive IARS gene mutations with hepatopathy, hypotonia, intellectual disability and growth retardation.* Clin Genet, 2017. **92**(6): p. 671-673.
15. Dowling, J.J., et al., *Oxidative stress and successful antioxidant treatment in models of RYR1-related myopathy.* Brain, 2012. **135**(Pt 4): p. 1115-27.
16. Todd, J.J., et al., *Randomized controlled trial of N-acetylcysteine therapy for RYR1-related myopathies.* Neurology, 2020. **94**(13): p. e1434-e1444.
17. Tetreault, M., et al., *Whole-exome sequencing as a diagnostic tool: current challenges and future opportunities.* Expert Rev Mol Diagn, 2015. **15**(6): p. 749-60.
18. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
19. Harding, K.E. and N.P. Robertson, *Applications of next-generation whole exome sequencing.* J Neurol, 2014. **261**(6): p. 1244-6.
20. Chin, E.L., C. da Silva, and M. Hegde, *Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations.* BMC Genet, 2013. **14**: p. 6.
21. Majewski, J. and D.S. Rosenblatt, *Exome and whole-genome sequencing for gene discovery: the future is now!* Hum Mutat, 2012. **33**(4): p. 591-2.
22. Ku, C.S., D.N. Cooper, and G.P. Patrinos, *The Rise and Rise of Exome Sequencing.* Public Health Genomics, 2016. **19**(6): p. 315-324.

23. Ng, S.B., et al., *Targeted capture and massively parallel sequencing of 12 human exomes.* Nature, 2009. **461**(7261): p. 272-6.
24. Rosenfeld, J.A., C.E. Mason, and T.M. Smith, *Limitations of the human reference genome for personalized genomics.* PLoS One, 2012. **7**(7): p. e40294.
25. Djebali, S., et al., *Bioinformatics Pipeline for Transcriptome Sequencing Analysis.* Methods Mol Biol, 2017. **1468**: p. 201-19.
26. Ku, C.S., et al., *Exome versus transcriptome sequencing in identifying coding region variants.* Expert Rev Mol Diagn, 2012. **12**(3): p. 241-51.
27. Byron, S.A., et al., *Translating RNA sequencing into clinical diagnostics: opportunities and challenges.* Nat Rev Genet, 2016. **17**(5): p. 257-71.
28. Cirulli, E.T., et al., *Screening the human exome: a comparison of whole genome and whole transcriptome sequencing.* Genome Biol, 2010. **11**(5): p. R57.
29. Consortium, G.T., *Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-60.
30. Gonorazky, H.D., et al., *Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease.* Am J Hum Genet, 2019. **104**(5): p. 1007.
31. Cummings, B.B., et al., *Improving genetic diagnosis in Mendelian disease with transcriptome sequencing.* Sci Transl Med, 2017. **9**(386).
32. Kremer, L.S., et al., *Genetic diagnosis of Mendelian disorders via RNA sequencing.* Nat Commun, 2017. **8**: p. 15824.
33. Zhong, J., et al., *Use of RNAsequencing to detect abnormal transcription of the collagen alpha2 (VI) chain gene that can lead to Bethlem myopathy.* Int J Mol Med, 2021. **47**(3).
34. Gonorazky, H., et al., *RNAseq analysis for the diagnosis of muscular dystrophy.* Ann Clin Transl Neurol, 2016. **3**(1): p. 55-60.
35. Conesa, A., et al., *A survey of best practices for RNA-seq data analysis.* Genome Biol, 2016. **17**: p. 13.
36. Park, Y.S., et al., *Comparison of library construction kits for mRNA sequencing in the Illumina platform.* Genes Genomics, 2019. **41**(10): p. 1233-1240.
37. Hrdlickova, R., M. Toloue, and B. Tian, *RNA-Seq methods for transcriptome analysis.* Wiley Interdiscip Rev RNA, 2017. **8**(1).
38. Head, S.R., et al., *Library construction for next-generation sequencing: overviews and challenges.* Biotechniques, 2014. **56**(2): p. 61-4, 66, 68, passim.
39. Piskol, R., G. Ramaswami, and J.B. Li, *Reliable identification of genomic variants from RNA-seq data.* Am J Hum Genet, 2013. **93**(4): p. 641-51.
40. Li, X., et al., *Quality control of RNA-seq experiments.* Methods Mol Biol, 2015. **1269**: p. 137-46.
41. Ku, C.S., et al., *Exome sequencing: dual role as a discovery and diagnostic tool.* Ann Neurol, 2012. **71**(1): p. 5-14.
42. Geniza M, J.P., *Tools for building de novo trasncriptome assembly.* Current Plant Biology, 2017. **11-12**: p. 41-45.
43. Park, E., et al., *The Expanding Landscape of Alternative Splicing Variation in Human Populations.* Am J Hum Genet, 2018. **102**(1): p. 11-26.
44. Sudhagar, A., G. Kumar, and M. El-Matbouli, *Transcriptome Analysis Based on RNA-Seq in Understanding Pathogenic Mechanisms of Diseases and the Immune System of Fish: A Comprehensive Review.* Int J Mol Sci, 2018. **19**(1).
45. Adetunji, M.O., et al., *Variant analysis pipeline for accurate detection of genomic variants from transcriptome sequencing data.* PLoS One, 2019. **14**(9): p. e0216838.
46. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.

47. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.
48. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
49. Karczewski, K.J., et al., *The mutational constraint spectrum quantified from variation in 141,456 humans.* Nature, 2020. **581**(7809): p. 434-443.
50. Kircher, M., et al., *A general framework for estimating the relative pathogenicity of human genetic variants.* Nat Genet, 2014. **46**(3): p. 310-5.
51. Schwarz, J.M., et al., *MutationTaster2: mutation prediction for the deep-sequencing age.* Nat Methods, 2014. **11**(4): p. 361-2.
52. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
53. Amberger, J.S., et al., *OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders.* Nucleic Acids Res, 2015. **43**(Database issue): p. D789-98.
54. Landrum, M.J., et al., *ClinVar: public archive of relationships among sequence variation and human phenotype.* Nucleic Acids Res, 2014. **42**(Database issue): p. D980-5.
55. Costa-Silva, J., D. Domingues, and F.M. Lopes, *RNA-Seq differential expression analysis: An extended review and a software tool.* PLoS One, 2017. **12**(12): p. e0190152.
56. Nagalakshmi, U., et al., *The transcriptional landscape of the yeast genome defined by RNA sequencing.* Science, 2008. **320**(5881): p. 1344-9.
57. Zhao, S., et al., *Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells.* PLoS One, 2014. **9**(1): p. e78644.
58. Dong, X., Y. You, and J.Q. Wu, *Building an RNA Sequencing Transcriptome of the Central Nervous System.* Neuroscientist, 2016. **22**(6): p. 579-592.
59. Dundar Friederike, S.L., Zumbo Paul, *Introduction to differential gene expression analysis using RNA-seq.* Weil Cornell Medical College, 2015.
60. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
61. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
62. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.
63. Huan, T., et al., *A whole-blood transcriptome meta-analysis identifies gene expression signatures of cigarette smoking.* Hum Mol Genet, 2016. **25**(21): p. 4611-4623.
64. Fernandes, M., A. Patel, and H. Husi, *C/VDdb: A multi-omics expression profiling database for a knowledge-driven approach in cardiovascular disease (CVD).* PLoS One, 2018. **13**(11): p. e0207371.
65. Bouquet, J., et al., *Longitudinal Transcriptome Analysis Reveals a Sustained Differential Gene Expression Signature in Patients Treated for Acute Lyme Disease.* MBio, 2016. **7**(1): p. e00100-16.
66. Cancer Genome Atlas Research Network. Electronic address, w.b.e. and N. Cancer Genome Atlas Research, *Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma.* Cell, 2017. **169**(7): p. 1327-1341 e23.
67. Chen, B., et al., *Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets.* Nat Commun, 2017. **8**: p. 16022.
68. Fu, J., et al., *Gene expression profiling leads to discovery of correlation of matrix metalloproteinase 11 and heparanase 2 in breast cancer progression.* BMC Cancer, 2015. **15**: p. 473.

69. Koch, C., et al., *Myostatin: a Circulating Biomarker Correlating with Disease in Myotubular Myopathy Mice and Patients.* Mol Ther Methods Clin Dev, 2020. **17**: p. 1178-1189.
70. Djeddi, S., et al., *Multi-omics comparisons of different forms of centronuclear myopathies and the effects of several therapeutic strategies.* Mol Ther, 2021.
71. Butterfield, R.J., et al., *Transcriptome profiling identifies regulators of pathogenesis in collagen VI related muscular dystrophy.* PLoS One, 2017. **12**(12): p. e0189664.
72. Luco, R.F., et al., *Epigenetics in alternative pre-mRNA splicing.* Cell, 2011. **144**(1): p. 16-26.
73. Kamps, R., et al., *Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification.* Int J Mol Sci, 2017. **18**(2).
74. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-6.
75. Nakka, K., et al., *Diversification of the muscle proteome through alternative splicing.* Skelet Muscle, 2018. **8**(1): p. 8.
76. Baralle, D. and E. Buratti, *RNA splicing in human disease and in the clinic.* Clin Sci (Lond), 2017. **131**(5): p. 355-368.
77. Costa, C., et al., *Comprehensive molecular screening: from the RT-PCR to the RNA-seq.* Transl Lung Cancer Res, 2013. **2**(2): p. 87-91.
78. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.
79. Kernohan, K.D., et al., *Whole-transcriptome sequencing in blood provides a diagnosis of spinal muscular atrophy with progressive myoclonic epilepsy.* Hum Mutat, 2017. **38**(6): p. 611-614.
80. Tang, J.Y., et al., *Alternative splicing for diseases, cancers, drugs, and databases.* ScientificWorldJournal, 2013. **2013**: p. 703568.
81. Savol, A.J., et al., *Genome-wide identification of autosomal genes with allelic imbalance of chromatin state.* PLoS One, 2017. **12**(8): p. e0182568.
82. Zhao, D., et al., *Characteristics of allelic gene expression in human brain cells from single-cell RNA-seq data analysis.* BMC Genomics, 2017. **18**(1): p. 860.
83. Jamieson, T.A., et al., *M6P/IGF2R loss of heterozygosity in head and neck cancer associated with poor patient prognosis.* BMC Cancer, 2003. **3**: p. 4.
84. Rose, A.M., et al., *Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa.* Sci Rep, 2016. **6**: p. 19450.
85. de la Chapelle, A., *Genetic predisposition to human disease: allele-specific expression and low-penetrance regulatory loci.* Oncogene, 2009. **28**(38): p. 3345-8.
86. Llavona, P., et al., *Allelic Expression Imbalance in the Human Retinal Transcriptome and Potential Impact on Inherited Retinal Diseases.* Genes (Basel), 2017. **8**(10).
87. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.* Genet Med, 2015. **17**(5): p. 405-24.
88. Fusto, A., et al., *Cored in the act: the use of models to understand core myopathies.* Dis Model Mech, 2019. **12**(12).
89. Kraeva, N., et al., *Novel excitation-contraction uncoupled RYR1 mutations in patients with central core disease.* Neuromuscul Disord, 2013. **23**(2): p. 120-32.
90. Lawal, T.A., et al., *Preclinical model systems of ryanodine receptor 1-related myopathies and malignant hyperthermia: a comprehensive scoping review of works published 1990-2019.* Orphanet J Rare Dis, 2020. **15**(1): p. 113.
91. Takeshima, H., et al., *Excitation-contraction uncoupling and muscular degeneration in mice lacking functional skeletal muscle ryanodine-receptor gene.* Nature, 1994. **369**(6481): p. 556-9.

92. Kornegay, J.N., *The golden retriever model of Duchenne muscular dystrophy.* Skelet Muscle, 2017. **7**(1): p. 9.
93. Howe, K., et al., *The zebrafish reference genome sequence and its relationship to the human genome.* Nature, 2013. **496**(7446): p. 498-503.
94. Sztal, T.E., et al., *Zebrafish models for nemaline myopathy reveal a spectrum of nemaline bodies contributing to reduced muscle function.* Acta Neuropathol, 2015. **130**(3): p. 389-406.
95. Gurung, R., et al., *A Zebrafish Model for a Human Myopathy Associated with Mutation of the Unconventional Myosin MYO18B.* Genetics, 2017. **205**(2): p. 725-735.
96. Sztal, T.E., et al., *Testing of therapies in a novel nebulin nemaline myopathy model demonstrate a lack of efficacy.* Acta Neuropathol Commun, 2018. **6**(1): p. 40.
97. Brenner, S., *The genetics of Caenorhabditis elegans.* Genetics, 1974. **77**(1): p. 71-94.
98. Gieseler, K., H. Qadota, and G.M. Benian, *Development, structure, and maintenance of C. elegans body wall muscle.* WormBook, 2017. **2017**: p. 1-59.
99. Yuan, J., et al., *The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme.* Cell, 1993. **75**(4): p. 641-52.
100. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-54.
101. Wightman, B., I. Ha, and G. Ruvkun, *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.* Cell, 1993. **75**(5): p. 855-62.
102. Hedgecock, E.M., et al., *Genetics of cell and axon migrations in Caenorhabditis elegans.* Development, 1987. **100**(3): p. 365-82.
103. White, J.G., et al., *The structure of the nervous system of the nematode Caenorhabditis elegans.* Philos Trans R Soc Lond B Biol Sci, 1986. **314**(1165): p. 1-340.
104. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.
105. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression.* Science, 1994. **263**(5148): p. 802-5.
106. Boulin, T. and O. Hobert, *From genes to function: the C. elegans genetic toolbox.* Wiley Interdiscip Rev Dev Biol, 2012. **1**(1): p. 114-37.
107. Conte, D., Jr., et al., *RNA Interference in Caenorhabditis elegans.* Curr Protoc Mol Biol, 2015. **109**: p. 26 3 1-26 3 30.
108. Neve, I.A.A., et al., *Escherichia coli Metabolite Profiling Leads to the Development of an RNA Interference Strain for Caenorhabditis elegans.* G3 (Bethesda), 2020. **10**(1): p. 189-198.
109. Papic, L., et al., *Double-stranded RNA production and the kinetics of recombinant Escherichia coli HT115 in fed-batch culture.* Biotechnol Rep (Amst), 2018. **20**: p. e00292.
110. Tabara, H., et al., *The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans.* Cell, 2002. **109**(7): p. 861-71.
111. Kamath, R.S. and J. Ahringer, *Genome-wide RNAi screening in Caenorhabditis elegans.* Methods, 2003. **30**(4): p. 313-21.
112. Qadota, H., et al., *Establishment of a tissue-specific RNAi system in C. elegans.* Gene, 2007. **400**(1-2): p. 166-73.
113. Firnhaber, C. and M. Hammarlund, *Neuron-specific feeding RNAi in C. elegans and its use in a screen for essential genes required for GABA neuron function.* PLoS Genet, 2013. **9**(11): p. e1003921.
114. Ishino, Y., et al., *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* J Bacteriol, 1987. **169**(12): p. 5429-33.

115. Stern, M.J., et al., *Repetitive extragenic palindromic sequences: a major component of the bacterial genome.* Cell, 1984. **37**(3): p. 1015-26.
116. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
117. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
118. Deveau, H., et al., *Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.* J Bacteriol, 2008. **190**(4): p. 1390-400.
119. Maggio, I., et al., *Integrating gene delivery and gene-editing technologies by adenoviral vector transfer of optimized CRISPR-Cas9 components.* Gene Ther, 2020. **27**(5): p. 209-225.
120. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-7.
121. Guo, T., et al., *Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing.* Genome Biol, 2018. **19**(1): p. 170.
122. Stewart, J., et al., *Modelling the Cancer Phenotype in the Era of CRISPR-Cas9 Gene Editing.* Clin Oncol (R Coll Radiol), 2020. **32**(2): p. 69-74.
123. Symington, L.S. and J. Gautier, *Double-strand break end resection and repair pathway choice.* Annu Rev Genet, 2011. **45**: p. 247-71.
124. Liu, M., et al., *Methodologies for Improving HDR Efficiency.* Front Genet, 2018. **9**: p. 691.
125. Arribere, J.A., et al., *Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans.* Genetics, 2014. **198**(3): p. 837-46.
126. Paix, A., et al., *High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes.* Genetics, 2015. **201**(1): p. 47-54.
127. Falsaperla, R., et al., *Congenital muscular dystrophy: from muscle to brain.* Ital J Pediatr, 2016. **42**(1): p. 78.
128. Fu, X.N. and H. Xiong, *Genetic and Clinical Advances of Congenital Muscular Dystrophy.* Chin Med J (Engl), 2017. **130**(21): p. 2624-2631.
129. Valencia, C.A., et al., *Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel.* PLoS One, 2013. **8**(1): p. e53083.
130. Gao, N. and J. Frank, *A library of RNA bridges.* Nat Chem Biol, 2006. **2**(5): p. 231-2.
131. Moore, P.B. and T.A. Steitz, *The structural basis of large ribosomal subunit function.* Annu Rev Biochem, 2003. **72**: p. 813-50.
132. Anfinsen, C.B., *The formation and stabilization of protein structure.* Biochem J, 1972. **128**(4): p. 737-49.
133. Rubio Gomez, M.A. and M. Ibba, *Aminoacyl-tRNA synthetases.* RNA, 2020. **26**(8): p. 910-936.
134. Chapeville, F., et al., *On the role of soluble ribonucleic acid in coding for amino acids.* Proc Natl Acad Sci U S A, 1962. **48**: p. 1086-92.
135. Buehner, M., et al., *D-glyceraldehyde-3-phosphate dehydrogenase: three-dimensional structure and evolutionary significance.* Proc Natl Acad Sci U S A, 1973. **70**(11): p. 3052-4.
136. Cusack, S., et al., *A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 A.* Nature, 1990. **347**(6290): p. 249-55.
137. Zhang, C.M., et al., *Distinct kinetic mechanisms of the two classes of Aminoacyl-tRNA synthetases.* J Mol Biol, 2006. **361**(2): p. 300-11.
138. Ruff, M., et al., *Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp).* Science, 1991. **252**(5013): p. 1682-9.
139. Safro, M. and L. Klipcan, *The mechanistic and evolutionary aspects of the 2'- and 3'-OH paradigm in biosynthetic machinery.* Biol Direct, 2013. **8**: p. 17.

140. Wolfenden, R., et al., *Temperature dependence of amino acid hydrophobicities.* Proc Natl Acad Sci U S A, 2015. **112**(24): p. 7484-8.
141. Tang, S.N. and J.F. Huang, *Evolution of different oligomeric glycyl-tRNA synthetases.* FEBS Lett, 2005. **579**(6): p. 1441-5.
142. Lo, W.S., et al., *Human tRNA synthetase catalytic nulls with diverse functions.* Science, 2014. **345**(6194): p. 328-32.
143. Eldred, E.W. and P.R. Schimmel, *Rapid deacylation by isoleucyl transfer ribonucleic acid synthetase of isoleucine-specific transfer ribonucleic acid aminoacylated with valine.* J Biol Chem, 1972. **247**(9): p. 2961-4.
144. Hirano, T., et al., *Mapping and exome sequencing identifies a mutation in the IARS gene as the cause of hereditary perinatal weak calf syndrome.* PLoS One, 2013. **8**(5): p. e64036.
145. Kopajtich, R., et al., *Biallelic IARS Mutations Cause Growth Retardation with Prenatal Onset, Intellectual Disability, Muscular Hypotonia, and Infantile Hepatopathy.* Am J Hum Genet, 2016. **99**(2): p. 414-22.
146. Orenstein, N., et al., *Bi-allelic IARS mutations in a child with intra-uterine growth retardation, neonatal cholestasis, and mild developmental delay.* Clin Genet, 2017. **91**(6): p. 913-917.
147. Casey, J.P., et al., *Identification of a mutation in LARS as a novel cause of infantile hepatopathy.* Mol Genet Metab, 2012. **106**(3): p. 351-8.
148. Fagbemi, A., et al., *Refractory very early-onset inflammatory bowel disease associated with cytosolic isoleucyl-tRNA synthetase deficiency: A case report.* World J Gastroenterol, 2020. **26**(15): p. 1841-1846.
149. Noguchi, E., et al., *Skeletal Muscle Involvement in Antisynthetase Syndrome.* JAMA Neurol, 2017. **74**(8): p. 992-999.
150. Fuchs, S.A., et al., *Aminoacyl-tRNA synthetase deficiencies in search of common themes.* Genet Med, 2019. **21**(2): p. 319-330.
151. Yang, W., et al., *Transcriptome analysis of the toxic mechanism of nanoplastics on growth, photosynthesis and oxidative stress of microalga Chlorella pyrenoidosa during chronic exposure.* Environ Pollut, 2021. **284**: p. 117413.
152. Sztretye, M., et al., *From Mice to Humans: An Overview of the Potentials and Limitations of Current Transgenic Mouse Models of Major Muscular Dystrophies and Congenital Myopathies.* Int J Mol Sci, 2020. **21**(23).
153. Bertini, E., et al., *Congenital muscular dystrophies: a brief review.* Semin Pediatr Neurol, 2011. **18**(4): p. 277-88.
154. National Organization for Rare Disorders, *Congenital Muscular Dystrophy* 2013 [cited 2021 27/06/2021]; Available from: https://rarediseases.org/rare-diseases/congenital-muscular-dystrophy/.
155. Martin, P.T., *Dystroglycan glycosylation and its role in matrix binding in skeletal muscle.* Glycobiology, 2003. **13**(8): p. 55R-66R.
156. Awano, H., et al., *Restoration of Functional Glycosylation of alpha-Dystroglycan in FKRP Mutant Mice Is Associated with Muscle Regeneration.* Am J Pathol, 2015. **185**(7): p. 2025-37.
157. Lapidos, K.A., R. Kakkar, and E.M. McNally, *The dystrophin glycoprotein complex: signaling strength and integrity for the sarcolemma.* Circ Res, 2004. **94**(8): p. 1023-31.
158. Hernan D. Gonorazky, C.G.B., James J. Dowling, *Handbook of Clinical Neurology*. Chapter 36 - The genetics of congenital myopathies. Vol. 148. 2018: Elsevier.
159. Moghadaszadeh, B., et al., *Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome.* Nat Genet, 2001. **29**(1): p. 17-8.
160. Vasli, N., et al., *Recessive mutations in the kinase ZAK cause a congenital myopathy with fibre type disproportion.* Brain, 2017. **140**(1): p. 37-48.

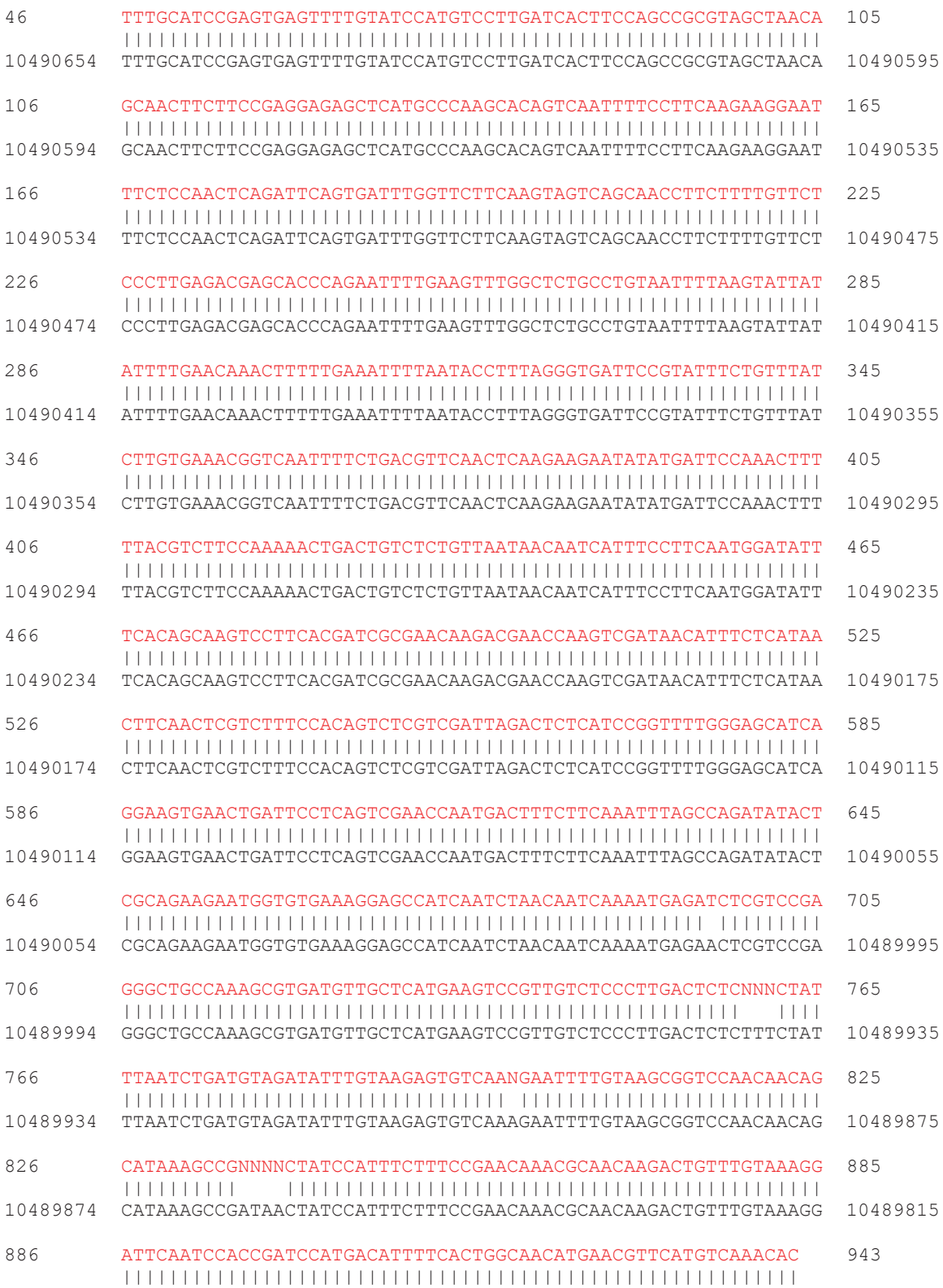
161. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
162. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
163. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.
164. (ESP), N.E.S.P., *Exome Variant Server (ESP)*. November 7, 2011.
165. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
166. Kamath, R.S., et al., *Systematic functional analysis of the Caenorhabditis elegans genome using RNAi.* Nature, 2003. **421**(6920): p. 231-7.
167. Sonnichsen, B., et al., *Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans.* Nature, 2005. **434**(7032): p. 462-9.
168. Lai, C.H., et al., *Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics.* Genome Res, 2000. **10**(5): p. 703-13.
169. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
170. Meissner, B., et al., *An integrated strategy to study muscle development and myofilament structure in Caenorhabditis elegans.* PLoS Genet, 2009. **5**(6): p. e1000537.
171. Small, T.M., et al., *Three new isoforms of Caenorhabditis elegans UNC-89 containing MLCK-like protein kinase domains.* J Mol Biol, 2004. **342**(1): p. 91-108.
172. Spooner, P.M., et al., *Large isoforms of UNC-89 (obscurin) are required for muscle cell architecture and optimal calcium release in Caenorhabditis elegans.* PLoS One, 2012. **7**(7): p. e40182.
173. Glenn, C.F., et al., *Behavioral deficits during early stages of aging in Caenorhabditis elegans result from locomotory deficits possibly linked to muscle frailty.* J Gerontol A Biol Sci Med Sci, 2004. **59**(12): p. 1251-60.
174. Cornwall, K.M., et al., *A Qualitative Approach to Health Related Quality-of-Life in Congenital Muscular Dystrophy.* J Neuromuscul Dis, 2018. **5**(2): p. 251-255.
175. Farrell, P.M., et al., *Evidence on improved outcomes with early diagnosis of cystic fibrosis through neonatal screening: enough is enough!* J Pediatr, 2005. **147**(3 Suppl): p. S30-6.
176. Elder, J.H., et al., *Clinical impact of early diagnosis of autism on the prognosis and parent-child relationships.* Psychol Res Behav Manag, 2017. **10**: p. 283-292.
177. Cassim, S., et al., *Patient and carer perceived barriers to early presentation and diagnosis of lung cancer: a systematic review.* BMC Cancer, 2019. **19**(1): p. 25.
178. Merelle, M.E., et al., *Early versus late diagnosis: psychological impact on parents of children with cystic fibrosis.* Pediatrics, 2003. **111**(2): p. 346-50.
179. Waterston, R.H., J.N. Thomson, and S. Brenner, *Mutants with altered muscle structure of Caenorhabditis elegans.* Dev Biol, 1980. **77**(2): p. 271-302.
180. Ferrara, T.M., D.B. Flaherty, and G.M. Benian, *Titin/connectin-related proteins in C. elegans: a review and new findings.* J Muscle Res Cell Motil, 2005. **26**(6-8): p. 435-47.
181. Randazzo, D., et al., *The potential of obscurin as a therapeutic target in muscle disorders.* Expert Opin Ther Targets, 2017. **21**(9): p. 897-910.
182. Agrawal, P.B., et al., *SPEG interacts with myotubularin, and its deficiency causes centronuclear myopathy with dilated cardiomyopathy.* Am J Hum Genet, 2014. **95**(2): p. 218-26.
183. Rossi, D., et al., *A novel FLNC frameshift and an OBSCN variant in a family with distal muscular dystrophy.* PLoS One, 2017. **12**(10): p. e0186642.
184. Watts, J.S., et al., *New Strains for Tissue-Specific RNAi Studies in Caenorhabditis elegans.* G3 (Bethesda), 2020. **10**(11): p. 4167-4176.

185. Tabara, H., et al., *The rde-1 gene, RNA interference, and transposon silencing in C. elegans.* Cell, 1999. **99**(2): p. 123-32.
186. Grishok, A., J.L. Sinskey, and P.A. Sharp, *Transcriptional silencing of a transgene by RNAi in the soma of C. elegans.* Genes Dev, 2005. **19**(6): p. 683-96.
187. De-Souza, E.A., et al., *RNA interference may result in unexpected phenotypes in Caenorhabditis elegans.* Nucleic Acids Res, 2019. **47**(8): p. 3957-3969.
188. Asad, N., Yih Aw W., and Timmons L. Natural and unanticipated modifiers of RNAi activity in Caenorhabditis elegans. PLoS One, 2012. 7(11): e50191
189. Zhang, Q., et al., *Genome-wide open chromatin regions and their effects on the regulation of silk protein genes in Bombyx mori.* Sci Rep, 2017. **7**(1): p. 12919.
190. Ibanez-Ventoso, C., et al. Automated Analysis of C. elegans Swim Behaviour Using CeleST Software. J VIs Exp, 2016. 7(118): 54359
191. Kai Zhang, J.D.H., Michael Miller, Xiaomeng Hou, Joshua Chiou, Olivier B. Poirion, Yunjiang Qiu, Yang E. Li, Kyle J. Gaulton, Allen Wang, Sebastian Preissl, Bing Ren., *A cell atlas of chromatin accessibility across 25 adult human tissues.* bioRxiv.
192. Adli, M., *The CRISPR tool kit for genome editing and beyond.* Nat Commun, 2018. **9**(1): p. 1911.
193. Campenhout, C.V., et al., *Guidelines for optimized gene knockout using CRISPR/Cas9.* Biotechniques, 2019. **66**(6): p. 295-302.
194. Cieslar-Pobuda, A., et al., *Transdifferentiation and reprogramming: Overview of the processes, their similarities and differences.* Biochim Biophys Acta Mol Cell Res, 2017. **1864**(7): p. 1359-1369.
195. Boularaoui, S.M., et al., *Efficient transdifferentiation of human dermal fibroblasts into skeletal muscle.* J Tissue Eng Regen Med, 2018. **12**(2): p. e918-e936.
196. Kabadi, A.M., et al., *Enhanced MyoD-induced transdifferentiation to a myogenic lineage by fusion to a potent transactivation domain.* ACS Synth Biol, 2015. **4**(6): p. 689-99.
197. Almeida, C.F., et al., *Direct Reprogramming of Human Fibroblasts into Myoblasts to Investigate Therapies for Neuromuscular Disorders.* J Vis Exp, 2021(170).
198. Calderon, J.C., P. Bolanos, and C. Caputo, *The excitation-contraction coupling mechanism in skeletal muscle.* Biophys Rev, 2014. **6**(1): p. 133-160.
199. Bezprozvanny, I.B., et al., *Activation of the calcium release channel (ryanodine receptor) by heparin and other polyanions is calcium dependent.* Mol Biol Cell, 1993. **4**(3): p. 347-52.
200. Primeau, J.O., et al., *The SarcoEndoplasmic Reticulum Calcium ATPase.* Subcell Biochem, 2018. **87**: p. 229-258.
201. Amburgey, K., et al., *Prevalence of congenital myopathies in a representative pediatric united states population.* Ann Neurol, 2011. **70**(4): p. 662-5.
202. Robinson, R., et al., *Mutations in RYR1 in malignant hyperthermia and central core disease.* Hum Mutat, 2006. **27**(10): p. 977-89.
203. Du, G.G., et al., *Topology of the Ca2+ release channel of skeletal muscle sarcoplasmic reticulum (RyR1).* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16725-30.
204. Yan, Z., et al., *Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution.* Nature, 2015. **517**(7532): p. 50-55.
205. Futatsugi, A., G. Kuwajima, and K. Mikoshiba, *Tissue-specific and developmentally regulated alternative splicing in mouse skeletal muscle ryanodine receptor mRNA.* Biochem J, 1995. **305 ( Pt 2)**: p. 373-8.
206. Kimura, T., et al., *Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca2+-ATPase in myotonic dystrophy type 1.* Hum Mol Genet, 2005. **14**(15): p. 2189-200.

207. Litman, R.S., et al., *Malignant Hyperthermia Susceptibility and Related Diseases.* Anesthesiology, 2018. **128**(1): p. 159-167.
208. National Organization for Rare Disorders. *RYR1-Related Diseases*. 2021 [cited 2021 7/7/2021]; Available from: https://rarediseases.org/rare-diseases/ryr-1-related-diseases/.
209. Rosenberg, H., et al., *Malignant hyperthermia.* Orphanet J Rare Dis, 2007. **2**: p. 21.
210. Nelson, T.E., *Porcine malignant hyperthermia: critical temperatures for in vivo and in vitro responses.* Anesthesiology, 1990. **73**(3): p. 449-54.
211. Michelucci, A., et al., *Strenuous exercise triggers a life-threatening response in mice susceptible to malignant hyperthermia.* FASEB J, 2017. **31**(8): p. 3649-3662.
212. Gronert, G.A., *Dantrolene in malignant hyperthermia (MH)-susceptible patients with exaggerated exercise stress.* Anesthesiology, 2000. **93**(3): p. 905.
213. Zhou, H., et al., *Molecular mechanisms and phenotypic variation in RYR1-related congenital myopathies.* Brain, 2007. **130**(Pt 8): p. 2024-36.
214. Engel, W.K., *The essentiality of histo- and cytochemical studies of skeletal muscle in the investigation of neuromuscular disease. 1962.* Neurology, 1998. **51**(3): p. 655 and 17 pages following.
215. Clarke, N.F. and K.N. North, *Congenital fiber type disproportion--30 years on.* J Neuropathol Exp Neurol, 2003. **62**(10): p. 977-89.
216. Clarke, N.F., *Congenital fiber-type disproportion.* Semin Pediatr Neurol, 2011. **18**(4): p. 264-71.
217. Kuhn, J.H., et al., *New filovirus disease classification and nomenclature.* Nat Rev Microbiol, 2019. **17**(5): p. 261-263.
218. Boycott, K.M., et al., *International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases.* Am J Hum Genet, 2017. **100**(5): p. 695-705.
219. King, W. and J. Kissel, *Multidisciplinary approach to the management of myopathies.* Continuum (Minneap Minn), 2013. **19**(6 Muscle Disease): p. 1650-73.
220. Graham, B., M.A. Shaw, and I.A. Hope, *Single Amino Acid Changes in the Ryanodine Receptor in the Human Population Have Effects In Vivo on Caenorhabditis elegans Neuro-Muscular Function.* Front Genet, 2020. **11**: p. 37.
221. Morck, C. and M. Pilon, *C. elegans feeding defective mutants have shorter body lengths and increased autophagy.* BMC Dev Biol, 2006. **6**: p. 39.
222. Bandyopadhyay, J., et al., *Calcineurin, a calcium/calmodulin-dependent protein phosphatase, is involved in movement, fertility, egg laying, and growth in Caenorhabditis elegans.* Mol Biol Cell, 2002. **13**(9): p. 3281-93.
223. Gonorazky, H.D., et al., *Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease.* Am J Hum Genet, 2019. **104**(3): p. 466-483.
224. Hirsch, C.D., N.M. Springer, and C.N. Hirsch, *Genomic limitations to RNA sequencing expression profiling.* Plant J, 2015. **84**(3): p. 491-503.
225. Spinozzi, G., et al., *ARPIR: automatic RNA-Seq pipelines with interactive report.* BMC Bioinformatics, 2020. **21**(Suppl 19): p. 574.
226. Wolfien, M., et al., *TRAPLINE: a standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation.* BMC Bioinformatics, 2016. **17**: p. 21.
227. Tissenbaum, H.A., *Using C. elegans for aging research.* Invertebr Reprod Dev, 2015. **59**(sup1): p. 59-63.
228. Jirka, C., et al., *Dysregulation of NRAP degradation by KLHL41 contributes to pathophysiology in nemaline myopathy.* Hum Mol Genet, 2019. **28**(15): p. 2549-2560.
229. D'Avila, F., et al., *Exome sequencing identifies variants in two genes encoding the LIM-proteins NRAP and FHL1 in an Italian patient with BAG3 myofibrillar myopathy.* J Muscle Res Cell Motil, 2016. **37**(3): p. 101-15.

230. Braungart, E., et al., *Caenorhabditis elegans MPP+ model of Parkinson's disease for high-throughput drug screenings.* Neurodegener Dis, 2004. **1**(4-5): p. 175-83.
231. EmelyneTeo, S.Y.J.L., Sheng Fong, Anis Larbi, Graham D.Wright, NicholasTolwinski, Jan Gruber, *A high throughput drug screening paradigm using transgenic Caenorhabditis elegans model of Alzheimer's disease.* Translational Medicine of Aging, 2020. **4**: p. 11-21.
232. Volpatti, J.R., et al., *Identification of drug modifiers for RYR1-related myopathy using a multi-species discovery pipeline.* Elife, 2020.

# Appendix A: Supplementary Figures



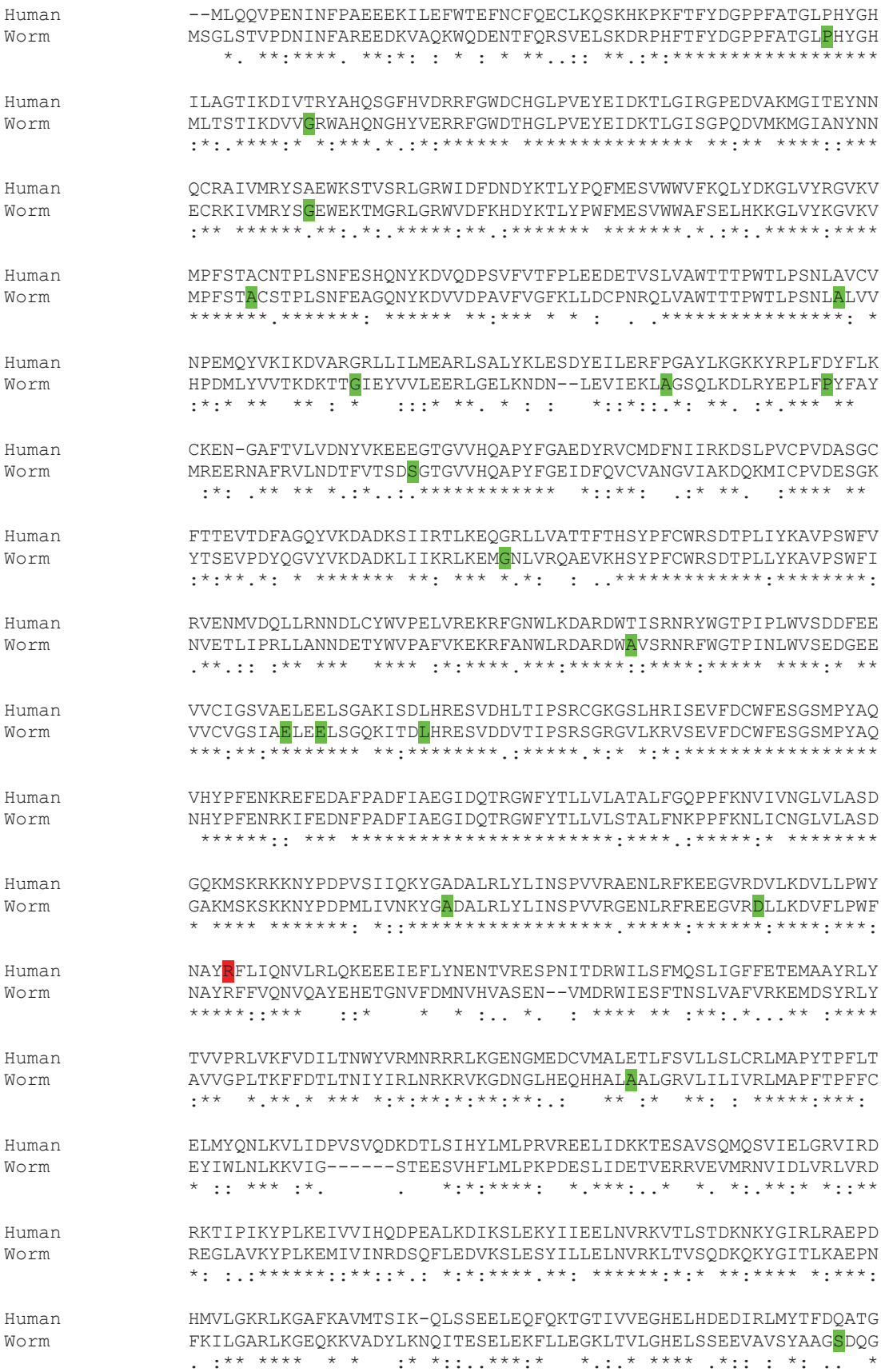
```
10489814   ATTCAATCCACCGATCCATGACATTTTCACTGGCAACATGAACGTTCATGTCAAACAC    10489757
```

**Blastn report**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 1624 bits (879) | 0.0 | 889/898 (99%) | 0/898 (0%) | Plus/Minus |

**Supplementary Figure 1. *iars-1* RNAi shares 99 percent identity with the *C. elegans iars-1* gene**

Comparing Sanger sequencing results of *iars-1* RNAi vector (red) to the *C. elegans* reference genome results in a 99% match with the *iars-1* gene (black).


Human --MLQQVPENINFPAEEEKILEFWTEFNCFQECLKQSKHKPKFTFYDGPPFATGLPHYGH
Worm MSGLSTVPDNINFAREEDKVAQKWQDENTFQRSVELSKDRPHFTFYDGPPFATGLPHYGH
Human ILAGTIKDIVTRYAHQSGFHVDRRFGWDCHGLPVEYEIDKTLGIRGPEDVAKMGITEYNN
Worm MLTSTIKDVVGRWAHQNGHYVERRFGWDTHGLPVEYEIDKTLGISGPQDVMKMGIANYNN
Human QCRAIVMRYSAEWKSTVSRLGRWIDFDNDYKTLYPQFMESVWWVFKQLYDKGLVYRGVKV
Worm ECRKIVMRYSGEWEKTMGRLGRWVDFKHDYKTLYPWFMESVWWAFSELHKKGLVYKGVKV
Human MPFSTACNTPLSNFESHQNYKDVQDPSVFVTFPLEEDETVSLVAWTTTPWTLPSNLAVCV
Worm MPFSTACSTPLSNFEAGQNYKDVVDPAVFVGFKLLDCPNRQLVAWTTTPWTLPSNLALVV
Human NPEMQYVKIKDVARGRLLILMEARLSALYKLESDYEILERFPGAYLKGKKYRPLFDYFLK
Worm HPDMLYVVTKDKTTGIEYVVLEERLGELKNDN--LEVIEKLAGSQLKDLRYEPLFPYFAY
Human CKEN-GAFTVLVDNYVKEEEGTGVVHQAPYFGAEDYRVCMDFNIIRKDSLPVCPVDASGC
Worm MREERNAFRVLNDTFVTSDSGTGVVHQAPYFGEIDFQVCVANGVIAKDQKMICPVDESGK
Human FTTEVTDFAGQYVKDADKSIIRTLKEQGRLLVATTFTHSYPFCWRSDTPLIYKAVPSWFV
Worm YTSEVPDYQGVYVKDADKLIIKRLKEMGNLVRQAEVKHSYPFCWRSDTPLLYKAVPSWFI
Human RVENMVDQLLRNNDLCYWVPELVREKRFGNWLKDARDWTISRNRYWGTPIPLWVSDDFEE
Worm NVETLIPRLLANNDETYWVPAFVKEKRFANWLRDARDWAVSRNRFWGTPINLWVSEDGEE
Human VVCIGSVAELEELSGAKISDLHRESVDHLTIPSRCGKGSLHRISEVFDCWFESGSMPYAQ
Worm VVCVGSIAFLEFLSGQKITDLHRESVDDVTIPSRSGRGVLKRVSEVFDCWFESGSMPYAQ
Human VHYPFENKREFEDAFPADFIAEGIDQTRGWFYTLLVLATALFGQPPFKNVIVNGLVLASD
Worm NHYPFENRKIFEDNFPADFIAEGIDQTRGWFYTLLVLSTALFNKPPFKNLICNGLVLASD
Human GQKMSKRKKNYPDPVSIIQKYGADALRLYLINSPVVRAENLRFKEEGVRDVLKDVLLPWY
Worm GAKMSKSKKNYPDPMLIVNKYGADALRLYLINSPVVRGENLRFREEGVRDLLKDVFLPWF
Human NAYRFLIQNVLRLQKEEEIEFLYNENTVRESPNITDRWILSFMQSLIGFFETEMAAYRLY
Worm NAYRFFVQNVQAYEHETGNVFDMNVHVASEN--VMDRWIESFTNSLVAFVRKEMDSYRLY
Human TVVPRLVKFVDILTNWYVRMNRRRLKGENGMEDCVMALETLFSVLLSLCRLMAPYTPFLT
Worm AVVGPLTKFFDTLTNIYIRLNRKRVKGDNGLHEQHHALAALGRVLILIVRLMAPFTPFFC
Human ELMYQNLKVLIDPVSVQDKDTLSIHYLMLPRVREELIDKKTESAVSQMQSVIELGRVIRD
Worm EYIWLNLKKVIG------STEESVHFLMLPKPDESLIDETVERRVEVMRNVIDLVRLVRD
Human RKTIPIKYPLKEIVVIHQDPEALKDIKSLEKYIIEELNVRKVTLSTDKNKYGIRLRAEPD
Worm REGLAVKYPLKEMIVINRDSQFLEDVKSLESYILLELNVRKLTVSQDKQKYGITLKAEPN
Human HMVLGKRLKGAFKAVMTSIK-QLSSEELEQFQKTGTIVVEGHELHDEDIRLMYTFDQATG
Worm FKILGARLKGEQKKVADYLKNQITESELEKFLLEGKLTVLGHELSSEEVAVSYAAGSDQG


**Supplementary Figure 2. Complete protein alignment of human IARS and *C. elegans* iars-1**
Protein alignment of human IARS (NP_001365503.1) and *C. elegans* iars-1 (NP_001366774.1). The position of the human mutation p.R661H is highlighted in red. The position of all 26 known worm mutations in iars-1 recorded by the Million Mutation Project are highlighted in green.

> PHX1948 *unc-68*(*syb1948*)

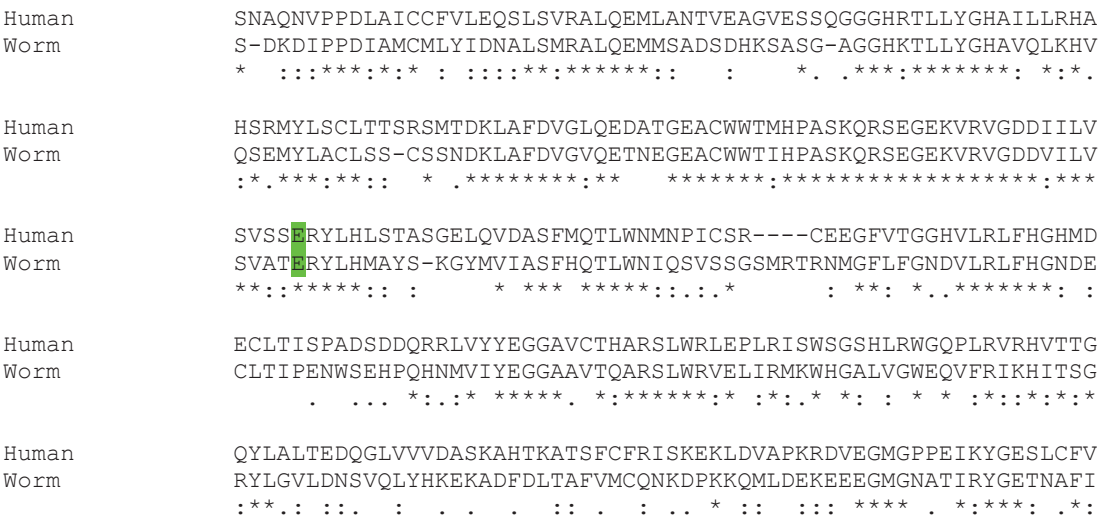
CGATAACGCACTGTCGATGAGAGCATTACAGGAAATGATGTCAGCGGATAGTGATCATgtgagttttttttttaattaaaaaa
taatttttcgcatataatggaaagatacagtttttgttaatactgaaacaatttcaaattttagAAGTCTGCATCCGGCGCTGGTGGTCACA
AAACACTGTTATACGGACACGCGGTACAGCTGAAGCATGTTCAGAGTGAAATGTACCTCGCTTGTCTATCATCATG
CTCTTCAAATGACAAACTTGCTTTTGATGTAGGAGTTCAAGAAACTAATGAAGgttggtacaacttttaaggcaaaagttttga
aaaataatttttagGAGAAGCATGTTGGTGGACAATTCATCCAGCATCCAAACAAAGATCAGAAGGAGAAAAGGTGCG
CGTCGGTGATGACGTCATT**TTA**GTCTCAGTTGCAACA**AAGCGC**TATTTGgtaagtatttttttattattaaaagagtttgtttgaat
gtaaaatttttttttaatttagcacttaaaaaagggtaaaaccgaagaagaggcaagtttaatgttgttgttttgtggtagaaatgtattattattatttca
tcaagttcatttgaactctaaataaaaccgtaatcttgaaatttttcagCACATGGCCTACAGTAAAGGCTACATGGTGATTGCCTCAT
TTCACCAAACACTATGGAATATCCAGTCAGTCAGTTCTGGAAGCATGAGAACAAGGAATATGGgttagttaacactgac
aatgagatttttttaaactttgaaactttagGTTTCCTCTTCGGAAATGACGTACTGAGGTTGTTCCACGGAAATGATGAGTGC
TTGACAATTCCAGAAAACTGGAGTGAGCATCCACAACACAAgtgagtaagaatatcaatttctgaattctgaaattctgttaattag
gaacttgtaatttggcaaaatattgtaatctaacgttccaattttgcagCATGGTAATCTACGAAGGAGGCGCTGCAGTGACACAAG
CCAGGTCACTATGGCGAGTAGAATTGATTCGAATGAAATGGCATGGTGCACTTGTCGGATGGGAGCAAGTGTTTC
G

>N2 wild type

CGATAACGCACTGTCGATGAGAGCATTACAGGAAATGATGTCAGCGGATAGTGATCATgtgagttttttttttaattaaaaaa
taatttttcgcatataatggaaagatacagtttttgttaatactgaaacaatttcaaattttagAAGTCTGCATCCGGCGCTGGTGGTCACA
AAACACTGTTATACGGACACGCGGTACAGCTGAAGCATGTTCAGAGTGAAATGTACCTCGCTTGTCTATCATCATG
CTCTTCAAATGACAAACTTGCTTTTGATGTAGGAGTTCAAGAAACTAATGAAGgttggtacaacttttaaggcaaaagttttga
aaaataatttttagGAGAAGCATGTTGGTGGACAATTCATCCAGCATCCAAACAAAGATCAGAAGGAGAAAAGGTGCG
CGTCGGTGATGACGTCATT**TTG**GTCTCAGTTGCAACA**GAAAGA**TATTTGgtaagtatttttttattattaaaagagtttgtttgaat
gtaaaatttttttttaatttagcacttaaaaaagggtaaaaccgaagaagaggcaagtttaatgttgttgttttgtggtagaaatgtattattattatttca
tcaagttcatttgaactctaaataaaaccgtaatcttgaaatttttcagCACATGGCCTACAGTAAAGGCTACATGGTGATTGCCTCAT
TTCACCAAACACTATGGAATATCCAGTCAGTCAGTTCTGGAAGCATGAGAACAAGGAATATGGgttagttaacactgac
aatgagatttttttaaactttgaaactttagGTTTCCTCTTCGGAAATGACGTACTGAGGTTGTTCCACGGAAATGATGAGTGC
TTGACAATTCCAGAAAACTGGAGTGAGCATCCACAACACAAgtgagtaagaatatcaatttctgaattctgaaattctgttaattag
gaacttgtaatttggcaaaatattgtaatctaacgttccaattttgcagCATGGTAATCTACGAAGGAGGCGCTGCAGTGACACAAG
CCAGGTCACTATGGCGAGTAGAATTGATTCGAATGAAATGGCATGGTGCACTTGTCGGATGGGAGCAAGTGTTTC
G

**Supplementary Figure 3. PHX1948 Mutagenesis**

Sanger sequencing of PHX1948 *unc-68*(*syb1948*) mutants and N2 wt worms as reported by SunyBiotech. Base pairs labeled in blue and red represent synonymous and non-synonymous mutations, respectively, introduced during CRISPR/Cas9 gene editing. Codon triplets affected by a mutation are underlined and the variant under scrutiny (c.526G>A) is highlighted in yellow. Forward and reverse primers are as follows: CGATAACGCACTGTCGATGA and CGAAACACTTGCTCCCATCC.




**Supplementary Figure 4. Protein alignment of a section of human RYR-1 and *C. elegans* UNC-68**

Protein alignment of human RYR1 (NP_000531.2) and *C. elegans* UNC-68 (NP_001256074.1). The position of the human mutation p.E176K and the *C.* elegans mutation p.E182K are highlighted in green.

> PHX2444 *unc-68*(*syb1948*)

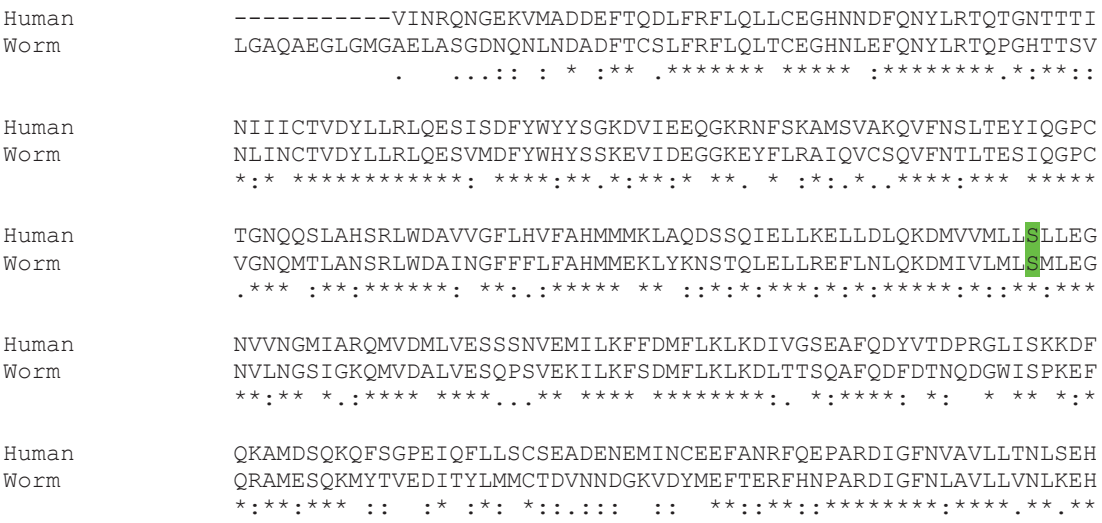
CGGTTATGGACTTCTACTGGCATTATTCTAGCAAAGAAGTAATTGATGAAGGTGGAAAAGAATACTTCCTTCGTGC
AATTCAAGTGTGTAGTCAGGTGTTCAATACACTCACTGAGTCGATTCAGgtaataatatctaaattaggaagtcaactattgaat
tttaaaatttaagGGTCCATGCGTTGGTAACCAGATGACTTTGGCCAACTCTCGCCTTTGGGATGCCATCAACGGTTTCT
TCTTTCTTTTTGCTCACATGATGGAGAAACTTTACAAGAATTCTACCCAACTTGAACTTTTACGAGAGTTCCTTAACT
TACAAAAAGACATGATTGTTCTCATGTTGCTGATGCTTGAAGGGAATGTTTTGAACGGTTCAATTGGTAAGCAGAT
GGTAGACGCTCTCGTCGAATCCCAACCATCCGTGGAGAAGATTTTGAAATTCTCGGATATGTTCTTGAAGTTGAAG
GATTTGACCACATCTCAAGCCTTCCAAGACTTTGATACAAATCAAGACGGATGGATTAGTCCTAAAGAATTCCAAC
GTGCAATGGAAAGTCAGAAGATGTACACTGTAGAAGATATTACATATCTGATGATGTGTACTGATGTTAACAACG
ATGGAAAAGTTGACTACATGGAATTCACTGAACGATTCCACAATCCTGCGAGAGACATTGGTTTTAACTTGGCAGT
GCTGCTTGTCAACCTGAAGGAGCACATCACAAATGATCCAAGACTCGAAAAGATCATAGAGAAGGCTCAAACACT
TCTGGAATACTTTGACCCATTCCTTGGAAGAATCGAGATTATGGGATCATCAAAGAGAGTTGAGAAAATTTACTTT
GAAATTCAAGAATCGTGGCTGGAACAATGGGGAAAACAACAGATTAGGGATTCGAAGAACTCGTTCCTTTTCAAT
GTTTTGCAAGACGATGGCGGTGACCAAGGAAAACTTGAAGCTTTCATCAATTTCTGTGAAGACACTATTTTTGAAA
TGCAGCATGCCGCGGCGATTTCTTCTGGGGATAGTGACACCAAGATGG

>N2 wild type

CGGTTATGGACTTCTACTGGCATTATTCTAGCAAAGAAGTAATTGATGAAGGTGGAAAAGAATACTTCCTTCGTGC
AATTCAAGTGTGTAGTCAGGTGTTCAATACACTCACTGAGTCGATTCAGgtaataatatctaaattaggaagtcaactattgaat
tttaaaatttaagGGTCCATGCGTTGGTAACCAGATGACTTTGGCCAACTCTCGCCTTTGGGATGCCATCAACGGTTTCT
TCTTTCTTTTTGCTCACATGATGGAGAAACTTTACAAGAATTCTACCCAACTTGAACTTCTCCGAGAGTTCCTTAACT
TACAAAAAGACATGATTGTTCTCATGCTTTCCATGCTTGAAGGGAATGTTTTGAACGGTTCAATTGGTAAGCAGAT
GGTGGACGCTCTCGTGGAATCCCAACCATCCGTGGAGAAGATTTTGAAATTCTCGGATATGTTCTTGAAGTTGAAG
GATTTGACCACATCTCAAGCCTTCCAAGACTTTGATACAAATCAAGACGGATGGATTAGTCCTAAAGAATTCCAAC
GTGCAATGGAAAGTCAGAAGATGTACACTGTAGAAGATATTACATATCTGATGATGTGTACTGATGTTAACAACG
ATGGAAAAGTTGACTACATGGAATTCACTGAACGATTCCACAATCCTGCGAGAGACATTGGTTTTAACTTGGCAGT
GCTGCTTGTCAACCTGAAGGAGCACATCACAAATGATCCAAGACTCGAAAAGATCATAGAGAAGGCTCAAACACT
TCTGGAATACTTTGACCCATTCCTTGGAAGAATCGAGATTATGGGATCATCAAAGAGAGTTGAGAAAATTTACTTT
GAAATTCAAGAATCGTGGCTGGAACAATGGGGAAAACAACAGATTAGGGATTCGAAGAACTCGTTCCTTTTCAAT
GTTTTGCAAGACGATGGCGGTGACCAAGGAAAACTTGAAGCTTTCATCAATTTCTGTGAAGACACTATTTTTGAAA
TGCAGCATGCCGCGGCGATTTCTTCTGGGGATAGTGACACCAAGATGG

**Supplementary Figure 5. PHX2444 Mutagenesis**

Sanger sequencing of PHX2444 *unc-68*(*syb1948*) mutants containing the c.12083C>T knock-in and N2 wt worms as reported by SunyBiotech. Base pairs labeled in blue and red represent synonymous and non-synonymous mutations, respectively, introduced during CRISPR/Cas9 gene editing. Codon triplets affected by a mutation are underlined and the variant under scrutiny (c.12083C>T) is highlighted in yellow. Forward and reverse primers are as follows: CGGTTATGGACTTCTACTGG and CCATCTTGGTGTCACTATCC.




**Supplementary Figure 6. Protein alignment of a section of human RYR-1 and *C. elegans* UNC-68**

Protein alignment of human RYR1 (NP_000531.2) and *C. elegans* UNC-68 (NP_001256074.1). The position of the human mutation p.S4028L and the *C.* elegans mutation p.S4273L are highlighted in green.

# Appendix B: Supplementary Tables

**Supplementary Table 1. Parameters used while tracking worms with WormLab software for crawling and swimming assays**

| | Parameter | Min | Max |
|---|---|---|---|
| **Crawling** | Area | 125.78 | 2775.43 |
| | Length | 55.44 | 119.70 |
| | Width | 3.12 | 7.26 |
| | Width/length ratio | 0.03 | 0.09 |
| | Detection fit | 0.45 | |
| | Registration fit | 0.25 | |
| **Swimming** | Area | 93.19 | 1732.47 |
| | Length | 47.21 | 95.49 |
| | Width | 2.72 | 5.87 |
| | Width/length ratio | 0.03 | 0.10 |
| | Detection fit | 0.47 | |
| | Registration fit | 0.25 | |