Using long-read sequencing to diagnose rare neuromuscular disorders: a pilot study

by

Alexandria Brooke Martin

A thesis submitted to the Department of Biological Sciences, University of Manitoba,

in partial fulfilment of the requirements for the course

BIOL 4100 (Honours Thesis)

for the degree of

Bachelor of Science (Honours)

©March, 2025

ABSTRACT

Many individuals with suspected genetic neuromuscular disorders (NMDs) remain undiagnosed despite extensive conventional genetic testing, which generally employs short-read sequencing (SRS) methods. Long-read sequencing (LRS) is a recently introduced technology which offers improved genome coverage and variant detection missed in standard approaches. Therefore, this pilot study aimed to investigate the application of long-read whole genome sequencing (LR-WGS) with the PacBio Revio platform for five participants (four families) with undiagnosed NMDs to identify causal genetic variants. Retrospective participant chart reviews were first conducted to assess phenotype, inheritance patterns, and other clinical factors. Following this, a novel bioinformatics pipeline was developed to enhance variant annotation and prioritization by integrating LRS population data with established genomic population frequency databases and *in-silico* predictions of deleteriousness. Finally, manual filtration was performed to further refine variant prioritization based on the clinical insights gained from chart reviews. Candidate variants of interest were found in three of four families analyzed (75%), including strong candidates in the *FLNC* and *DMD* genes, and a potential candidate in the *TTN* gene requiring functional validation. This pilot validates the pipeline for analysis of LRS data and provides evidence that LR-WGS could offer improvements over traditional SRS approaches in diagnosing NMDs, particularly in detecting complex variants. Improving accurate molecular diagnosis ultimately leads to better clinical outcomes for affected patients, including personalized treatment strategies and interventions. Future directions include the study of additional participants with NMDs to allow for a more comprehensive evaluation of diagnostic yield and applicability across diverse genetic backgrounds.

ACKNOWLEDGEMENTS

A sincere thank you to my thesis advisor, Jessica Hartley, for her continued guidance and mentorship throughout this project. Her expertise, patience, and feedback were invaluable in shaping the quality of my work, and words cannot express my gratitude for her efforts.

Thank you to my thesis committee, Dr. Galen Wright, Dr. Mark Belmonte, and Dr. Jae-Hyeok Lee, for your individual contributions and insights. It was a privilege to work with each of you.

I would also like to extend my appreciation to the research team, including Dr. Patrick Frosk, Dr. Britt Drögemöller, Rissa Hamlin, Alana Lamont, John Pham, Sophia Perron, and Alejandra Llanes-Cuesta for their collaboration.

Thank you to the study participants and their families for their involvement in this research. Their trust and openness to sharing their experiences has made this work possible, and I am deeply grateful.

Thank you to my friends and family for their support and willingness to listen to me talk about my research and everything that went into it. I wouldn't be where I am without you. Finally, thank you to my dog, Charlie, who was with me through many long nights and difficult days of this journey. Though you are no longer here, your memory and unconditional love has stayed beside me. This one's for you, buddy.

TABLE OF CONTENTS

ABSTRACT……..……..**i**
ACKNOWLEDGEMENTS……...**ii**
TABLE OF CONTENTS……..**iii**
LIST OF TABLES……..**v**
LIST OF FIGURES……...**vi**
LIST OF ABBREVIATIONS……..**vii**
1. INTRODUCTION……..**1**
1.1. Research objectives and hypothesis……..**9**
2. MATERIALS & METHODS……..**10**
2.1. Ethics approval and funding……**10**
2.2. Participant recruitment and sample collection……**10**
2.3. Retrospective chart review and phenotype extraction……**12**
2.4. Long-read whole-genome sequencing data transfer and bioinformatics pipeline……...**13**
2.5. Genomic variant filtration and analysis……...**15**
2.6. Manual genomic variant prioritization and validation……**21**
3. RESULTS……..**23**
3.1. Case 1……..**27**
3.1.1. Case Description……...**27**
3.1.2. Investigations……**30**
3.1.3. Variant Filtration……..**31**
3.1.4. *TTN*……...**37**
3.2. Case 2……..**41**
3.2.1. Case Description……..**41**
3.2.2. Investigations……...**44**
3.2.3. Variant Filtration……**44**
3.3. Case 3……..**47**
3.3.1. Case Description……...**47**
3.3.2. Investigations……**49**
3.3.3. Variant Filtration……..**51**
3.3.4. *FLNC*……**54**
3.4. & 3.5. Case 4 and Case 5……**56**
3.4. Case 4……..**58**
3.4.1. Case Description……...**58**
3.4.2. Investigations……**60**
3.5. Case 5……..**61**
3.5.1. Case Description……...**61**
3.5.2. Investigations……**62**

3.5.3. Variant Filtration……..**63**
3.5.4. *DMD*……**68**

4. DISCUSSION……**70**
4.1. Discussion of candidate variants……**71**
4.1.1. Case 1……...**71**
4.1.2. Case 3……...**73**
4.1.3. Cases 4 and 5……**75**
4.1.4. Summary……...**77**
4.2. Limitations……...**79**
4.3. Significance and conclusions……..**81**
5. LITERATURE CITED……**82**
6. APPENDICES……...**91**

LIST OF TABLES

Table 1: Categories in the GeneTable of Neuromuscular Disorders……1
Table 2: Study inclusion and exclusion criteria……11
Table 3: Small variant predicted deleteriousness classifications……18
Table 4: Number of variants identified across participants……25
Table 5: Potential inheritance patterns and HPO terms for Case 1……28
Table 6: Clinical and bioinformatic information for small variants in Case 1……...33
Table 7: Clinical and bioinformatic information for CNVs in Case 1...[REDACTED]
Table 8: Clinical and bioinformatic information for SVs in Case 1…...**[REDACTED]**
Table 9: Potential inheritance patterns and HPO terms for Case 2……43
Table 10: Clinical and bioinformatic information for small variants in Case 2……**[REDACTED]**
Table 11: Potential inheritance patterns and HPO terms for Case 3……..48
Table 12: Clinical and bioinformatic information for small variants in Case 3…….52
Table 13: Clinical and bioinformatic information for SVs in Case 3….**[REDACTED]**
Table 14: Potential inheritance patterns and HPO terms for Cases 4 and 5……..59
Table 15: Clinical and bioinformatic information for small variants in Case 4……**[REDACTED]**
Table 16: Clinical and bioinformatic information for CNVs in Case 4……...65
Table 17: Clinical and bioinformatic information for SVs in Case 4….**[REDACTED]**
Table 18: Clinical and bioinformatic information for small variants in Case 5……**[REDACTED]**
Table 19: Clinical and bioinformatic information for CNVs in Case 5……...67

LIST OF FIGURES

Figure 1: Examples of variant types……**5**
Figure 2: Genetic variant filtration pipeline……..**16**
Figure 3: ACMG evidence framework……..**[REDACTED]**
Figure 4: Pedigree for Case 1……**[REDACTED]**
Figure 5: IGV snapshot of exonic *TTN* variant in Case 1……**38**
Figure 6: Franklin by Genoox snapshot of exonic *TTN* variant in Case 1……..**39**
Figure 7: IGV snapshot of intronic *TTN* variant in Case 1……..**40**
Figure 8: Franklin by Genoox snapshot of intronic *TTN* variant in Case 1……...**41**
Figure 9: Pedigree for Case 2……**[REDACTED]**
Figure 10: Pedigree for Case 3……...**[REDACTED]**
Figure 11: IGV snapshot of *FLNC* variant in Case 3……**55**
Figure 12: Franklin by Genoox snapshot of *FLNC* variant in Case 3……**56**
Figure 13: Pedigree for Cases 4 and 5……...**[REDACTED]**
Figure 14: IGV snapshot of *DMD* variant in Cases 4 and 5……**69**
Figure 15: UCSC Genome Browser snapshot of *DMD* variant in Cases 4 and 5…...**70**
Figure 16: Franklin by Genoox snapshot of *DMD* variant in Cases 4 and 5……..**70**
Supplementary Figure 1: Standardized data collection sheet……..**91**
Supplementary Figure 2: UCSC Genome Browser snapshot of exonic *TTN* variant in Case 1……...**101**
Supplementary Figure 3: UCSC Genome Browser snapshot of intronic *TTN* variant in Case 1……...**102**
Supplementary Figure 4: UCSC Genome Browser snapshot of *FLNC* variant in Case 3……**102**

LIST OF ABBREVIATIONS

ACMG	American College of Medical Genetics and Genomics
ADHD	Attention-deficit/hyperactivity disorder
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
CK	Creatine kinase
CNS	Central nervous system
CNV(s)	Copy number variation(s)
EMG	Electromyography
GA4K	Genomic Answers for Kids
gDNA	Genomic DNA
HPO	Human Phenotype Ontology
IGV	Integrative Genomics Viewer
LRS	Long-read sequencing
LR-WGS	Long-read whole-genome sequencing
MRI	Magnetic resonance imaging
NCV	Nerve conduction velocity
NGS	Next-generation sequencing
NMD(s)	Neuromuscular disorder(s)

OCD	Obsessive-compulsive disorder
OMIM	Online Mendelian Inheritance in Man
ONT	Oxford Nanopore Technologies
PacBio	Pacific Biosciences
PI	Principal Investigator
RNA-seq	RNA sequencing
SickKids	The Hospital for Sick Children
SNP	Single nucleotide polymorphism
SRS	Short-read sequencing
STR(s)	Short tandem repeat(s)
SV(s)	Structural variant(s)
TCAG	The Centre for Applied Genomics
VCF	Variant Call Format
VUS	Variant of Uncertain Significance
WGS	Whole genome sequencing

1. INTRODUCTION

Neuromuscular disorders (NMDs) are a heterogeneous group of over 150 conditions causing chronic and progressive skeletal muscle weakness and atrophy, which can be life-limiting (Benarroch et al., 2024; Chardon et al., 2016; Zatz et al., 2016). They are often caused by genetic variants leading to dysfunction of the muscles, peripheral nerves, anterior horn cells, and/or the neuromuscular junction (Dowling et al., 2018). NMDs broadly include disorders across numerous categories, such as muscular dystrophies, myopathies, hereditary ataxias, and several others, as listed in **Table 1** (Benarroch et al., 2024). While individual NMDs are rare, their collective impact is substantial, affecting over one million people in Canada and the United States and approximately one in 1,000 individuals worldwide (Chardon et al., 2016; Deenen et al., 2015; Zatz et al., 2016).

Table 1. NMD categories as specified in the GeneTable of Neuromuscular Disorders (Benarroch et al., 2024).

Disease Group	Example(s)	Number of Genes Identified
1. Muscular dystrophies	Duchenne/Becker muscular dystrophies, Emery-Dreifuss muscular dystrophies, limb girdle muscular dystrophies	53
2. Congenital muscular dystrophies	Congenital muscular dystrophies, Bethlem myopathies, muscular dystrophy-dystroglycanopathies	38
3. Congenital myopathies	Congenital myopathies, nemaline myopathies, centronuclear myopathies	41
4. Distal myopathies	Miyoshi muscular dystrophies, distal myopathies	22
5. Other myopathies	Myofibrillar myopathies, oculopharyngodistal myopathies	45
6. Myotonic syndromes	Myotonic dystrophies, myotonias, rippling muscle diseases	6
7. Ion channel muscle	Chloride/sodium/potassium/calcium channel	8

	diseases	muscle diseases	
8.	Malignant hyperthermia	Malignant hyperthermia susceptibilities	2
9.	Metabolic myopathies	Glycogen storage diseases, glycolytic pathway diseases, disorders of lipid metabolism	29
10.	Hereditary cardiomyopathies	Hypertrophic/dilated/restrictive/arrythmogenic ventricular/other non-arrythmogenic hereditary cardiomyopathies, long/short QT syndromes, atrial fibrillation syndromes	116
11.	Congenital myasthenic syndromes	Slow-channel/fast-channel/acetylcholine receptor congenital myasthenic syndromes	35
12.	Spinal muscular atrophies motoneuron diseases	Proximal spinal muscular atrophies, distal hereditary motor neuropathies, amyotrophic lateral sclerosis	94
13.	Hereditary ataxias	Spinocerebellar ataxias, episodic ataxias	90
14.	Hereditary motor and sensory neuropathies	Charcot-Marie-Tooth neuropathies, Dejerine-Sottas syndromes, other hereditary sensory and autonomic neuropathies, other complex neuropathy syndromes	115
15.	Hereditary paraplegias	Spastic paraplegias, spastic ataxias	76
16.	Other neuromuscular disorders	Torsion/myotonic dystonias, fibrosis of extraocular muscles, distal arthrogryposis, fetal akinesia deformation sequences, progressive external opthalmoplegias, mitochondrial DNA depletion syndromes, etc.	75

Most NMDs are genetic and follow Mendelian inheritance, caused by inherited or spontaneously occurring (*de novo*) genetic variant(s) in a single gene (Darras et al., 2014; Dowling et al., 2018). There are currently over 600 genes associated with NMDs, and establishing a genetic diagnosis can improve disease management by providing access to targeted therapies and clinical trials, informing genetic counselling, and facilitating involvement in support of advocacy groups (Benarroch et al., 2024; Dratch et al., 2024). However, many affected individuals and their families undergo a prolonged and challenging journey toward obtaining an accurate diagnosis, often referred to as a

'diagnostic odyssey' (Bauskis et al., 2022). For those with NMDs, this journey can last between several years to over a decade, and usually involves numerous costly tests and procedures, along with difficulties accessing adequate support (Bauskis et al., 2022; Grigull et al., 2016). Thus, successfully identifying diagnostic genomic alterations—often through DNA sequencing methods—is essential to the molecular diagnosis and treatment of affected individuals.

Since the advent of exome sequencing in 2012, allowing for comprehensive assessment of all protein-coding regions of the genome, "next-generation" sequencing (NGS) technologies have been widely adopted into clinical diagnostics (Lee et al., 2019; VanOudenhove et al., 2023). Currently, NGS is employed in two clinical settings: (1) targeted gene panels, where specific sets of genes relevant to a particular clinical presentation are sequenced, and (2) entire exomes, where all coding regions (exons) of the genome are sequenced (i.e., where most pathogenic variants occur) (Antonarakis et al., 1995; Choi et al., 2009; Lee et al., 2019). While mostly applied in research settings, NGS approaches can also be employed in whole genome sequencing (WGS) to assay all genetic variants across coding and non-coding regions (introns, intergenic regions, regulatory elements, etc.) of an individual's entire genome (Mitsuhashi & Matsumoto, 2020). The increased sequencing coverage afforded by WGS approaches can therefore improve the likelihood of detecting a causal pathogenic variant in an affected individual, especially if such variants are located outside of coding regions or involve complex alterations that span several genes (Mitsuhashi & Matsumoto, 2020).

However, despite the development of affordable, high-throughput sequencing methods, over half of individuals presenting with an NMD remain without a molecular

diagnosis, likely influenced by the technical and interpretive challenges faced by current sequencing platforms (Boycott et al., 2019; Mantere et al., 2019; Thompson et al., 2020; Wojcik et al., 2023). Further, most NGS approaches utilize “short read” sequencing (SRS) platforms, referring to the length of the DNA fragments sequenced by the technology, which “read” fragments in small segments of 50-300 base pairs at a time (Mantere et al., 2019). These short reads struggle to reliably detect large variants, such as copy number variations (CNVs), structural variants (SVs) like deletions and duplications, and expanded short tandem repeats (STRs) or repeat expansions, due to issues such as non-uniform coverage of the genome and sequence alignments (see **Figure 1**; Marwaha et al., 2022; VanOudenhove et al., 2023; Wojcik et al., 2023). However, as NMDs can be caused by these variant types, a more effective approach is required to bridge this diagnostic gap and identify causal variants and genes refractory to detection in genetically undiagnosed cases.

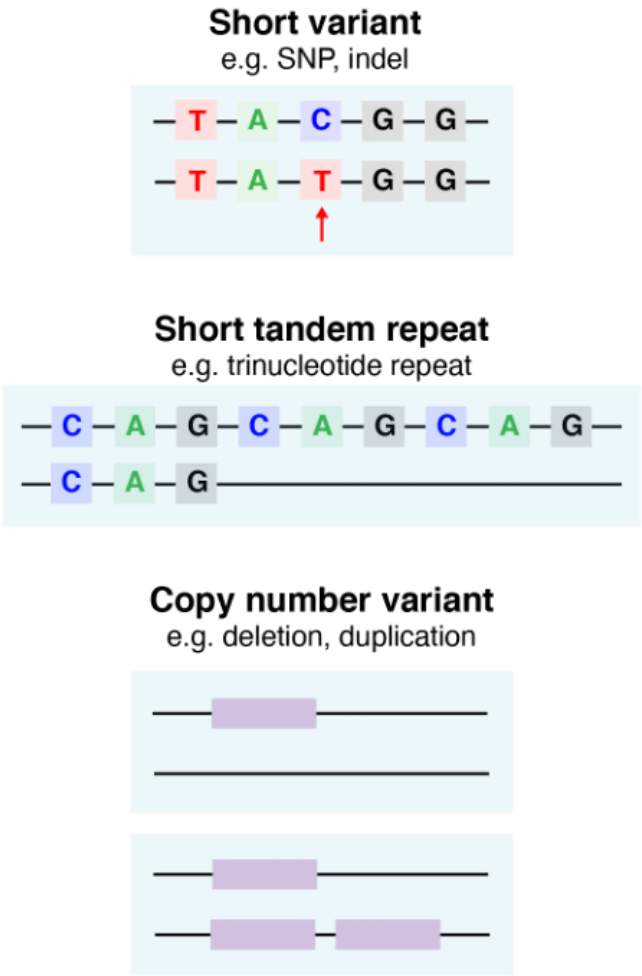


Figure 1. Examples of variant types considered for analysis (Wright, 2022).

Long read sequencing (LRS) platforms offer a promising solution to the limitations of SRS methods by generating overlapping reads longer than ten kilobases from a single molecule, thereby eliminating sequence gaps and improving genome assembly and alignment (Mantere et al., 2019; Owusu & Savarese, 2023). Detection of CNVs and SVs is hindered by the limited read lengths of short-read platforms, particularly in regions of repetitive DNA or abnormal GC content, where these variants are often enriched (Chaisson et al., 2019; Logsdon et al., 2020; Owusu & Savarese, 2023). In contrast, LRS platforms generate longer reads (>10 kb) compared to SRS

platforms (50-300 bp) that can transverse complex genomic regions and capture entire variants in a single read, making long reads especially adept at detecting CNVs and SVs, as well as variants within challenging-to-sequence regions such as repetitive or GC-rich areas (Logsdon et al., 2020; Owusu & Savarese, 2023). Therefore, the improved resolution afforded by long reads has the potential to capture variants that may be missed by traditional sequencing approaches in both known and novel NMD genes (Mantere et al., 2019).

Although LRS is still an emerging area in clinical diagnostics, its affordability and accuracy have improved over the last several years, and numerous studies have already demonstrated the capacity to improve the detection of the aforementioned variant types across a variety of genetic conditions (Cohen et al., 2022; Mantere et al., 2019; Merker et al., 2018; Miao et al., 2018; Sano et al., 2022). For example, in a study by Cohen et al. (2022), both SRS and LRS methods were employed to analyze the genomes of over 900 families with suspected genetic disorders. Variant discovery rate with LRS was markedly increased compared to SRS-based methods, identifying two-fold more SVs, over four-fold more transmitted rare SVs, and 5% more small variants, leading to an increase in diagnostic yield (Cohen et al., 2022).

Preliminary studies have also begun to employ LRS methods to detect disease-causing variants in individuals with NMDs. Examples demonstrating the strength of LRS technology include identification of novel rare CNVs, SVs, and splice-altering variants in the *DMD* gene, which is implicated in Duchenne and Becker muscular dystrophy (Bruels et al., 2022; Chin et al., 2021; Geng et al., 2023; Owusu & Savarese, 2023; Xie et al., 2020). Due to the large size of the *DMD* gene, these complex CNVs and SVs may evade

detection by SRS methods due to insufficient read length (Xie et al., 2020). Novel repeat variants involving NMD genes have also now been resolved with LRS approaches, including a GGC expansion in the *NOTCH2NLC* gene and CCG expansions in the *RILP1* and *ABCD3* genes, which are associated with oculopharyngodistal myopathy, and a D4Z4 contraction in *DUX4* causing facioscapulohumeral muscular dystrophy (Cortese et al., 2024; Owusu & Savarese, 2023; Yeetong et al., 2023; Yu et al., 2020; Yu et al., 2022). As previously mentioned, these variants pose a challenge for SRS platforms due to their repetitive nature, making them refractory to detection with traditional sequencing approaches (Logsdon et al., 2020; Owusu & Savarese, 2023).

While research using LRS platforms has begun to emerge, their application in the molecular diagnosis of NMDs remains relatively understudied, with only a few recent publications. Most of these studies also focus on genes involved in common NMDs (e.g., *DMD*, *SMN1* and *SMN2*, *DMPK*, etc.), and as a result, the full potential of LRS for diagnosing NMDs, particularly those caused by less common genes, has yet to be fully explored (Owusu & Savarese, 2023). In addition, only four peer-reviewed studies have utilized the latest high-throughput LRS platform, the Pacific Biosciences (PacBio) Revio (Pacific Biosciences of California, Inc., 2022). These studies have primarily focused on quality control and the development of tools for long-read data analysis, with no applications to human clinical data reported thus far (Choi et al., 2024; Duan et al., 2024; Kolesnikov et al., 2024; Okumura & Masuda, 2024). Thus, the utility of the PacBio Revio compared to other LRS platforms in detecting human disease-causing variation remains largely unknown. Currently, PacBio and Oxford Nanopore Technologies (ONT) are the two leading providers of LRS platforms (Oehler et al., 2023). While PacBio platforms

require a higher initial investment compared to ONT platforms, they are known for greater accuracy (99.9%), longer read lengths (15-20 kb), lower turnaround time for data analysis, and superior performance in sequencing complex genomic changes, which is well-suited for applications in precise variant detection (Kolesnikov et al., 2024; Oehler et al., 2023). Additionally, instruments such as the PacBio Revio have significantly reduced the cost of LRS in recent years by requiring less consumables than previous models, making studies based on this technology increasingly accessible (Kolesnikov et al., 2024; Oehler et al., 2023). These advancements highlight the growing potential of LRS to revolutionize genomic research by combining high sequencing accuracy with improved affordability.

Additionally, to accommodate the large amounts of data generated by LRS technologies, variant calling algorithms have recently evolved to include sophisticated statistical methods like machine learning to enable more accurate and efficient genotyping of complex regions (Poplin et al., 2018). These called variants can then be interpreted and ranked to yield a small list of probable disease-causing candidates in the affected individual utilizing the most up-to-date variant annotation tools, limiting interpretive hours (Stenton et al., 2024). These tools permit variant filtration and prioritization in accordance with contextual information from data repositories and predictive metrics, including gene and protein information, population frequency data, functional impact predictions, and existing clinical associations as described in the literature or query databases such as ClinVar, Online Mendelian Inheritance in Man (OMIM), and others (Stenton et al., 2024). Candidate variants can also be filtered based on phenotypic terms relevant to the clinical presentation of the affected individual to streamline analysis of

sequencing data to variants most relevant to the condition of interest that are contributing to the observed signs and symptoms (Köhler et al., 2019; Stenton et al., 2024). By applying these filtration steps to data from long-read whole genome sequencing (LR-WGS), potentially pathogenic variants can be identified that were not detected using traditional sequencing approaches, as well as variants in new disease genes.

Taken together, analysis of LR-WGS data with improved tools for variant calling and annotation has the potential to significantly increase the molecular diagnostic yield of NMDs, as prior research has shown. Therefore, the aim of this research project is to identify candidate pathogenic variants in a cohort of genetically undiagnosed individuals with suspected NMDs by developing a bioinformatic pipeline and variant filtration strategy based on LR-WGS data.

1.1. Research objectives and hypothesis

Study Rationale: Many individuals with NMDs remain genetically undiagnosed due to limitations in current genomic technologies used in clinical practice. Recent advancements, such as LRS platforms and improved variant calling and annotation tools, offer potential for enhancing diagnostic capabilities by identifying disease-causing variants in known and novel genes.

Research Question: How can integration of LR-WGS data and advanced bioinformatic strategies improve detection of candidate pathogenic variants and genes in individuals with NMDs that have a suspected genetic cause?

Hypothesis/Prediction: If integration of LRS platforms and advanced bioinformatic strategies affect the diagnostic yield for NMDs, then causative genetic variants should be identified within a cohort of undiagnosed individuals with suspected NMDs who have undergone prior traditional clinical genetic testing.

This will be investigated via the following objectives:

Objective 1: Conduct a clinical chart review to collect and curate phenotypic information to be used in variant prioritization and interpretation.

Objective 2: Develop and validate a pilot genetic variant filtration pipeline to prioritize candidate pathogenic variants from annotated LR-WGS data.

2. MATERIALS & METHODS

2.1. Ethics approval and funding

Ethical approval for this study has been obtained from the University of Manitoba Bannatyne Campus Health Research Ethics Board (REB approval number HS19854/H2016:247). This work is funded by a Research Manitoba New Investigator Operating Grant awarded to collaborator, Dr. Galen Wright.

2.2. Participant recruitment and sample collection

Participants with suspected genetic NMDs meeting the inclusion/exclusion criteria specified in **Table 2** were referred by clinicians in the Program of Genetics and Metabolism at Shared Health, Winnipeg, Manitoba. Each participant (or their legal guardian) met with the study genetic counsellor to obtain information about the study and

provide informed consent, including the release of relevant clinical specimens. Participants also consented to be contacted to receive the results of this study or if further information is required by the study team. Study participants were also informed of the potential of incidental findings, which is the detection of a variant known to cause a genetic disease that is unrelated to their NMD presentation, and have selected their preference to receive these results or not as a part of the consent process. Unaffected familial negative controls were not recruited in this study; however, affected family members were eligible for recruitment.

Table 2. Inclusion and exclusion criteria for recruitment into NMD genetic study.

Inclusion	**Exclusion**
1. The individual is referred for investigation of a potential **neurogenetic condition** to the Shared Health Program of Genetics and Metabolism	1. The affected individual/legal guardian is incapable of providing informed consent
2. The condition is muscular in origin. Has **two** of: ● Muscle weakness ● Episodic rhabdomyolysis ● Cramping or myalgia ● Marked hypotonia ● Persistent creatine kinase elevation ● Fibrofatty replacement on muscle MRI ● Myopathic changes on EMG ● Myopathic or dystrophic changes on muscle biopsy	2. The affected individual has already received a definitive genetic diagnosis
3. The condition has a high likelihood of being monogenic (i.e., younger age of onset, family history, parental consanguinity, etc.)	
4. The affected individual/legal guardian has consented to a skeletal muscle biopsy as part of clinical care	
5. Informed consent is obtained	

For each affected individual, genomic DNA (gDNA) was extracted from whole blood samples. Fresh frozen disease-relevant tissue from muscle biopsies were also released by the clinical laboratory at Shared Health (muscle biopsies will be used for future RNA sequencing (RNA-seq), beyond the scope of this pilot study). For this pilot project, five participants have been prioritized for analysis based on recommendations from Dr. Patrick Frosk (medical geneticist and clinician scientist), considering factors such as phenotype and family history. While this number is an insufficient sample size to calculate and draw meaningful conclusions regarding improvements in diagnostic yield, it is suitable to establish and validate a genetic variant filtration and prioritization pipeline for analysis of future samples.

2.3. Retrospective chart review and phenotype extraction

A retrospective review of each participant's electronic medical records was performed using a data collection form developed by the research team members to extract clinical information relevant to the variant filtration process (see **Supplementary Figure 1**). Information collected from charts included phenotypic descriptions of symptoms, family history and ethnicity, and previous laboratory, imaging, and genetic investigations. Phenotypic descriptions for each participant were then assigned standardized terms using the Human Phenotype Ontology (HPO), a hierarchical vocabulary for classifying human phenotype abnormalities (Köhler et al., 2019; Robinson et al., 2008). Each HPO term has a unique identifier that is linked to associated genes via the OMIM database and phenotype annotations through the Phenotate program, which can be used to clarify genotype-phenotype relationships (Amberger et al., 2015; Chang et al., 2020). The HPO assignment process involved assigning phenotypic terms tailored to

each individual based on their clinical presentation, in accordance with the most specific term, which is then categorized under the broader hierarchy (e.g., an individual presenting with proximal weakness of the lower limbs is assigned the term proximal muscle weakness (HP:0003701) which is categorized under the broader term abnormality of the musculoskeletal system (HP:0033127) within the hierarchy). A genetic counsellor, Jessica Hartley, with consultation with clinical geneticist Patrick Frosk, then independently verified the accuracy of these terms for use in variant filtration. Terms defining the participant's unique phenotype were included and terms relating to incidental findings or previous unrelated medical history were excluded. The final HPO terms were then used to aid in prioritizing genetic variants identified by LR-WGS. All potentially identifying information pertaining to participants has been redacted from this thesis to ensure the protection of their privacy and confidentiality.

2.4. Long-read whole-genome sequencing data transfer and bioinformatics pipeline

Participant samples were sent to The Centre for Applied Genomics (TCAG) facility at The Hospital for Sick Children (SickKids) in Toronto, Ontario for WGS on the PacBio Revio LRS system. Upon arrival, gDNA quality assessment was performed using the Agilent TapeStation 4200 automated electrophoresis system with the gDNA ScreenTape analysis method (Pellecchia, 2024a; Pellecchia, 2024b). Samples were verified to have a concentration between 20-200 ng/µl with a total amount of 5-7 µg of DNA, peak sizes >60kb, and an A260/A280 of 1.8 or greater prior to downstream analysis (Pellecchia, 2024a; Pellecchia, 2024b). For each gDNA sample, a HiFi SMRTbell library was generated using the PacBio SMRTbell Express Template Prep Kit 3.0 prior to sequencing on one Revio SMRT Cell per sample on the PacBio Revio system

with an expected yield of 90GB (30x coverage) with 15-20 kb reads (Pellecchia, 2024a; Pellecchia, 2024b). Sequencing data for each sample was aligned to the most recent human reference genome (GRCh38/hg38) before variants were called using a customized PacBio workflow from TCAG (Pellecchia, 2024a; Pellecchia, 2024b). Sequencing data for related affected individuals was integrated as available in the case of duos or trios.

Variants were called and annotated at the TCAG facility. Small variants were called by machine learning-based tool DeepVariant, developed by Google (Pellecchia, 2024a). CNVs were called by HiFiCNV, and SVs were called by pbsv, all of which were returned as Variant Call Format (VCF) files (Pellecchia, 2024b). Small variant calls were then annotated using a custom small variant annotation pipeline (rev28.2) developed at TCAG (Pellecchia, 2024a; Yuen et al., 2015) based on the previously established ANNOVAR tool (Pellecchia, 2024a; Wang et al., 2010). CNV and SV calls were annotated using a custom annotation pipeline developed at TCAG (Pellecchia, 2024b). Annotated information provided in the annotation pipelines for both small variants and CNVs/SVs includes gene and protein information, disease and phenotype associations (i.e., HPO terms, OMIM, ClinVar, etc.), and population allele frequencies obtained from data repositories and independent cohorts (Pellecchia, 2024a; Pellecchia, 2024b).

Small variant annotated information also included predicted deleteriousness scores from several *in silico* tools (i.e., CADD, REVEL, MPC, MutationTaster, etc.), as well as predicted splicing impact using the SpliceAI prediction tool for splice site variants (Pellecchia, 2024a). Annotated files and raw sequencing data for small variants and CNVs/SVs were transferred via a secure protocol established by TCAG for input to a local computing infrastructure at the University of Manitoba, which is accessed by the

student principal investigator (PI) and research team (Pellecchia, 2024a; Pellecchia, 2024b).

2.5. Genomic variant filtration and analysis

Filtration of annotated variants was performed for all samples as illustrated in the pipeline in **Figure 2** for both small variants and CNVs/SVs. All analyses were performed in R v4.3.3 utilizing the tidyverse v2.0.0 package for data handling and visualization (R Core Team, 2024; Wickham et al., 2019). The filtration strategy utilized a set of consistent "hard" filters applied uniformly across all individuals and variant types, complemented by optional clinical filters tailored to individual features as informed by the chart review process (see **Section 2.2**). Each variant annotation file was analyzed independently for each individual based on variant type (i.e., small variants, CNVs, SVs).

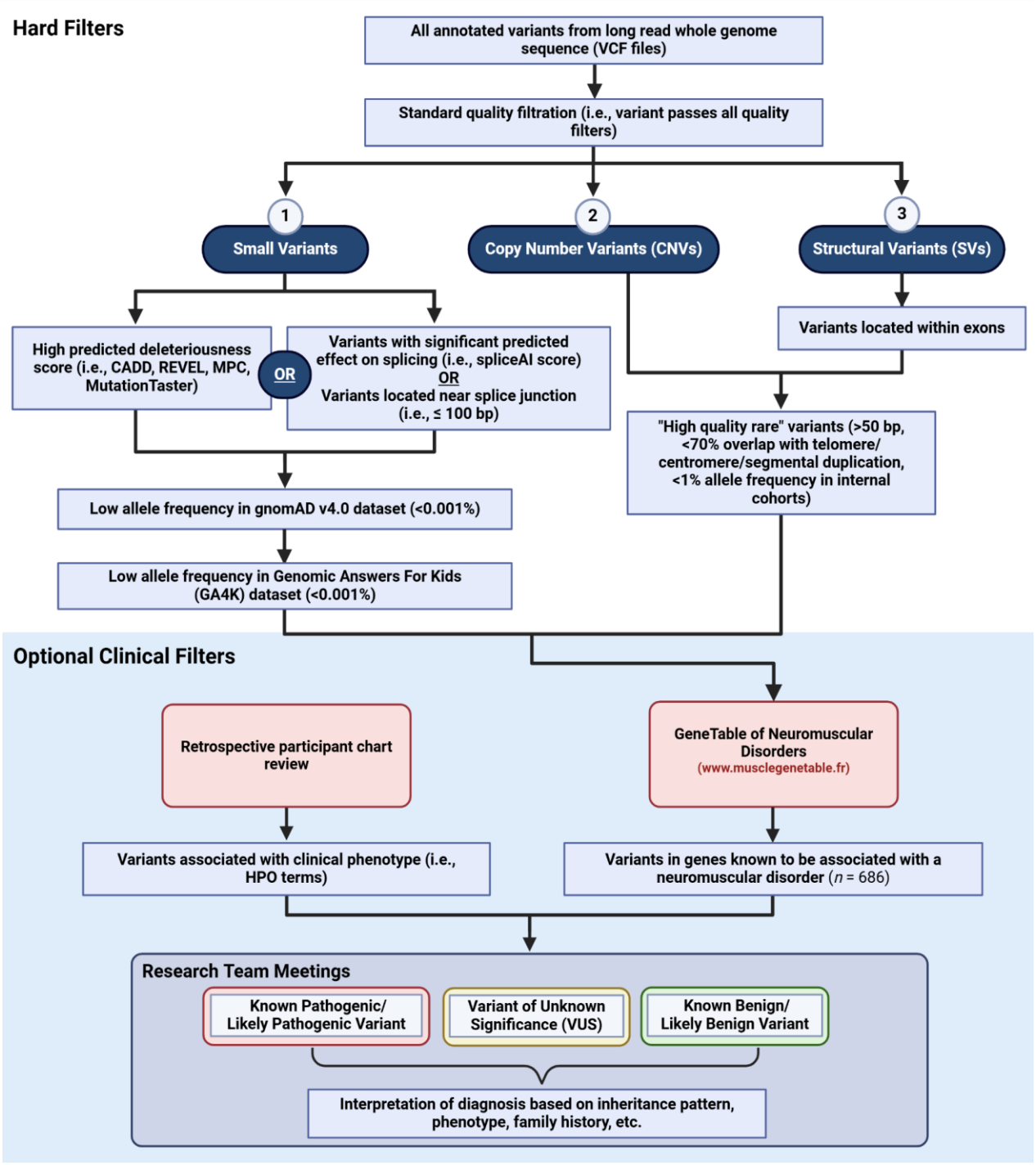


Figure 2. Genetic variant filtration pipeline as applied to small variants and CNVs/SVs (see Section 2.5).

Initially, small variants were filtered to retain only those predicted to be highly deleterious based on scores from various *in-silico* tools such as CADD, REVEL, MPC,

PhyloP, and MutationTaster (Ioannidis et al., 2016; Pollard et al., 2010; Rentzsch et al., 2019; Samocha et al., 2017; Schwarz et al., 2014). The filtering criteria followed thresholds established by SickKids TCAG, based on the American College of Medical Genetics and Genomics (ACMG) guidelines for pathogenicity interpretation, which can be found in **Table 3** for each variant type (Pejaver et al., 2022; Pellecchia, 2024a; Richards et al., 2015; Stenton et al., 2024). For small splice variants, scores determined by dbNSFP v.3.0 (Liu et al., 2016) and SpliceAI (Jaganathan et al., 2019) were also used to determine predicted impact on splicing (Pellecchia, 2024a). To evaluate the potential impact of a genetic variant on splicing, dbNSFP combines outputs from multiple *in-silico* prediction tools that assess alterations to donor and acceptor splice sites as well as overall splicing patterns (Liu et al., 2016). In contrast, SpliceAI employs a deep learning-based neural network to detect splicing variants by predicting splice site junction positions directly from a pre-mRNA sequence (Jaganathan et al., 2019). Therefore, small splice variants with significant predicted impacts on splicing based on these scores or those located within 100 bp of a splice site junction were included based on recommendations from SickKids TCAG (Jaganathan et al., 2019; Liu et al., 2016; Pellecchia, 2024a). CNVs/SVs were also incorporated into the pipeline at this stage (Pellecchia, 2024b).

Next, small variants were filtered to include only rare variants based on minor population allele frequency using a conservative threshold of 0.001% (Stenton et al., 2024). Population allele frequencies for small variants were sourced from the gnomAD v4.0 genome dataset (Karczewski et al., 2020) and from technology-matched LRS data for a small, unrelated cohort of individuals ($n \sim 500$) from the Genomic Answers for Kids (GA4K) initiative at the Children's Mercy Hospital Research Institute (Cohen et al.,

2022; Pellecchia, 2024a). Population allele frequency filtration followed a two-step approach: small variants were first filtered based on their frequency in the larger gnomAD dataset (short reads), followed by their frequency in the smaller, technology-matched GA4K dataset (long reads). This approach allowed for assessing the number of small variants excluded by each individual dataset across individuals with each population control database.

For CNVs/SVs, variants were filtered to include only those annotated as "*high quality rare*" by TCAG, which applies to variants >50 bp with <70% overlap with telomeres, centromeres, or segmental duplications, and a population allele frequency <1% based on internal cohorts available to SickKids TCAG (Pellecchia, 2024b). For CNVs, TCAG internal cohorts included a database of unrelated samples sequenced using the Illumina NovaSeq 6000 at an average coverage of 35x, called by Dragen 3.8.4, with at least 90% reciprocal overlap (Pellecchia, 2024b). Additional allele frequencies for CNVs/SVs were obtained from a long-read analysis of SVs called from over 500 unrelated samples sequenced by Children's Mercy Hospital, with at least 50% reciprocal overlap, as well as NTSD186, a heterogeneous dataset from NCBI (Pellecchia, 2024b). These datasets serve as interim references until a specific dataset for CNVs/SVs based on PacBio LRS data becomes available. As a large number of SVs was returned, CNVs/SVs larger than 1 kb that encompass at least one gene, as recommended by TCAG, were also incorporated into the pipeline at this stage (Pellecchia, 2024b).

Table 3. Small variant predicted deleteriousness classifications based on thresholds defined by SickKids TCAG (Pellecchia, 2024a).

Variant type	Predicted functional impact	Threshold(s)
Missense (non-synonymous single nucleotide variants)	**High**	REVEL_score ≥ 0.45
	Medium	REVEL_score ≥ 0.25 **OR** phylopMam_avg ≥ 1.30 **OR** phylopVert100_avg ≥ 3.90 **OR** MPC_score ≥ 2.00 **OR** CADD_phred ≥ 30.0
	Low	Missense variants that do not meet the above two criteria.
Other (non-frameshift deletions, insertions, substitutions, and stoploss variants)	Medium	((phylopMam_avg ≥ 1.30 **OR** phylopVert100_avg ≥ 3.90 **OR** CADD_phred ≥ 21.1) **AND** (dbsnp_common != "NA")) **OR** ((phylopMam_avg ≥ 1.10 **OR** phylopVert100_avg ≥ 1.60 **OR** CADD_phred ≥ 13.7) **AND** (dbsnp_region != "NA"))
	Low	'Other' variants that do not meet the above criteria.
Predicted splicing	**High**	dbscSNV_ADA_SCORE > 0.6 **OR** dbscSNV_RF_SCORE > 0.6 **OR** (spliceAI_DS_AG > 0.5 **AND** spliceAI_DP_AG ≤ 50) **OR** (spliceAI_DS_AL > 0.5 **AND** spliceAI_DP_AL ≤ 50) **OR** (spliceAI_DS_DG > 0.5 **AND** spliceAI_DP_DG ≤ 50) **OR** (spliceAI_DS_DL > 0.5 **AND** spliceAI_DP_DL ≤ 50)
	Medium	dbscSNV_ADA_SCORE > 0.6 **OR** dbscSNV_RF_SCORE > 0.6

		OR (spliceAI_DS_AG > 0.2 **AND** spliceAI_DP_AG ≤ 50) **OR** (spliceAI_DS_AL > 0.2 **AND** spliceAI_DP_AL ≤ 50) **OR** (spliceAI_DS_DG > 0.2 **AND** spliceAI_DP_DG ≤ 50) **OR** (spliceAI_DS_DL > 0.2 **AND** spliceAI_DP_DL ≤ 50)
	Low	Predicted splicing variants that do not meet the above two criteria.
Non-coding RNA	**High**	((phylopMam_avg ≥ 1.30 **OR** phylopVert100_avg ≥ 3.90 **OR** CADD_phred ≥ 21.1) **AND** (phastCons_placental != "NA"))
	Medium	((phylopMam_avg ≥ 1.10 **OR** phylopVert100_avg ≥ 1.60 **OR** CADD_phred ≥ 13.7) **AND** (phastCons_placental != "NA"))
	Low	Non-coding RNA variants that do not meet the above two criteria.
3' or 5' UTR	Low	Any UTR variants.

*__REVEL_score__ = Rare Exome Variant Ensemble Learner (REVEL) scores; **phylopMam_avg** = average value of PhyloP nucleotide-level conservation inferred from the Placental Mammal genome group; **phylopVert100_avg** = average value of PhyloP nucleotide-level conservation inferred from the 100 Vertebrate genome group; **MPC_score** = MPC (**M**issense badness, **P**olyPhen-2, and **C**onstraint) score based on the Exome Aggregation Consortium (ExAC) database; **CADD_phred** = PHRED-like c-score ranking from CADD (Combined Annotation Dependent Depletion); **dbsnp_common** = exact match (position, allele) to common dbSNP track USCS; **dbsnp_region** = overlap-based match for dbSNP;
dbscSNV_ADA_SCORE/dbscSNV_RF_SCORE = splice site prediction scores from dbscSNV;
spliceAI_DS_AG/spliceAI_DP_AG = SpliceAI Delta score (DS) and Delta position (DP), the probability of the variant being splice-altering via acceptor gain (AG);
spliceAI_DS_AL/spliceAI_DP_AL = SpliceAI Delta score (DS) and Delta position (DP), the probability of the variant being splice-altering via acceptor loss (AL);
spliceAI_DS_DG/spliceAI_DP_DG = SpliceAI Delta score (DS) and Delta position (DP), the probability of the variant being splice-altering via donor gain (DG);

spliceAI_DS_DL/spliceAI_DP_DL = SpliceAI Delta score (DS) and Delta position (DP), the probability of the variant being splice-altering via donor loss (DL); **phastCons_placental** = PhastCons score for the Placental Mammal genome group, based on UCSC track (calculated by HMM scan of PhyloP values to infer conserved status).

After application of these "hard" filters, additional clinical filters based on chart review data (defined below) were used on a case-by-case basis. For each affected individual, a tiered filtration approach was applied to obtain a manageable list of candidate variants in the following order:

(1) Filtration for variants matching at least one of the selected HPO terms for the participant. This approach targets the observed phenotype of the affected individual and limits incidental findings.

(2) Filtration for variants in genes known to be implicated in NMDs as obtained from the most recent version of the GeneTable of Neuromuscular Disorders (*n* = 686), a curated repository (Benarroch et al., 2024).

(3) If no candidate variants were identified, the filtration strategy was relaxed to examine variants identified by the HPO term filter alone. This approach allows for a higher number of variants to be considered, increasing the likelihood of capturing novel or previously unrecognized genetic associations by including variants in genes not yet linked to NMDs.

Finally, after applying the filtration steps described above, genetic variants likely resulting from sequencing artifacts or insufficient annotation (e.g., missing population allele frequency) were addressed by excluding those that appeared across two or more unrelated individuals, yielding a final shortlist of variants for interpretation.

2.6. Manual genomic variant prioritization and validation

Prioritized variants from filtration were reviewed for consistency with potential inheritance pattern(s) (i.e., confirming the presence of two affected alleles for an autosomal recessive pattern of inheritance), as well as information previously reported in OMIM and the gnomAD and ClinVar databases as available to identify potential candidates (Amberger et al., 2015; Karczewski et al., 2020; Landrum et al., 2014). Following this preliminary review, candidate variants were assessed for pathogenicity following ACMG guidelines as described below in **Figure 3 [REDACTED]** and visualized using the Integrative Genomics Viewer (IGV) for manual review and confirmation (Richards et al., 2015; Robinson et al., 2017). The University of Maryland Genetic Variant Interpretation Tool (Kleinberger et al., 2016) was used to assign the ACMG classification to each variant. Supplementary tools, such as Franklin by Genoox (Genoox, n.d.) and the UCSC Genome Browser (Perez et al., 2025) were used for additional visualization and variant information as required. The student PI then presented the analyzed variants to the research team at regular meetings, including a clinician scientist (Dr. Patrick Frosk), genetic counsellor (Jessica Hartley), bioinformaticians (John Pham), and human genetics researchers with expertise in neurogenetics and large-scale computational analysis of rare pathogenic variants (Drs. Galen Wright and Britt Drögemöller). The strength of evidence for each variant was discussed based on its predicted deleteriousness, mode of inheritance, and clinical relevance, considering factors such as the phenotype and family history of the affected individual. Findings from the analysis were then compared to the results of standard clinical sequencing previously used to assess LRS performance on a case-by-case basis. The team decided plans for follow-up with affected individuals as necessary to confirm the results (e.g., Sanger

sequencing, RNA-seq, model organism validation, etc.) or to request samples for familial studies as required.

Figure 3. The American College of Medical Genetics and Genomics (ACMG) criteria evidence framework (Figure 1 from: Richards et al., 2015). "This chart organizes each of the criteria by the type of evidence as well as the strength of the criteria for a benign (left side) or pathogenic (right side) assertion... BS, benign strong; BP, benign supporting; FH, family history; LOF, loss of function; MAF, minor allele frequency; path., pathogenic; PM, pathogenic moderate; PP, pathogenic supporting; PS, pathogenic strong; PVS, pathogenic very strong" (Richards et al., 2015). **[Figure REDACTED to comply with copyright requirements. The figure can be found at: Richards et al. (2015). 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. *Genetics in Medicine, 17*(5), 405–424. https://doi.org/10.1038/gim.2015.30]**

3. RESULTS

A total of five individuals were analyzed in this pilot study, with each case described herein corresponding to a study participant ID (i.e., Cases 1-5). The study cohort included three males and two females (two males were full siblings). Long-read whole genome sequencing (LR-WGS) yielded an average sequencing depth of **34.6X** and an average read length of **17.4 kb** across samples. Variant filtration, as previously described, was performed for each participant across three variant types: small variants, CNVs, and SVs. **Table 4** summarizes the number of variants identified after initial variant calling and the number removed at each filtration step across participants. After filtration and removal of duplicate variants, a total of 21 small variants, three CNVs, and three SVs were retained for further clinical interpretation and assessment of pathogenicity using ACMG guidelines across this cohort. Candidate variants were identified in four out of five cases in known disease-causing genes. In three cases, the variants were classified as

"***Pathogenic/Likely Pathogenic***," and in one case, the variants were classified as "***Variant of Uncertain Significance***." The phenotypic descriptions, family history, diagnostic testing history, and interpretation of the top candidate variants for each case are described below.

Table 4A. Number of small variants identified prior to and following variant filtration steps for Cases 1-5.

Case #	Total # of small variants	# after gnomAD frequency* filter (short reads)	# after GA4K frequency** filter (long reads)	# after HPO terms filter	# after NMD gene filter	# after NMD gene + HPO term filters
1	5,673,137	163	94	35	9	9
2	5,593,514	104	40	14	4	4
3	5,674,913	133	52	17	3	3
4	5,620,833	132	58	13	2	2
5	5,609,984	122	45	13	3	3

*****gnomAD frequency = <0.001 population allele frequency in gnomAD v4.0 database based on data from short reads; ******GA4K frequency = <0.001 population allele frequency in Genomic Answers for Kids (GA4K) database based on technology-matched data from long reads

Table 4B. Number of copy number variants (CNVs) identified prior to and following variant filtration steps for Cases 1-5.

Case #	Total # of CNVs	# after "*high quality rare*"* filter	# after HPO terms filter	# after NMD gene filter	# after NMD gene + HPO term filters
1	69	4	2	1	1
2	77	5	1	0	0
3	69	4	0	0	0

4	77	5	1	1	1
5	70	7	1	1	1

******high quality rare* = <1% population allele frequency in internal PacBio datasets, >50 bp, <70% overlap with segmental duplications/telomere/centromere

Table 4C. Number of structural variants (SVs) identified prior to and following variant filtration steps for Cases 1-5.

Case #	**Total # of SVs**	**# after "*high quality rare*"* filter**	**# after exons only filter**	**# after HPO terms filter**	**# after NMD gene filter**	**# after NMD gene + HPO term filters**
1	51,778	2,487	119	25	2	1
2	51,452	2,145	97	14	0	0
3	51,944	2,187	84	19	1	1
4	51,085	2,076	71	15	1	1
5	51,207	2,115	93	16	0	0

******high quality rare* = <1% population allele frequency in internal PacBio datasets, >50 bp, <70% overlap with segmental duplications/telomere/centromere

3.1. Case 1

3.1.1. Case Description: Case 1 is a 12-year-old female of Syrian and Palestinian descent. Her parents are first cousins (i.e., consanguineous), but otherwise, the family history was unremarkable (see pedigree; **Figure 4 [REDACTED]**). The potential inheritance patterns for this case are listed in order of likelihood in **Table 5A**. The pregnancy history was unremarkable, and she was born at 36 weeks gestation following a slow delivery where there were concerns for fetal decelerations. She required a three-week hospitalization due to feeding difficulties, fatigue, and hyperbilirubinemia. She had a birth weight of 2956 g (50 to 90th percentile), a length of 53 cm (50th percentile), and a head circumference of 34 cm (>98th percentile).

Figure 4. Pedigree for Case 1 (NM12; collected 05-Jan-2022). **[REDACTED]**

On neonatal examination, she had severe peripheral hypotonia and upper limb contractures. At three days old, she was evaluated by a genetics physician, who described mild facial dysmorphism including a high forehead, frontal bossing, high-arched eyebrows, a high-arched palate, microretrognathia, and thin-simple auricles. She also displayed dolichocephaly with a prominent occiput and metopic ridge, bilateral ptosis, bilateral infra-orbital creases, and a thin upper lip vermillion. By two years of age, she continued to exhibit mild arthrogryposis, feeding challenges, minimal finger joint contractures, and delays in gross motor skills. On most recent physical examination at 7 years old, she demonstrated diffuse, non-progressive weakness, particularly in the knee and hip flexors and shoulder abductors, along with poor muscle bulk, decreased deep tendon reflexes, a waddling gait, and a positive Gowers sign. She is also hyperflexible, exhibiting swan-necking of her distal joints, long fingers, and thoracic scoliosis. Currently, she completes a regular school program, however, she continues to struggle with gross motor skills to keep up with her peers, standing from a seated position, and balancing on one foot. The HPO terms selected for this case based on clinical presentation are listed in **Table 5B**.

Table 5A. Potential inheritance pattern(s) for Case 1 in order of likelihood (most to least likely).

Inheritance Pattern	**Symbol**
Autosomal recessive	AR
De novo autosomal dominant	*De novo* AD
X-linked	XL

Table 5B. Human Phenotype Ontology (HPO) terms selected for Case 1 based on their clinical presentation.[1]

HPO Term Description	HPO Code
Dolichocephaly	HP:0000268
Prominent occiput	HP:0000269
Frontal bossing	HP:0002007
Microretrognathia	HP:0000308
High forehead	HP:0000348
Prominent metopic ridge	HP:0005487
Highly arched eyebrow	HP:0002553
Thin upper lip vermilion	HP:0000219
High palate	HP:0000218
Simple ear	HP:0020206
Congenital bilateral ptosis	HP:0007911
Decreased muscle mass	HP:0003199
Fatty replacement of skeletal muscle	HP:0012548
Minicore myopathy	HP:0003789
Hypertrophied muscle fibers	HP:0100293
Increased variability in muscle fiber diameter	HP:0003557
Generalized muscle weakness	HP:0003324
Proximal muscle weakness	HP:0003701

Hypotonia	HP:0001252
Generalized joint hypermobility	HP:0002761
Congenital contracture	HP:0002803
Delayed gross motor development	HP:0002194
Waddling gait	HP:0002515
Reduced tendon reflexes	HP:0001315
Feeding difficulties	HP:0011968
Abnormal finger morphology	HP:0001167

[1]The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3.

3.1.2. Investigations: Metabolic studies of blood and urine were normal. Creatine kinase (CK) level was normal at two years and seven months. Magnetic resonance imaging (MRI) of the brain at eight days of age was unremarkable except for a mildly increased T1 signal within the globus pallidus, likely related to her hyperbilirubinemia at that time. A muscle MRI conducted in infancy and repeated at age 3 revealed symmetrical fibrofatty replacement and paucity of muscle tissue in the posterior and anterior compartments of both thighs, with severe replacement in the posterior thigh and compensatory hypertrophy of the adductor muscles. Partial fibrofatty replacement of the quadriceps and mild bilateral replacement in the soleus of the calves were also noted. Electromyography (EMG) and nerve conduction velocity (NCV) studies were normal. Muscle biopsy of the left quadriceps femoris showed myopathy with myofiber size variation, oxidative enzyme deficit, and multiple foci of contractile element disorganization suggestive of multi-minicore myopathy. Absence of myoadenylate deaminase activity was also

reported, but it is unclear whether this is an incidental finding or related to the myofiber abnormalities.

She had a normal karyotype, and a single nucleotide polymorphism (SNP) microarray revealed no copy number variants, but several regions of homozygosity were identified. Single-gene analysis of the *SMN1, CHST3, DMPK,* and *FBN2* genes were normal, which are known to cause spinal muscular atrophy, spondyloepiphyseal dysplasia with congenital joint dislocations, autosomal dominant myotonic dystrophy, and autosomal dominant contractual arachnodactyly, respectively. Clinical singleton exome sequencing performed in 2015 (GRCh37/hg19 genome assembly) found many variants of uncertain significance (VUSs), including a heterozygous variant in the gene *RYR1,* a heterozygous variant in the gene *CHRNE,* a heterozygous variant in the gene *SYNE2,* a heterozygous variant in the gene *RYR3,* and two homozygous variants in the gene *TTN* (c.11023G>A, p.Gly3675Ser and c.33340+3A>G). The c.33340+3A>G homozygous VUS in *TTN* was identified by the genetics clinician as a strong candidate based on the participant's phenotype requiring further investigation.

3.1.3. Variant Filtration:

Variant filtration yielded 12 variants for further assessment, including nine small variants, one CNV, and two SVs. The relevant bioinformatic and clinical information for the relevant variants are summarized in **Tables 6, 7, and 8 [Tables 7 and 8 REDACTED]**. It is worth noting that five of the six variants previously detected via exome sequencing were identified through the filtration approach, including variants in *CHRNE, SYNE2, RYR3,* and *TTN* (c.11023G>A, p.Gly3675Ser; c.33340+3A>G), whereas the *RYR1* variant was filtered out as a candidate as it

had a low predicted deleteriousness based on *in-silico* tools (Missense-Low; see **Table 3**), but its presence was verified in the sequencing data.

Table 6A. Clinical information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 1 (*n* = 9).

Gene symbol	Variant (cDNA; protein)	Chromosomal position	Variant type	Zygosity; Genotype	ClinVar classification	OMIM phenotype(s)	Inheritance pattern	Clinical impression	ACMG classification
*TTN**	c.34291+3A>G	chr2:178677618T>C	Intronic	Homo-zygous; C/C	Not reported	- Cardiomyopathy, dilated, 1G - Cardiomyopathy, familial hypertrophic, 9 - Congenital myopathy 5 with cardiomyopathy - Muscular dystrophy, limb-girdle, autosomal recessive 10 - Myopathy, myofibrillar, 9, with early respiratory failure - Tibial muscular dystrophy, tardive	Autosomal recessive; autosomal dominant	Condition matches phenotype, inheritance pattern possible	Uncertain Significance (PM2, PP3)
*TTN**	c.11023G>A; p.Gly3675Ser	chr2:178741259C>T	Exonic	Homo-zygous; T/T	Uncertain Significance (2 reports)	See above	Autosomal recessive; autosomal dominant	Condition matches phenotype, inheritance pattern possible	Uncertain Significance (PM2, PP3, BP1)

Table 6B. Bioinformatic information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 1 (*n* = 9).

Gene symbol	Variant (cDNA; protein)	Chromosomal position	Variant effect	Zygosity; Genotype	gnomAD allele frequency	GA4K allele frequency	Predicted functional impact	Impact score metrics
*TTN**	c.34291+3A>G	chr2:178677618T>C	Splice-altering	Homo-zygous; C/C	0	0	Predicted Splicing High, Rare0	dbscSNV_RF_ADA_SCORE = 0.9987; dbscSNV_RF_SCORE = 0.914
*TTN**	c.11023G>A; p.Gly3675Ser	chr2:178741259C>T	Missense	Homo-zygous; T/T	0.00001583	0	Missense High, Rare01	REVEL = 0.45

*Variants previously identified by clinical genetic testing approaches for this participant.

Table 7A. Clinical information for CNVs in known NMD genes that matched at least one of the selected HPO terms for Case 1 (*n* = 1). **[REDACTED]**

Table 7B. Bioinformatic information for CNVs in known NMD genes that matched at least one of the selected HPO terms for Case 1 (*n* = 1). **[REDACTED]**

Table 8A. Clinical information for SVs in known NMD genes that matched at least one of the selected HPO terms for Case 1 (*n* = 2). **[REDACTED]**

Table 8B. Bioinformatic information for SVs in known NMD genes that matched at least one of the selected HPO terms for Case 1 (n = 2). **[REDACTED]**

3.1.4. TTN

After review with the research team, two homozygous variants in the *TTN* gene were of greatest interest, which were observed in the participant's original clinical exome in 2015 and are reported in **Table 6**. Pathogenic variants in *TTN* have been reported in several NMD phenotypes with both autosomal dominant and recessive inheritance, including tibial and limb-girdle muscular dystrophies, as well as congenital and myofibrillar myopathies, which may match the phenotype of this participant (OMIM #: 611705, 608807, 603689, 600334; Amberger et al., 2015).

The first homozygous *TTN* variant is the missense variant c.11023G>A (p.Gly3675Ser). Visual confirmation of this variant was completed using IGV (**Figure 5**) and the Franklin by Genoox tool (**Figure 6**; see **Supplementary Figure 2** for additional visualization with the UCSC Genome Browser). According to the ACMG criteria, this variant is classified as a ***Variant of Uncertain Significance*** using the following evidence: absent from controls or at an extremely low frequency if recessive (**PM2**), and multiple lines of computational evidence support a deleterious effect (**PP3**). It was listed as a VUS in ClinVar twice by Eurofin EGL Genetics in 2017 and GeneDx in 2024, but the conditions for this classification were not provided (Accession: VCV000593018.5; Landrum et al., 2014).

This variant has been seen in 29 heterozygotes in the gnomAD database, but not in a homozygous state (Karczewski et al., 2020; **PM2**). *In-silico* software predicts a potential functional impact (PolyPhen: 1; MPC: 0.463; REVEL: 0.454; CADD: 18.68). PolyPhen strongly supports a deleterious effect, REVEL and MPC provide moderate predictions of deleteriousness, and CADD suggests a possible deleterious effect (**PP3**). However, according to the Atlas of Cardiac Genetic Variation (Cardiovascular Genetics & Genomics Group, n.d.), this variant does

not appear to impact the N2A soleus/skeletal long isoform transcript of the *TTN* gene within the skeletal muscle, and only impacts cardiac transcripts, as such, it may be less likely to cause a congenital neuromuscular phenotype making it a weaker candidate for consideration (Walsh et al., 2017).

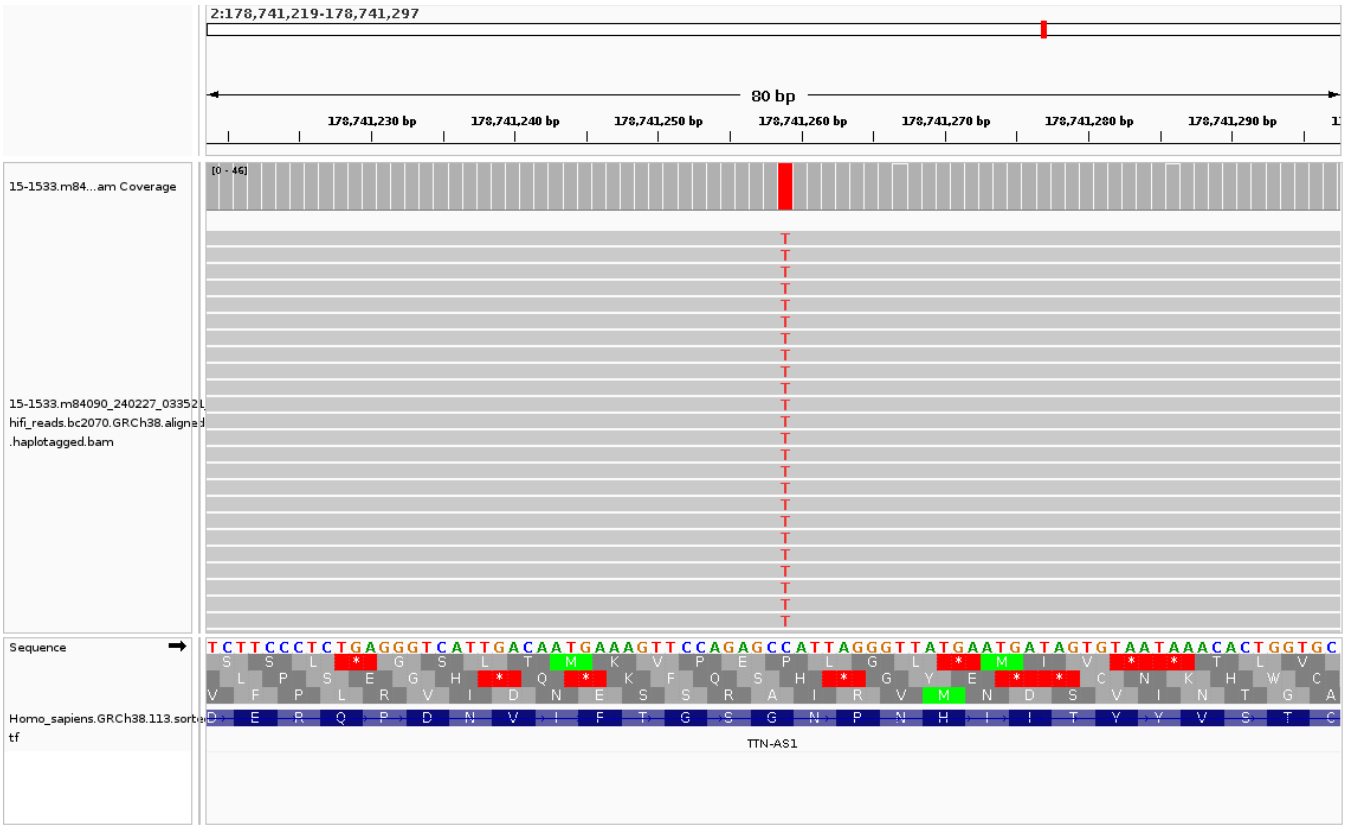


Figure 5. Integrative Genomics Viewer (IGV) snapshot of the sequencing read alignments for Case 1, showing the presence of an exonic variant in the *TTN* gene (c.11023G>A, p.Gly3675Ser) as indicated by the central red line.

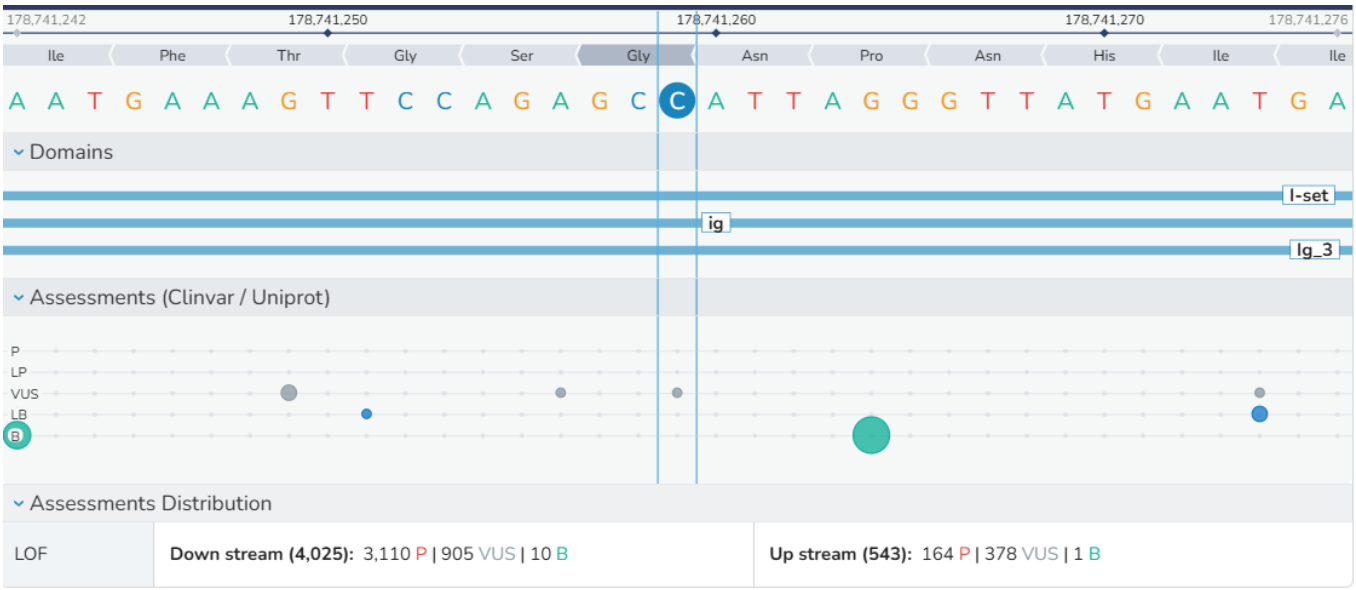


Figure 6. Franklin by Genoox snapshot visualizing the genomic regions affected by the exonic variant in the *TTN* gene (c.11023G>A, p.Gly3675Ser) as identified in Case 1. The boundaries of the variant are indicated by blue vertical lines. Dots under 'Assessments' indicate nearby variants, colour-coded by ACMG classification (red = *Pathogenic*, orange = *Likely Pathogenic*, grey = *Variant of Uncertain Significance*, blue = *Likely Benign*, green = *Benign*).

The second homozygous *TTN* variant is the intronic variant c.34291+3A>G (chr2:178677618T>C). This is the same intronic *TTN* variant as previously detected in the participant's clinical exome sequencing. Visual confirmation of the variant was completed using IGV (**Figure 7**) and the Franklin by Genoox tool (**Figure 8**; see **Supplementary Figure 3** for additional visualization using the UCSC Genome Browser). Based on the ACMG classification, this variant is predicted to be a ***Variant of Uncertain Significance*** based on the following criteria: absent from controls or at an extremely low frequency if recessive (**PM2**), and multiple lines of computational evidence support a deleterious effect (**PP3**).

This variant has not been previously reported in gnomAD or ClinVar (Karczewski et al., 2020; Landrum et al., 2014; **PM2**). It is located three base pairs upstream of a canonical splice site within intron 147, with exon 147 known to contribute to the N2A skeletal muscle transcript

of *TTN*, and *in-silico* tools predict a strong impact on normal splicing (dbscSNV_RF_ADA_SCORE = 0.9987; dbscSNV_RF_SCORE = 0.914; Cardiovascular Genetics & Genomics Group, n.d.; Walsh et al., 2017). Supporting evidence of deleteriousness is also provided by a CADD score of 23.5 (**PP3**).

The research team is interested in conducting additional functional studies for both *TTN* variants, however the intronic splicing variant (c.34291+3A>G) is of greatest interest as additional functional evidence (i.e., RNA-seq studies) could possibly change the classification to ***Likely Pathogenic***, suggesting a diagnosis for this individual.

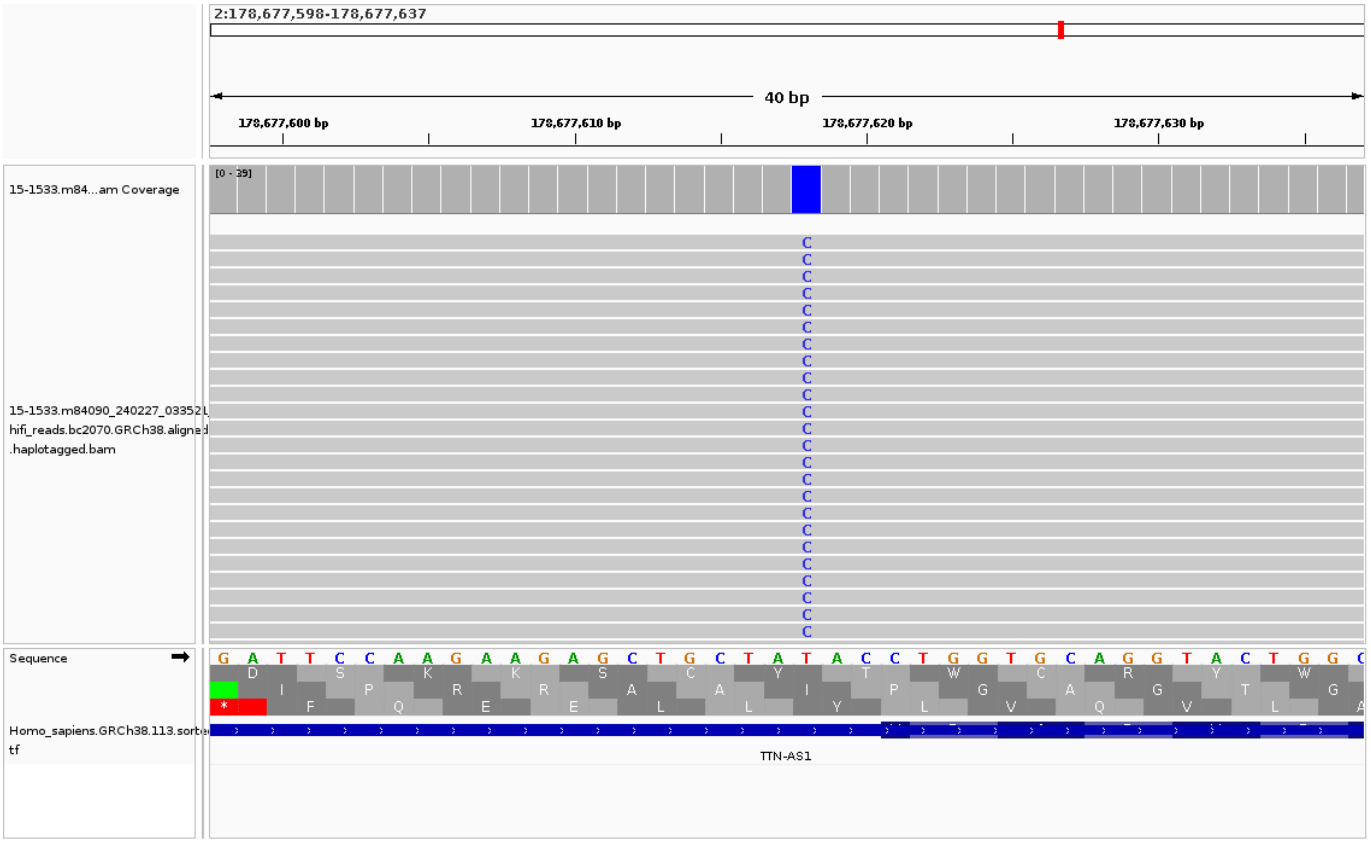


Figure 7. Integrative Genomics Viewer (IGV) snapshot of the sequencing read alignments for Case 1, showing the presence of an intronic variant in the *TTN* gene (c.34291+3A>G; chr2:178677618T>C) as indicated by the central blue line.

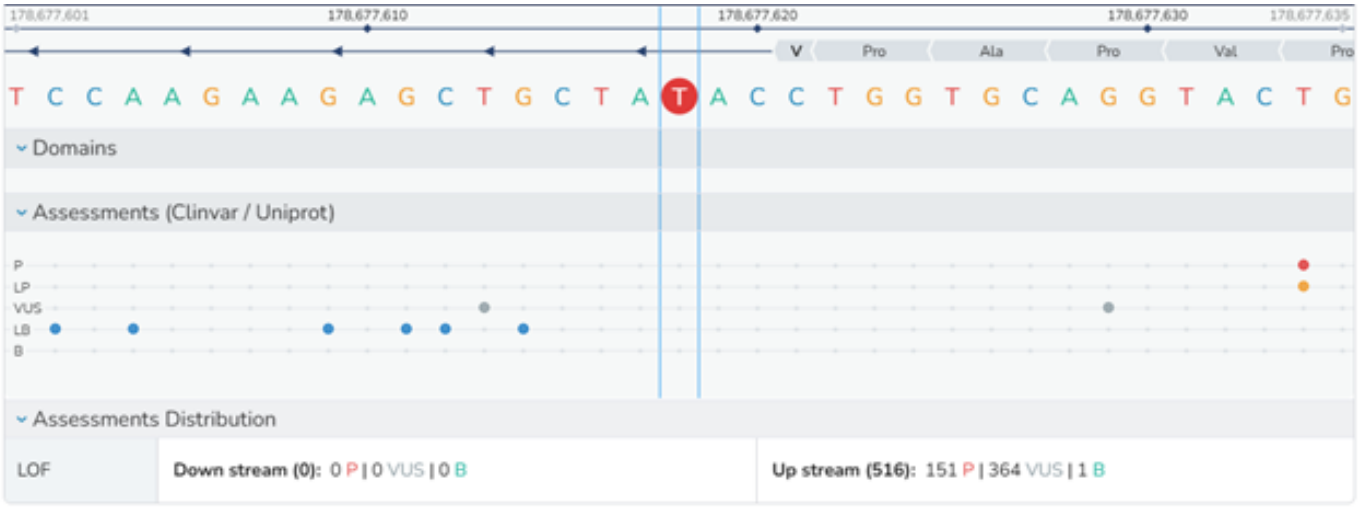


Figure 8. Franklin by Genoox snapshot visualizing the genomic regions affected by the intronic variant in the *TTN* gene (c.34291+3A>G; chr2:178677618T>C) as identified in Case 1. The boundaries of the variant are indicated by blue vertical lines. Dots under 'Assessments' indicate nearby variants, colour-coded by ACMG classification (red = *Pathogenic*, orange = *Likely Pathogenic*, grey = *Variant of Uncertain Significance*, blue = *Likely Benign*, green = *Benign*).

3.2. Case 2

3.2.1. Case Description: Case 2 is a 16-year-old male of German, Scottish, Irish, English, Finnish, and Mennonite descent. He has a healthy older sister, and the family history is remarkable for a great-great uncle, great aunt, and second cousin with polycystic kidney disease (see pedigree; **Figure 9 [REDACTED]**). His parents are non-consanguineous. The potential inheritance patterns for this case are listed in order of likelihood in **Table 9A**. He was born at term with an unremarkable pregnancy history and neonatal course. Birth weight was 4.337 kg. His mother indicates that his neuromuscular difficulties began around 3 years of age, presenting as a gait abnormality affecting predominantly the right leg.

Figure 9. Pedigree for Case 2 (NM7; collected 06-Aug-2015) **[REDACTED]**

The first assessment in genetics was at 7 years of age due to a history of frequent pain and fatiguability in his thighs, heels, and left shoulder exacerbated by exercise, along with occasional back pain. On examination, he was a healthy and non-dysmorphic child with a weight of 25.6 kg (50th percentile), height of 135 cm (95th percentile), and head circumference of 53.5 cm (50 to 75th percentile). He had low to normal muscle tone overall, an incomplete positive Gower sign, and a bilateral Trendelenburg gait with the right foot pointed outwards. He exhibited mild, non-progressive bilateral proximal weakness primarily affecting the deltoids, hip flexors, and plantar flexors. He also developed constipation and urinary issues with possible urinary retention and was found to have residual urine in his bladder. Follow-up examinations at 8 and 9 years old revealed hamstring tightness, progressive ankle flexion contractures, and slightly reduced muscle bulk. At age 10, he was also noted to have mild proximal and distal hand weakness, antalgic movements, and slight hyperflexibility in his hands. At age 14, a follow-up examination revealed mild kyphosis with reduced lumbar lordosis. Developmental delays have also been noted in both his gross and fine motor skills, and he continues to have difficulty with stairs, standing for long periods, and holding small objects such as writing and eating utensils. He does, however, complete a regular academic program at school with a modified gym class and uses a computer to complete assignments. His past medical history is remarkable for snoring and asthma, which is well-controlled. As of his most recent examination at 15 years old, he has remained stable, receiving regular physiotherapy and occupational therapy and utilizing ankle-foot orthoses. He has also received Botox treatment in his heels. The HPO terms selected for this case based on clinical presentation are listed in **Table 9B**.

Table 9A. Potential inheritance pattern(s) for Case 2 in order of likelihood (most to least likely).

Inheritance Pattern	**Symbol**
Autosomal recessive	AR
De novo autosomal dominant	*De novo* AD
X-linked	XL

Table 9B. Human Phenotype Ontology (HPO) terms selected for Case 2 based on their clinical presentation.[1]

HPO Term Description	**HPO Code**
Delayed gross motor development	HP:0002194
Delayed fine motor development	HP:0010862
Waddling gait	HP:0002515
Abnormal muscle tone	HP:0003808
Fatiguable weakness of skeletal muscles	HP:0030197
Proximal muscle weakness	HP:0003701
Decreased muscle mass	HP:0003199
Minicore myopathy	HP:0003789
Ankle flexion contracture	HP:0006466
Exercise-induced myalgia	HP:0003738

[1]The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3.

3.2.2. Investigations: Biochemical investigations, including creatine kinase levels, metabolic screens, and screening for congenital defects of glycosylation were normal. MRI of the brain at 6 years old suggested a mild Chiari Type 1 malformation. MRI of the spine was normal. MRI of the thigh and calf muscles at 7 years old demonstrated no fatty replacement or atrophy. EMG and NCV studies were normal. Muscle biopsy of the left quadriceps femoris at 8 years old revealed multiple foci with myofibrillar disarray, loss of oxidative enzymes, and sarcomeric abnormalities not shown with a modified Gomori stain, consistent with multi-minicore myopathy or a central core myopathy.

At 8 years old, a structural myopathy NextGen sequencing panel (Prevention Genetics, 2016) including 42 genes identified a heterozygous, maternally inherited, synonymous VUS in the *ACTA1* gene. *ACTA1* variants are known to cause both autosomal dominant and recessive congenital myopathies. Clinical exome sequencing performed in 2022 at 13 years old identified a heterozygous, paternally inherited VUS in *HSPB8*. Pathogenic variants in *HSPB8* are known to cause autosomal dominant Charcot-Marie-Tooth type 2L and distal hereditary motor neuronopathy type IIA. Exome sequencing reanalysis performed two years later did not detect any further variants of interest.

3.2.3. Variant Filtration:

Variant filtration yielded four variants for further assessment, including four small variants, no CNVs, and no SVs. The relevant bioinformatic and clinical information for the relevant variants is summarized in **Table 10 [REDACTED]**. While the variant filtration approach identified the *HSPB8* heterozygous variant previously detected through exome sequencing, the *ACTA1* heterozygous variant reported through a clinical NextGen panel was filtered out as it is a synonymous variant with no REVEL score, no MPC score, and low PhyloP and CADD scores

(phylopMam_avg = 1.097; phylopVert100_avg = 0.013; CADD_phred = 10.38; see **Table 3**), but its presence was verified in the sequencing data. After preliminary interpretation of the variants ascertained, no candidates to confirm a diagnosis were identified. The filtration criteria were subsequently relaxed to include only variants that matched at least one HPO term, however, no additional candidate variants were found. These findings suggest that further investigations like RNA-seq and further relaxation of the filtration strategy may be necessary to identify potential candidate variants for this participant.

Table 10A. Clinical information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 2 (*n* = 4). **[REDACTED]**

Table 10B. Bioinformatic information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 2 (*n* = 4). **[REDACTED]**

3.3. Case 3

3.3.1. Case Description: Case 3 is a 59-year-old female of Irish and Czech descent. She is the oldest of a large sibship of seven full siblings (four sisters, three brothers) and three half-siblings, who were reported to be healthy. She has one sister with a mildly elevated CK level who is asymptomatic (134 U/ml; normal: 28-110 IU/ml). Her parents were non-consanguineous. The maternal family history is remarkable for a first cousin once removed who is in a wheelchair for undetermined reasons. The paternal family history is unremarkable (see pedigree; **Figure 10 [REDACTED]**). The potential inheritance patterns for this case are listed in order of likelihood in **Table 11A**. No pregnancy, birth, or neonatal history was available. She reports that her symptoms began after a ligamentous knee injury at 42 years old in 2008, which required meniscal debridement. Since then, she has experienced increasing lower extremity weakness, and difficulty with stairs, running, jumping, standing from a seated position, and riding a bike. She also reports compromised balance and buckling of her left knee, numbness and tingling in her thighs, occasional night-time tingling in her toes, and numbness of her fingers and toes.

Figure 10. Pedigree for Case 3 (NM9; collected 28-Oct-2014). **[REDACTED]**

She first presented to genetics at 49 years old. Physical examination at that time described a non-dysmorphic woman with proximal lower limb weakness, particularly in the hip girdle muscles. She was unable to toe-walk or heel-walk, and no knee or ankle reflexes could be elicited. She had overriding great toes bilaterally and displayed a modified positive Gowers sign. Her gait also indicates hip stabilizer weakness, with a widened stance and knee instability, and she has reduced sensation to touch in her lower back, upper thighs, and anterior legs. Her medical history is remarkable for adolescent thoracic scoliosis, eczematoid dermatosis, and a 35-year history of tobacco smoking, although she has since quit. Over the last decade she has been unable to work due to a gradual worsening of her symptoms, with increased aching pain in her lower back and feet and the need for a walker or cane for ambulation. The HPO terms selected for this case based on clinical presentation are listed in **Table 11B**.

Table 11A. Potential inheritance pattern(s) for Case 3 in order of likelihood (most to least likely).

Inheritance Pattern	**Symbol**
De novo autosomal dominant	*De novo* AD
Autosomal recessive	AR
X-linked	XL

Table 11B. Human Phenotype Ontology (HPO) terms selected for Case 3 based on their clinical presentation.[1]

HPO Term Description	**HPO Code**
Postural instability	HP:0002172
Gait disturbance	HP:0001288

Areflexia of lower limbs	HP:0002522
Fatty replacement of skeletal muscle	HP:0012548
Abnormal muscle fiber morphology	HP:0004303
Muscle fiber atrophy	HP:0100295
Increased variability in muscle fiber diameter	HP:0003557
Nemaline bodies	HP:0003798
EMG: myopathic abnormalities	HP:0003458
Pelvic girdle muscle weakness	HP:0003749
Axial muscle weakness	HP:0003327
Proximal muscle weakness	HP:0003701
Mildly elevated creatine kinase	HP:0008180
Left ventricular hypertrophy	HP:0001712

[1]The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3.

Investigations: CK levels measured at age 49 were mildly elevated at 121 IU/ml (normal: 28-110 IU/ml), while other biochemical investigations were normal, including metabolic and Pompe disease screens. MRI of the thigh and calf muscles at age 50 showed severe bilateral fatty atrophy, notably in the quadriceps femoris, with some sparing of the vastus lateralis and medius. The biceps femoris and adductor magnus also displayed patchy fatty replacement, and the gluteus medius exhibited a core of fibrofatty atrophy. In the calves, significant fatty replacement occurred in the gastrocnemius and soleus, as well as moderate replacement in the peroneal

muscles, but the deep muscle compartments were mostly spared. MRI of the spine at age 48 indicated probable paraspinal involvement, with atrophy and fibrofatty changes, along with an unrelated low cord termination at the L2 level. EMG and NCV studies at age 47 showed low amplitude polyphasia in several muscles (left tibialis anterior, deltoid, bilateral vastus muscles) but no evidence of nerve damage. A follow-up at age 49 indicated myopathic features in motor unit potentials, as well as detection of electrosilent tissue in some lower extremity muscles indicative of fibrofatty replacement. Muscle biopsy of the quadriceps femoris at age 48 revealed chronic myopathic changes, including lobulated fibres and nemaline rods, increased fibre size variation, grouped atrophy, mild endomysial fibrosis, and fatty replacement with areas of streaking hyper-erythema, suggesting inflammation or trauma. Echocardiogram at age 49 was normal, but at age 52, it showed borderline left ventricular hypertrophy and mild heart valve issues (mild aortic stenosis, mitral and tricuspid regurgitation, trace pulmonary insufficiency).

At age 48 in 2014, a nemaline myopathy sequencing panel of *ACTA1*, *NEB*, and *TPM2* found no pathogenic variants. At age 50 in 2016, a myopathy panel of 43 genes revealed two VUSs in the *FLNC* and *NEB* genes. The VUS in *NEB* involves a heterozygous copy number gain in the triplicate region of the gene. Compound heterozygous or homozygous pathogenic variants in *NEB* are known to cause autosomal recessive nemaline myopathy. Familial screening in full siblings identified the same *NEB* variant in each of her three asymptomatic brothers and one of her sisters, with the latter having the mildly elevated creatine kinase levels. The VUS in *FLNC* is a rare heterozygous missense variant (c.577G>A; p.Ala193Thr) that has been previously associated with distal myopathy in one individual and their relatives, suggesting a potential impact on protein function (Duff et al., 2011). Pathogenic variants in *FLNC* are known to be linked to dominantly inherited cardiomyopathies, distal myopathies, or myofibrillar myopathies.

Familial testing in her asymptomatic full siblings was negative for this variant. While this variant remains classified as a VUS, the clinical team is strongly suspicious of a diagnosis of myofibrillar myopathy type 5 as her muscle biopsy findings are consistent with this condition.

3.3.3. Variant Filtration:

Variant filtration yielded four variants for further assessment, including three small variants, no CNVs, and one SV in known NMD genes that matched at least one of the selected HPO terms for this participant. The relevant bioinformatic and clinical information for the relevant variants are summarized in **Tables 12 and 13 [Table 13 REDACTED]**. While both variants reported in the previous NextGen sequencing panel were verified in the data, *NEB* and *FLNC* (c.577G>A; p.Ala193Thr), the filtering process only identified the *FLNC* variant as the *NEB* variant had a high allele frequency.

Table 12A. Clinical information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 3 (*n* = 3).

Gene symbol	Variant (cDNA; protein)	Chromosomal position	Variant type	Zygosity; Genotype	ClinVar classification	OMIM phenotype(s)	Inheritance pattern	Clinical impression	ACMG classification
*FLNC**	c.577G>A; p.Ala193Thr	chr7:128835550G>A	Exonic	Hetero-zygous; G/A	Pathogenic /Likely Pathogenic (4 Pathogenic reports, 3 Likely Pathogenic reports, 1 Uncertain Significance report, 1 report with no criteria)	- Cardiomyopathy, familial hypertrophic, 26 - Arrhythmogenic right ventricular dysplasia, familial - Cardiomyopathy, familial restrictive, 5 - Myopathy, distal, 4 - Myopathy, myofibrillar, 5	Autosomal dominant	Condition matches phenotype, inheritance pattern possible	Pathogenic (PS3, PS4, PM2, PP3, PM6, PP5)

Table 12B. Bioinformatic information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 3 (*n* = 3).

Gene symbol	Variant (cDNA; protein)	Chromosomal position	Variant effect	Zygosity; Genotype	gnomAD allele frequency	GA4K allele frequency	Predicted functional impact	Impact score metrics
*FLNC**	c.577G>A; p.Ala193Thr	chr7:128835550G>A	Missense	Hetero-zygous	0	0	Missense High, Rare0	REVEL = 0.91

*Variants previously identified by clinical genetic testing approaches for this participant.

Table 13A. Clinical information for SVs in known NMD genes that matched at least one of the selected HPO terms for Case 3 (*n* = 1). **[REDACTED]**

Table 13B. Bioinformatic information for SVs in known NMD genes that matched at least one of the selected HPO terms for Case 3 (*n* = 1). **[REDACTED]**

3.3.4. FLNC

After review with the research team, a heterozygous missense variant, c.577G>A (p.Ala193Thr) in the *FLNC* gene was of greatest interest, which was previously detected through a NextGen clinical panel and classified as a ***Variant of Uncertain Significance*** at that time. Pathogenic variants in *FLNC* are associated with several autosomal dominant cardiac disorders (i.e., familial cardiomyopathies and ventricular dysplasias), as well as autosomal dominant distal myopathy type 4 and myofibrillar myopathy type 5 (OMIM #: 617047, 614065, 609524; Amberger et al., 2015). The latter closely aligns with the phenotype and inheritance pattern observed in the participant. Visual confirmation of this variant was completed using IGV (**Figure 11**) and the Franklin by Genoox tool (**Figure 12**; see **Supplementary Figure 4** for additional visualization using the UCSC Genome Browser).

With integration of the evidence below, and in accordance with the ACMG criteria, this variant was upgraded from a VUS and is now predicted to be ***Pathogenic* (II)**. Criteria applied includes: well-established functional studies supportive of a damaging effect (**PS3**), prevalence in affected individuals is significantly increased compared with controls (**PS4**), absent from controls (**PM2**), assumed *de novo* without confirmation of paternity/maternity (**PM6**), multiple lines of computational evidence support a deleterious effect (**PP3**), and a reputable source recently reports variant as pathogenic (**PP5**).

A functional study suggests this variant may impact the protein product via actin-binding dysregulation or a similar mechanism (Duff et al., 2011; **PS3**). This variant is absent from the gnomAD database but has been reported as ***Pathogenic/Likely Pathogenic*** in the ClinVar database based on nine submissions (four *Pathogenic*, three *Likely Pathogenic*, one *Uncertain Significance*, and one with no criteria provided; Accession: VCV000029592.20; Landrum et al.,

2014; **PS4**, **PM2**, **PP5**). Family studies conducted indicate that this variant affects only this single proband within a large sibship, where all other siblings tested negative for the variant, suggesting it likely arose *de novo* (**PM6**). *In-silico* software predicts a highly deleterious effect (SIFT: 0; PolyPhen: 0.999; MA: 3.685; MT: 1; MPC: 2.152; REVEL: 0.909; Alpha Missense: 0.7975; CADD: 26) and intolerance of the *FLNC* gene to loss-of-function variants (pLI = 1; o/e LoF = 0.25; Decipher Haploinsufficiency = 21.56; **PP3**).

Therefore, with reclassification based on the new evidence from the literature and family studies, this variant in *FLNC* is now a strong candidate for contributing to the phenotype of this individual. Based on clinical best practice, validation of the variant in the clinical laboratory is required prior to making an official diagnosis.

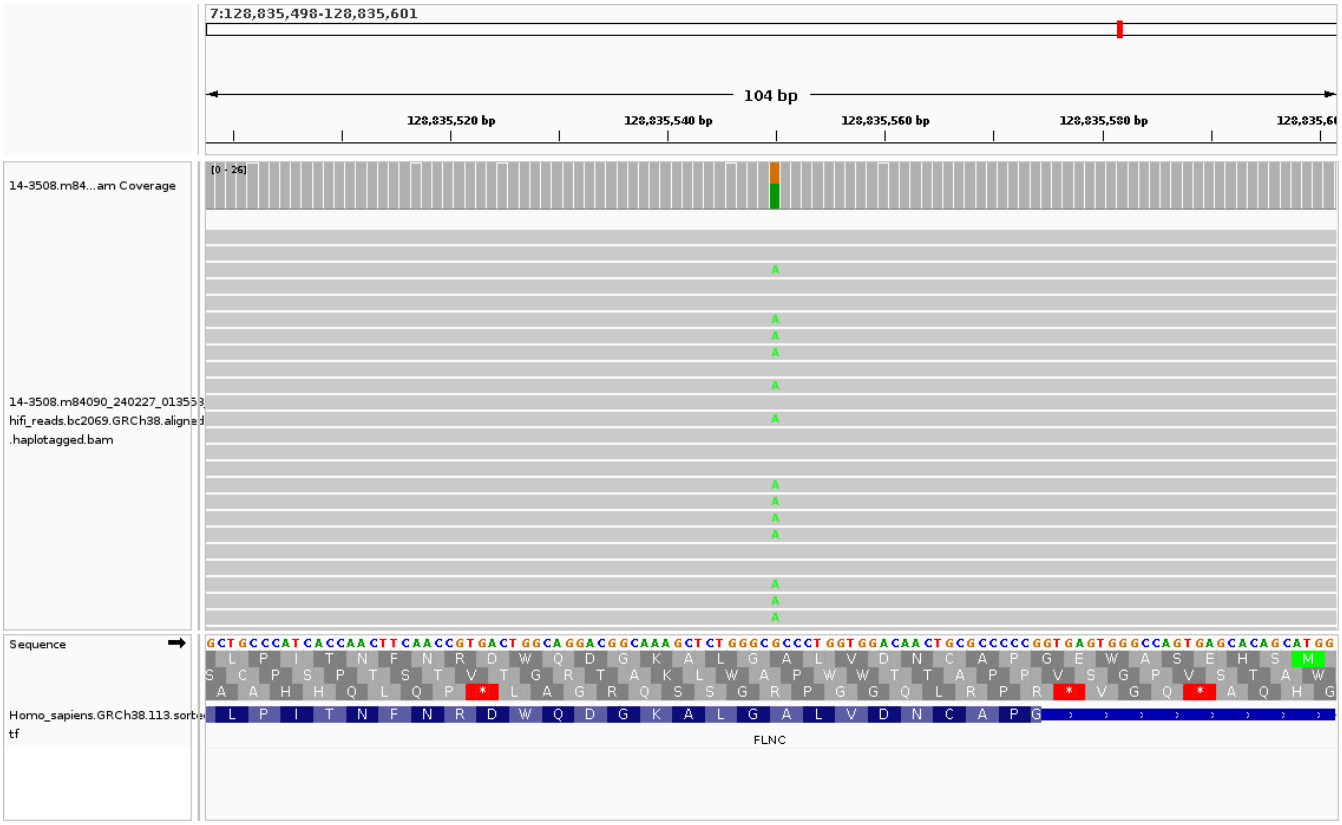


Figure 11. Integrative Genomics Viewer (IGV) snapshot of the sequencing read alignments for Case 3, showing the presence of a missense variant in the *FLNC* gene (c.577G>A; p.Ala193Thr) as indicated by the central green line.

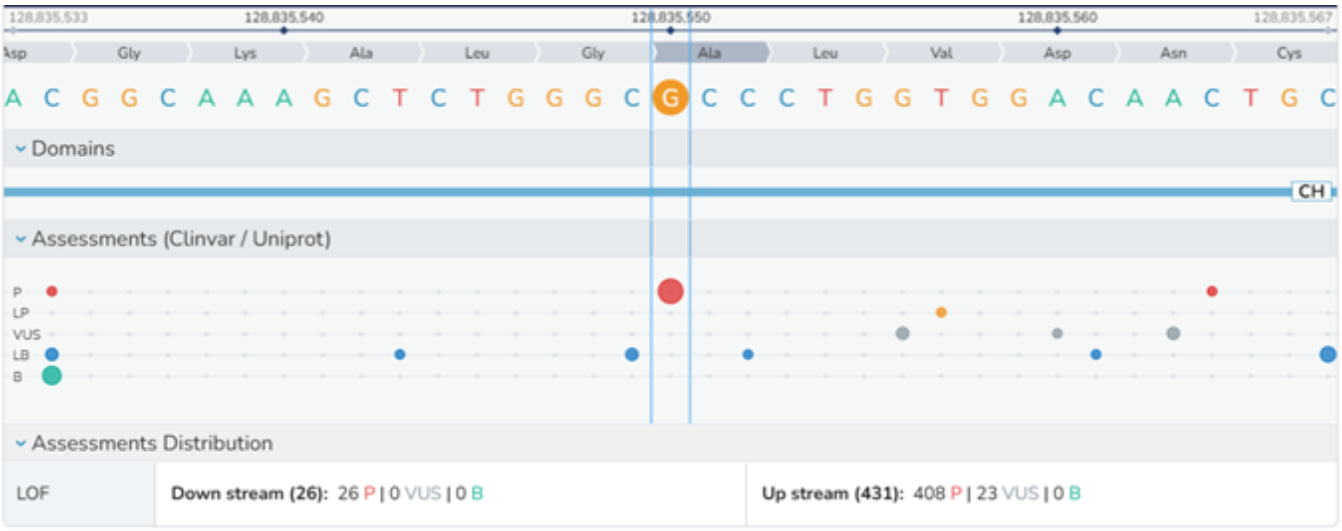


Figure 12. Franklin by Genoox snapshot visualizing the genomic regions affected by the missense variant in the *FLNC* gene (chrX:31,730,001-32,076,000) as identified in Case 3 (red dot). The boundaries of the variant are indicated by blue vertical lines. Dots under 'Assessments' indicate nearby variants, colour-coded by ACMG classification (red = Pathogenic, orange = Likely Pathogenic, grey = VUS, blue = Likely Benign, green = Benign).

3.4. & 3.5. Case 4 and Case 5

Case 4 and Case 5 consist of an affected sibling pair of Dutch Mennonite/Scottish/Irish/English/Spanish descent. Their parents are non-consanguineous. Their mother has von Willebrand's disease and attention-deficit/hyperactivity disorder (ADHD) and their father has autosomal dominant dilated cardiomyopathy due to a heterozygous *TTN* variant. The family history is remarkable for a younger brother with a history of stuttering, fatigue, mild gross motor delay, and complaints of leg pain (not yet formally evaluated). In addition, there is a history of ADHD, potential bipolar disorder, febrile seizures, and speech delay in their older brother, and a history of obsessive-compulsive disorder (OCD) and possible celiac disease in their older sister. Familial studies for the *TTN* variant were conducted on both parents and four

out of the six siblings (excluding the eldest and youngest brothers) in 2018 after the variant was identified in the father as described in **Figure 13 [REDACTED]**. The paternal family history is remarkable for congenital heart disease and an uncle with ADHD but is otherwise unknown due to lack of contact. The maternal family history is remarkable for depression, OCD, a great aunt with systemic lupus erythematosus, an uncle with ADHD, and a cousin with cerebral palsy (see pedigree; **Figure 13 [REDACTED]**). The potential inheritance patterns for both cases (Case 4 and Case 5) are listed in order of likelihood in **Table 14A**.

Figure 13. Pedigree for Case 4 and Case 5 (NM3 and NM6; collected Jun-2016).
[REDACTED]

3.4. Case 4

3.4.1. Case Description: The elder of the two siblings is Case 4, a 20-year-old male who was born at term following an unremarkable pregnancy history. He was induced at term and required suction for shoulder dystocia. He was treated with phototherapy for jaundice within the first week. He had a birth weight of 4.520 kg, a length of 57.5 cm, and a head circumference of 36 cm. He first presented to medical attention at 18 months of age with concerns of lethargy, veil-type rashes, and mild jaundice around his mouth and nose. Investigations revealed persistent elevation of serum liver enzymes aspartate aminotransferase (AST) and alanine aminotransferase (ALT); he was assessed by a gastroenterologist at age 3 who performed a non-diagnostic liver biopsy. Also at age 3, his parents reported a two-year history of intermittent exercise- and temperature-exacerbated migratory joint and arm pain with fatigue and frequent falls. In retrospect, his parents reflected he also had some differences in early walking compared to his siblings (a waddle and did not bend his knees).

He was first assessed by genetics at 4 years old and was remarked to be non-dysmorphic with a normal physical examination. He was described to have some concerns with developmental delays including speech articulation, and delays in fine and gross motor skills. In addition, he was described to have frequent illnesses, poor weight gain, and episodic diarrhea and nausea. At a follow-up examination by genetics at age 5, he was found to have evidence of mild proximal weakness particularly affecting the shoulder abductor and elbow flexor, with an abnormal stiff gait during periods of weakness. Mild weakness of the hip and knee flexors was also noted on additional examination at the same age. At ages 7 and 8 he continued to have recurrent rashes and hives, fevers, malaise and myalgia of unknown etiology, which occasionally required hospitalization for management. He was started on a trial of ribose supplementation and

naproxen for pain management, neither of which appeared to be effective. At age 10, he started on creatinine and noted some improvements in pain levels after activity. A follow-up examination at age 13 did not find evidence of objective muscle weakness or myoglobinuria. However, at his most recent encounter at 16 years old, he was found to experience periods of weakness and temperature sensitivity with occasional myoglobinuria. His crossed adductor reflex was also found to be absent, and he displayed some joint hyperflexibility. He is currently able to run and walk short distances and plays football but needs to use a wheelchair when he experiences more severe myalgias. He has been attending public school with educational assistance since Grade 4, before which he was home-schooled. He has longstanding difficulties with reading and consistently remains below grade level. His past medical history is also remarkable for eczema, mild lumbar scoliosis, sleep difficulties, and requiring glasses. The HPO terms selected for this case based on clinical presentation are listed in **Table 14B**.

Table 14A. Potential inheritance pattern(s) for Case 4 and Case 5 in order of likelihood (most to least likely).

Inheritance Pattern	**Symbol**
Autosomal recessive	AR
X-linked	XL
De novo or germline mosaicism autosomal dominant	*De novo*/GM AD

Table 14B. Human Phenotype Ontology (HPO) terms selected for Case 4 and Case 5 based on their clinical presentation.[1] The presence of an 'X' indicates that the term is applicable to that case.

HPO Term Description	**HPO Code**	**Case 4**	**Case 5**
Exercise intolerance	HP:0000118	X	X

Exercise-induced myalgia	HP:0003738	X	X
Abnormal muscle fiber morphology	HP:0004303	X	
Reduced muscle fiber alpha dystroglycan	HP:0030099	X	
Muscle stiffness	HP:0003552	X	
Proximal muscle weakness	HP:0003701	X	X
Hip flexor weakness	HP:0012515	X	
Neck flexor weakness	HP:0003722		X
Pelvic girdle muscle weakness	HP:0003749		X
Cold-induced muscle cramps	HP:0003449	X	X
Exercise-induced leg cramps	HP:0008991	X	X
Motor delay	HP:0001270	X	
Easy fatiguability	HP:0003388	X	X
Areflexia of lower limbs	HP:0002522	X	
Pes planus	HP:0001763		X
Gait disturbance	HP:0001288	X	X
Pronunciation difficulties	HP:5200237	X	X
Dyslexia	HP:0010522	X	
Elevated circulating creatine kinase concentration	HP:0003236	X	X
Reduced muscle myoadenylate deaminase activity	HP:0034529	X	X
Myoglobinuria	HP:0002913	X	

[1]The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3.

3.4.2. Investigations: CK levels have remained consistently elevated since first measurement at age 3, ranging from 283 U/L at age 12 to 5268 U/L at age 8 (normal: <350 U/L). His serum liver enzymes (ALT, AST) have also been persistently elevated since first measurement, ranging from

85 to 103 U/L for AST (normal: 10-50 U/L) and 34 to 107 for ALT (normal: 4-30 U/L). Other remarkable biochemical investigations include elevated serum myoglobin at age 8 (88 ug/L; normal: <50 ug/L), along with metabolic tests that were largely non-diagnostic. MRI of the brain at age 4 and abdominal ultrasound at age 2 were both normal. MRI of the thighs and calves at ages 9 and 11 were normal. No EMG or NCV studies were performed. A percutaneous liver biopsy at age 3 revealed somewhat swollen hepatocytes with scattered glycogenated nuclei and possibly increased hepatocellular glycogen in the cytoplasm. A muscle biopsy of the right quadriceps femoris at age 4 revealed decreased myoadenylate deaminase activity, two small foci of inflammatory cell aggregates, a slight increase in the portion of fibers with internal nuclei, and rare atrophic and degenerating fibers. Recent re-staining also showed an abnormal patchy reduction in alpha-dystroglycan.

He had a normal karyotype. Sequencing of *AMPD1* following the muscle biopsy results detected two compound heterozygous variants, confirming a definite molecular diagnosis of AMPD1 deficiency. He was enrolled in a research study offering whole exome sequencing, which did not detect any causative variants.

3.5. Case 5

3.5.1. Case Description: The younger of the two siblings is Case 5, a 16-year-old male who was born at term following an unremarkable pregnancy history. Birth weight was 3.629 kg. He was jaundiced requiring phototherapy in the neonatal period. He first presented to genetics at 3 years 5 months of age with a several-month history of cramping and myalgia associated with exercise and temperature changes, a stiff walk, reduced exercise tolerability, fatigue, and a history of temperature sensitivity as a toddler. At 9 years old, he reported experiencing occasional episodes

of exertional chest pain, but no cardiac abnormalities were found. At that time, hyperflexibility in his joints and pes planus with a mild valgus deformity at the ankle were noted on examination.

He has a history of speech delay and difficulties with articulation and is currently home-schooled in a regular academic program. He requires reading glasses, and his past medical history is remarkable for a left inguinal hernia and hydrocele surgery at age 2, ADHD, headaches, celiac disease, and suspected autoimmune overlap disease. He carries the familial *TTN* cardiomyopathy variant and is undergoing cardiac screening.

At his last clinical encounter at 13 years old, he was found to have a positive Trendelenburg gait, slight weakness in the neck flexors, hip flexors and abductors, and serratus anterior muscles. He reports that his symptoms continue to occur sporadically, often worsened in high temperatures. However, they have remained stable and are generally less severe than his brother. He can complete activities such as a short run or bike without requiring a recovery day but does experience decreased sweating and occasional myalgia after such activities. In cold weather, he also describes leg stiffness and an abnormal gait. The HPO terms selected for this case based on clinical presentation are listed in **Table 14B**.

3.5.2. Investigations: CK levels have remained consistently elevated, ranging from 901-1968 U/L (normal: <350 U/L). Two mild elevations in serum liver enzymes AST (53, 143 U/L [normal: 10-50 U/L] and ALT (55, 76 U/L; [normal: 4-30 U/L]) were reported, which normalized at age 11. Metabolic investigations were normal. MRI of the brain and calf muscles at 8 years old were normal. No muscle biopsies, EMG, or NCV studies were performed. Testing for the familial *AMPD1* gene variants detected in his brother indicated compound heterozygosity for two variants, confirming a definite molecular diagnosis of AMPD1

deficiency. No other genetic testing relating to his presenting symptoms has been pursued to date for this individual.

3.5.3. Variant Filtration:

Variant filtration for **Case 4** yielded four variants for further assessment, including two small variants, one CNV, and one SV in known NMD genes that matched at least one of the selected HPO terms for this participant. The relevant bioinformatic and clinical information for the relevant variants are summarized in **Tables 15, 16, and 17 [Tables 15 and 17 REDACTED]**. The presence of compound heterozygous *AMPD1* variants previously detected by Sanger sequencing were verified in the sequencing data, but filtered out as each variant had a high population allele frequency.

Variant filtration for **Case 5** yielded four variants for further assessment, including three small variants, one CNV, and no SVs in known NMD genes that matched at least one of the selected HPO terms for this participant. The relevant bioinformatic and clinical information for the relevant variants are summarized in **Tables 18** and **19 [Table 18 REDACTED]**. The presence of compound heterozygous *AMPD1* variants previously detected by Sanger sequencing and the familial *TTN* variant were verified in the sequencing data, however, the *AMPD1* variants were filtered out as each variant had a high population allele frequency. See **Section 3.5.4** for the collective interpretation of variants for Cases 4 and 5.

Table 15A. Clinical information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 4 (*n* = 2). **[REDACTED]**

Table 15B. Bioinformatic information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 4 (*n* = 2). **[REDACTED]**

Table 16A. Clinical information for CNVs in known NMD genes that matched at least one of the selected HPO terms for Case 4 (*n* = 1).

Gene symbol(s)	Chromosomal position	Variant type	Zygosity	ClinVar classification	OMIM phenotype(s)	Inheritance pattern	Clinical impression	ACMG classification
DMD†	chrX:31730001 -32076000	Deletion	Hemi-zygous	Not reported	- Duchenne muscular dystrophy - Becker muscular dystrophy - Cardiomyopathy, dilated, 3B	X-linked recessive	Condition matches phenotype, inheritance pattern possible	Pathogenic (PVS1, PP1, PM2, PP3, PP4)

Table 16B. Bioinformatic information for CNVs in known NMD genes that matched at least one of the selected HPO terms for Case 4 (n = 1).

Gene symbol(s)	Chromosomal position	Variant type	Zygosity	Variant length (bp)	Cytogenic location	Internal database allele frequency (50% reciprocal overlap)	Internal Children's Mercy Hospital allele frequency (50% reciprocal overlap)
DMD†	chrX:31730001 -32076000	Deletion	Hemi-zygous	346000	Xp21.1	0.02	0

† = shared variants between Case 4 and Case 5 (full brothers)

Table 17A. Clinical information for SVs in known NMD genes that matched at least one of the selected HPO terms for Case 4 (*n* = 1). **[REDACTED]**

Table 17B. Bioinformatic information for SVs in known NMD genes that matched at least one of the selected HPO terms for Case 4 (*n* = 1). **[REDACTED]**

Table 18A. Clinical information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 5 (*n* = 3). **[REDACTED]**

Table 18B. Bioinformatic information for small variants in known NMD genes that matched at least one of the selected HPO terms for Case 5 (*n* = 3). **[REDACTED]**

Table 19A. Clinical information for CNVs in known NMD genes that matched at least one of the selected HPO terms for Case 5 (*n* = 1).

Gene symbol(s)	Chromosomal position	Variant type	Zygosity (Hom, Het)	ClinVar classification	OMIM phenotype(s)	Inheritance pattern	Clinical impression	ACMG classification
DMD†	chrX:31730001 -32076000	Deletion	Hemi-zygous	Not reported	- Duchenne muscular dystrophy - Becker muscular dystrophy - Cardiomyopathy, dilated, 3B	X-linked recessive	Condition matches phenotype, inheritance pattern possible	Pathogenic (PVS1, PP1, PM2, PP3, PP4)

Table 19B. Bioinformatic information for CNVs in known NMD genes that matched at least one of the selected HPO terms for Case 5 (n = 1).

Gene symbol(s)	Chromosomal position	Variant type	Zygosity (Hom, Het)	Variant length (bp)	Cytogenic location	Internal database allele frequency (50% reciprocal overlap)	Internal Children's Mercy Hospital allele frequency (50% reciprocal overlap)
DMD†	chrX:31730001 -32076000	Deletion	Hemi-zygous	346000	Xp21.1	0.02	0

† = shared variants between Case 4 and Case 5 (full brothers)

3.5.4. DMD

After review with the research team, a hemizygous *DMD* variant (chrX:31,730,001-32,076,000) identified in both siblings was of greatest interest. This variant is a large (346 Kb) in-frame deletion spanning exons 45-52 on Xp21.1 with a breakpoint occurring between exons 44 and 45 within intron 44 and extending into exon 52 (see **Figure 14** for visualization by IGV and **Figure 15** for additional visualization via the UCSC Genome Browser). Further investigation with the Franklin by Genoox tool revealed several other deletions classified as *Pathogenic/Likely Pathogenic* within the range of this deletion based on curated submissions (see **Figure 16**). Pathogenic variants in *DMD* are known to cause X-linked dystrophinopathies (Duchenne/Becker muscular dystrophy), as well as dilated cardiomyopathy 3B (OMIM #: 300376, 302045, 310200; Amberger et al., 2015).

Strictly following the ACMG criteria, this variant is predicted to be ***Pathogenic*** **(Ia)** using the following evidence: null variant (multiexon deletion) in a gene where loss-of-function is a known mechanism of disease (**PVS1**), co-segregation with disease in multiple affected family members (**PP1**), absent from controls (**PM2**), multiple lines of computational evidence support a deleterious effect (**PP3**), and the phenotype or family history of the participants is highly specific for a disease with a single genetic etiology (**PP4**).

Given the strong association of loss-of-function variants in *DMD* with dystrophinopathies and the well-established pathogenicity of multiexon deletions in the *DMD* gene, this CNV is likely to contribute to disease (**PVS1**). However, as the reading frame is maintained, the resulting dystrophin protein may retain partial function and lead to a milder phenotype such as Becker muscular dystrophy, which is consistent with the phenotype and inheritance pattern of the participants. This variant has not been previously reported in the gnomAD database; however,

gnomAD reports that the *DMD* gene is highly intolerant to loss-of-function variants (pLI = 1, o/e LoF = 0.15, ClinGen haploinsufficiency = 3; Karczewski et al., 2020; **PM2**, **PP3**). Clinical assessment with an elevated CK level and an X-linked inheritance pattern also aligns with this monogenic condition (**PP4**).

These findings suggest that the identified variant in the *DMD* gene is a strong candidate for contributing to the phenotype of these participants. Based on clinical best practice, validation of the variant in a clinical laboratory is required prior to making an official diagnosis.



Figure 14. Integrative Genomics Viewer (IGV) snapshot of the sequencing read alignments for Cases 4 and 5 confirming the presence of a large deletion in the *DMD* gene (chrX:31,730,001-32,076,000) as indicated by the white space.

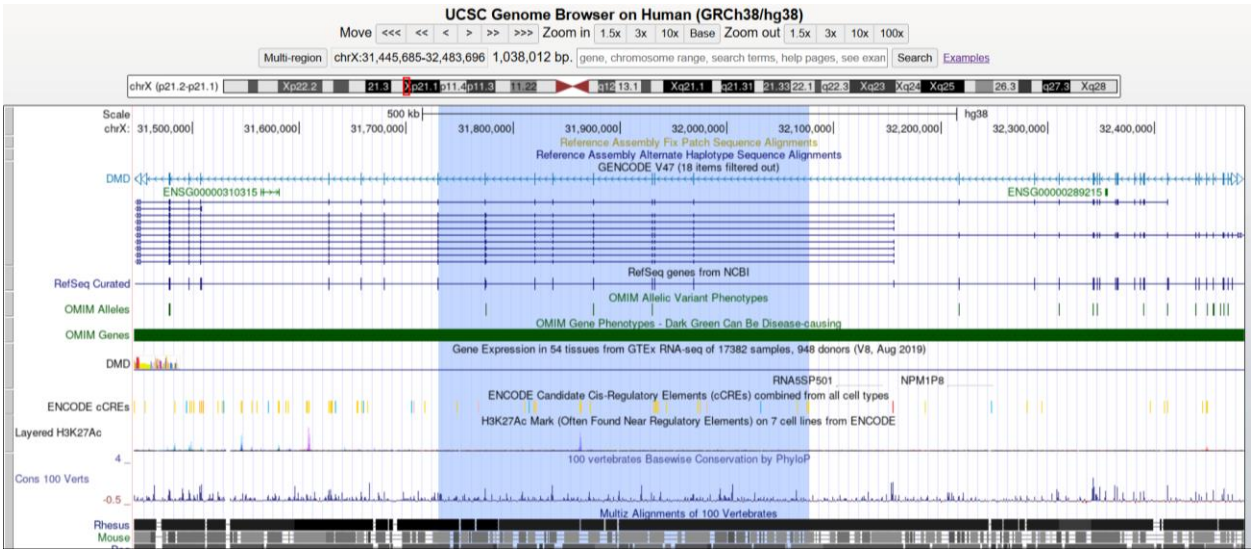


Figure 15. UCSC Genome Browser snapshot visualizing the genomic regions affected by the large deletion in the *DMD* gene (chrX:31,730,001-32,076,000) as identified in Case 4 and Case 5. The range of the deletion is indicated in blue.

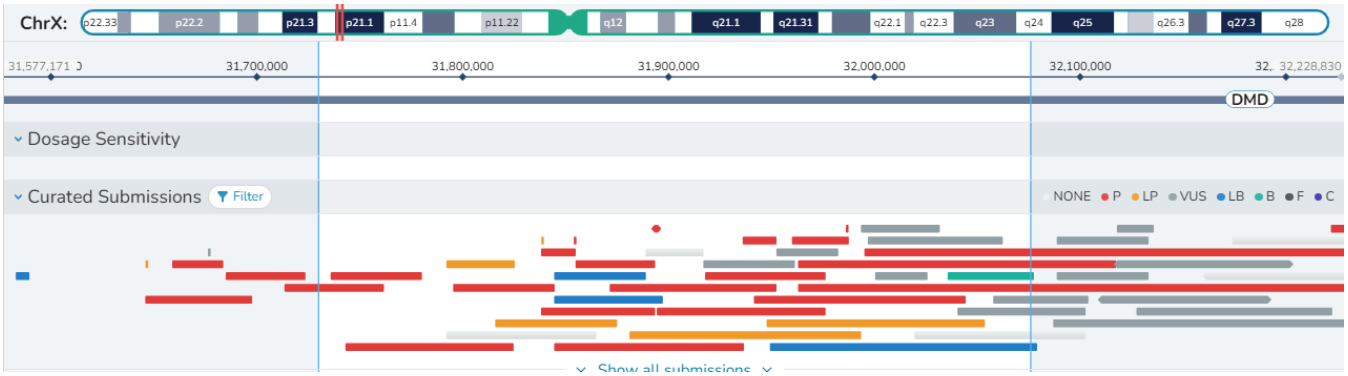


Figure 16. Franklin by Genoox snapshot visualizing the genomic regions affected by the large deletion in the *DMD* gene (chrX:31,730,001-32,076,000) as identified in Case 4 and Case 5. The boundaries of the deletion are indicated by blue vertical lines. Horizontal bars under 'Curated Submissions' indicate other variants within this genomic region, colour-coded by ACMG classification (red = *Pathogenic*, orange = *Likely Pathogenic*, grey = *Variant of Uncertain Significance*, blue = *Likely Benign*, green = *Benign*).

4. DISCUSSION

This study aimed to evaluate the performance of LR-WGS with the latest LRS platform, the PacBio Revio, for variant identification in a small cohort of undiagnosed individuals with suspected NMDs. This pilot study represents an initial step in evaluating the effectiveness of LR-

WGS along with improved bioinformatic tools for NMD diagnostics compared to traditional short-read approaches. Moving forward, expanding the analysis to include additional participants (22 individuals originally recruited and sequenced; five analyzed in this study) will provide a more comprehensive assessment of the diagnostic yield and allow for direct comparisons with SRS methods. Increasing the sample size will also help determine the broader applicability of this approach in identifying pathogenic variants across diverse cases in regard to both disease etiology and ancestry.

LR-WGS, in conjunction with enhanced variant calling and annotation tools, identified candidate variants of interest in three out of four families analyzed (three singletons, one duo; 75%). To qualify as a strong candidate following established approaches widely used in studies of Mendelian disorders, a variant had to be rare in the population, occur in a known disease-associated gene linked to the participant's phenotype, match the expected inheritance pattern, and have evidence of pathogenicity in accordance with established criteria. Rarity, evidence of pathogenicity from *in-silico* prediction tools, and known gene-disease association were incorporated into the variant filtration steps as described in **Section 2.5**, while assessment of phenotypic consistency, inheritance pattern, and evidence of pathogenicity was based on manual clinical judgement. Among these four participants, three strong candidate variants were identified in the *TTN*, *FLNC*, and *DMD* genes, which are discussed in detail below.

4.1. Discussion of candidate variants

4.1.1. Case 1

The strongest candidate for Case 1 is a homozygous intronic variant in the *TTN* gene (c.34291+3A>G, chr2:178677618T>C), which is predicted *in-silico* as splice-disrupting. The

TTN gene encodes titin, a key structural protein in muscle, and pathogenic variants in this gene are associated with a range of neuromuscular and cardiac conditions depending on the affected transcript (Bang et al., 2001; Gohlke et al., 2024). While this variant has not been previously described, LRS approaches have been successful in identifying other novel variants in the *TTN* gene (Perrin et al., 2022).

This *TTN* variant is currently classified as a VUS as its functional impact has not been established beyond *in-silico* predictions. Unfortunately, as many intronic *TTN* variants lack sufficient mechanistic evidence, most remain classified as a VUS (Campuzano et al., 2015). While a nearby variant, c.34291+19G>A (chr2:178677602C>T), is reported as ***Likely Benign*** in ClinVar (Accession: VCV002923747.2; Landrum et al., 2014), this variant is 19 base pairs away from the nearest splice site and as such, has less evidence to suggest it would impact splicing as compared to the variant found in Case 1.

The splice variant identified in *TTN* qualifies as a so-called 'hot' VUS due to strong supporting evidence but does not yet meet the ACMG criteria for classification as pathogenic (Joynt et al., 2021; Richards et al., 2015). Additional data, such as evidence from functional studies or familial segregation analysis could lead to its reclassification (Joynt et al., 2021). As RNA-seq analysis for this variant is currently underway for Case 1, this could provide functional evidence for whether the variant causes alterations in titin isoform expression due to a splicing defect. To date, RNA has been successfully extracted, and library preparation is in progress at SickKids TCAG. The addition of this supporting evidence could allow for the application of the **PS3** criteria (well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect) following ACMG guidelines, which would move the variant classification from a VUS to ***Likely Pathogenic*** **(II)**. While outside the scope of this study, further supporting experiments,

such as expressing the variant in a relevant model organism to observe the resulting phenotype, could also provide evidence of pathogenicity. Additionally, segregation analysis with familial samples, particularly for siblings, is important to determine if other family members are affected and whether this variant co-segregates with disease, which would support applying **PP1** under ACMG guidelines and upgrade the classification to ***Likely Pathogenic* (II)**. Thus, these next steps are crucial in determining whether this variant is causative for Case 1. As this *TTN* variant was validated using LRS in this pilot study after prior identification by short-read methods, this case supports the accuracy of long reads in detecting intronic splice-altering variants. As mentioned, further functional studies will be essential in assessing the true impact of this variant on splicing and its potential role in disease, providing a better understanding of its clinical significance moving forward.

4.1.2. Case 3

The strongest candidate for Case 3 is a heterozygous missense variant in the *FLNC* gene (c.577G>A, p.Ala193Thr) classified as ***Pathogenic* (II)**. This variant was also identified by prior clinical genetic testing, however, evidence at that time classified it as a VUS.

Since that time, additional evidence from the literature, family studies, and population databases can now be integrated to support the pathogenicity of this *FLNC* variant to cause a heterogeneous myopathic phenotype, contributing to its reclassification as *Pathogenic/Likely Pathogenic*. Other cases of this *FLNC* variant now reported in the literature include a study of three affected individuals (mother/daughter duo and a single proband) who presented with progressively worsening proximal and limb-girdle weakness between the fourth and fifth decade of life, reduced sensation and reflexes in the lower limbs, and normal to mildly elevated CK

levels, progressing to distal and respiratory muscle involvement (van den Bogaart et al., 2017). NCV studies were normal in all cases and muscle MRI showed fatty degeneration of several lower limb muscles in the daughter and single proband. For the daughter, a muscle biopsy revealed minor abnormalities, including hypertrophic and atrophic fibers, fiber type grouping, and a core-like appearance, with one unremarkable EMG and a follow-up describing mildly myopathic changes (van den Bogaart et al., 2017). In addition, in an affected Italian family, four individuals with the *FLNC* variant (mother and three out of her four children - two sisters, one brother) are described to present with distal myopathy and thenar muscle weakness in the third decade of life, which progressed to posterior weakness, and cardiomyopathy in two individuals (Duff et al., 2011). In addition, Previtali et al. (2019) describes affected Italian siblings (brother and sister) with the variant who presented with a proximal-distal myofibrillar myopathy phenotype in the third decade of life, with no respiratory or cardiac involvement, a normal to mildly elevated CK, and involvement of the central nervous system (CNS) with cerebellar ataxia and pyrimidal signs (Previtali et al., 2019).

As our Case 3 presented with progressive proximal lower limb weakness, myopathic features, fatty atrophy, and a lack of respiratory involvement at age 49, this aligns with reported disease presentation between the third to fifth decade of life in the literature (Duff et al., 2011; Previtali et al., 2019; van den Bogaart et al., 2017). The late-onset cardiac abnormalities in this participant also aligns with the cardiomyopathy phenotype described by Duff et al. (2011). While the cerebellar ataxia described by Previtali et al. is unlike the sensory deficits observed in our participant, which are unusual features not previously reported in *FLNC* myopathies, transcripts of FLNC have been previously detected in the CNS, so it is possible this finding could be either

part of the phenotype or the result of a coexisting condition (Hawrylycz et al., 2012; Previtali et al., 2017).

Functional studies were conducted by Duff et al. (2011) to study the effect of this variant in the binding domain. Transient expression of the mutated FLNC protein in transfected cells resulted in small protein aggregates, suggesting a potential disease mechanism involving abnormal protein accumulation or actin-binding dysregulation (Duff et al., 2011). However, no protein accumulations were found in the muscle biopsies by immunohistochemical staining, which could indicate the existence of compensatory mechanisms preventing their formation *in vivo* (Duff et al., 2011). Therefore, while there is some evidence that the variant affects protein aggregation and actin binding affinity, the muscle biopsy findings necessitate further validation (Duff et al., 2011). As such, the next steps for Case 3 could include performing additional functional studies (i.e., RNA-seq, which is currently in progress) to provide support to the conclusions made by Duff et al. in relation to the functional aspects of the variant relevant to the disease pathology. Overall, this case illustrates the utility of periodic variant re-analysis as new evidence becomes available, as the reclassification of this variant in *FLNC* was made possible by both family studies and recent publications describing new cases and functional evidence. This reinforces the importance of continually integrating new data into the diagnostic process to ensure accurate and up-to-date variant interpretation, as it may lead to a definitive diagnosis for the affected individual.

4.1.3. Cases 4 and 5

In Cases 4 and 5 (full brothers), a hemizygous deletion affecting exons 45-52 of the *DMD* gene (chrX:31,730,001-32,076,000) was identified and classified as ***Pathogenic*** **(Ia)** according to

ACMG guidelines. Although this exact variant has not been previously documented, multiexonic deletions represent a major cause of *DMD*-related dystrophinopathies (Bladen et al., 2015; Geng et al., 2021). These are most prevalent within two mutational hotspots: exons 2-20 and exons 45-55, with the latter comprising over half of all reported deletions for *DMD*-related dystrophinopathies (Bladen et al., 2015; Geng et al., 2021). The deletion identified in Cases 4 and 5 was found to have a breakpoint within intron 44, the largest intron in the *DMD* gene, which has been identified as a frequent deletion breakpoint site (Geng et al., 2021; Tong et al., 2020). Geng et al. (2021) used LRS methods to analyze 22 dystrophinopathy patients with breakpoints in intron 44, and found that specific genomic features, such as a high density of repetitive elements (e.g., LINEs, Alu sequences) and secondary DNA structures (e.g., hairpin loops), likely contribute to the instability that predisposes this region to large deletions. As both brothers are affected, suggesting maternal inheritance rather than a *de novo* variant, the hypothesis of instability would likely not be relevant for these brothers. However, as the high proportion of repetitive sequences in intron 44 can impede detection using SRS methods, this may be an explanation why the deletion was not identified in previous exome sequencing in another research study for Case 4, demonstrating the limitations of SRS in accurately capturing certain CNVs (Geng et al., 2021). In line with Geng et al., and supported by several other studies that have also successfully identified novel *DMD* variants using LRS approaches, our study further supports the effectiveness of LRS for detecting novel *DMD* deletions beginning in intron 44, reinforcing its potential clinical utility in diagnosing NMDs (Bruels et al., 2022; Chin et al., 2021; Geng et al., 2023; Owusu & Savarese, 2023; Xie et al., 2020). Thus, this case supports the added value of LRS in detecting complex variants that may be missed by traditional approaches,

ultimately improving the diagnostic accuracy for *DMD*-related dystrophinopathies and other NMDs.

The next steps for Cases 4 and 5 involve conducting Sanger sequencing to confirm the presence of the deletion and verify the breakpoint locations at higher resolution, ensuring the accuracy of the variant call before reporting the results. Given this is an X-linked condition with variable penetrance in females, after clinical validation of the variant, familial studies can be offered to screen female relatives, including the mother and sisters. Female carriers of pathogenic *DMD* variants can manifest disease symptoms, including muscle weakness, gait abnormalities, myalgia, and cardiac issues (Cho & Choi, 2013). In addition, this information could be important for risk assessment and family planning.

4.1.4. Summary

Two candidate variants were identified that were previously reported in the participant charts from prior genetic testing using short reads: *TTN* (c.34291+3A>G, chr2:178677618T>C) and *FLNC* (c.577G>A, p.Ala193Thr). The *TTN* variant was detected through clinical exome sequencing, while the *FLNC* variant was identified using a NextGen panel, demonstrating the comparable effectiveness of long reads in detecting certain genetic changes. While the *TTN* variant remains classified as a “hot” VUS, as additional studies (e.g., RNA-seq) are needed to assess its functional impact, additional evidence has since emerged for the *FLNC* variant leading to its reclassification as pathogenic. This finding underscores the potential value of re-analyzing sequencing data for undiagnosed cases, as new information can provide further insights into variant pathogenicity and classification.

It is also important to note that a stringent filtering strategy could limit the detection of causative variants, as some variants identified by prior genetic testing approaches were filtered out due to factors such as a high population allele frequency (i.e., *NEB* variant in Case 3, *AMPD1* variants in Cases 4 and 5), lack of classification as deleterious based on *in-silico* scores (i.e., *RYR1* variant in Case 1), or lacking *in-silico* deleteriousness scores as a synonymous variant (i.e., *ACTA1* variant in Case 2). While these variants would not have been considered strong candidates in this study, this highlights the need for careful consideration of the filtering criteria in future analyses as some variants with potential clinical relevance may still be excluded. Further, since our filtration strategy prioritized known NMD-associated genes and computational predictions of deleteriousness, if the individual has a condition caused by a novel gene, this would not be identified with the current approach due to a lack of annotated information. To address this, refining or relaxing variant prioritization strategies could be beneficial in unresolved cases. In addition, while prediction tools are valuable in evaluating pathogenicity of genetic variants and were included in the filtering strategy, they are insufficient on their own for providing a classification as they may produce false positives and are not always reliable in predicting the biological consequences of all variant types. To fully assess variant pathogenicity, these predictions must therefore be complemented by experimental validation, familial segregation data, or other supporting information. Furthermore, if the disease of the participant follows non-Mendelian mechanisms that were not fully assessed in this analysis, such as somatic mosaicism or mitochondrial DNA variants, the causal variant could also be missed. Thus, continuing to refine the bioinformatic approaches could enhance detection of candidate variants and provide further insights for individuals without any candidates identified, such as Case 2.

4.2. Limitations

As LRS has only recently become affordable, technology-matched population data for interpreting LRS is limited by factors such as incomplete representation of diverse populations and the lack of large-scale, well-annotated datasets for certain variant types, as most genetic databases (e.g., gnomAD) primarily rely on data generated from SRS. For small variants, more information is available in gnomAD v4.0, which includes exome sequencing data from >730,000 individuals and genome sequencing data from >76,000 individuals (Broad Institute, 2023). This large sample size provides more reliable frequency data for common variants (i.e., small variants) from years of research and data collection, making it a widely accepted reference database for genetic variation in the general population (Broad Institute, 2023). However, as gnomAD continues to rely on short-read data, long-read data from the GA4K database was used as a supplement. This is currently the only available long-read dataset, which comes from a relatively small cohort, demonstrating the need for more expanded LRS reference datasets (Pellecchia, 2024b). Additionally, there is often limited information available for certain variants, such as CNVs and SVs, compared to small variants. Consequently, the frequency data applied for CNVs and SVs in this study was derived from an even smaller database of available internal cohorts available to SickKids TCAG, as no comprehensive large-scale dataset for these variant types exists (Pellecchia, 2024b). Annotated information from these databases is also commonly Eurocentric, with Europeans being the largest ancestry group represented in gnomAD v4.0. Information from GA4K and these small internal cohorts also suffer from a lack of representation, with most data based on individuals of European ancestry (Kane et al., 2023; Pellecchia, 2024b). Eurocentricity of genetic data poses an issue for variant interpretation, as the frequency of certain variants can vary significantly between ancestral groups, yielding a higher

rate of VUSs, especially among non-European populations, or further variants misclassified as rare or disease-causing (Berger et al., 2022). This necessitates the importance of technology-matched population data to appropriately assess allele frequencies across diverse populations.

Variants in certain genes also appear to suffer from a disproportionate lack of information, also known as '*dark genes*,' for which allele frequency data or *in-silico* prediction scores were often absent in the annotation files. Dark genes are described by Twist Bioscience as genes that are "difficult or impossible to fully sequence" using SRS approaches, with 389 genes included in this classification (Twist Bioscience, 2024). As a result, these genes are typically poorly annotated and understood, causing duplicate variants to appear across multiple unrelated participants. While these duplicate variants were manually excluded as a part of the filtration strategy, issues relating to dark genes and annotated information from databases are recognized as limitations of this study. However, we are hopeful that continual contributions to available information will be made as more research and functional studies are performed using LRS methods, which includes the data generated as a part of this study.

Due to the large number of variants called by LR-WGS, this also necessitated a stringent filtration strategy to yield a manageable number for interpretation. However, the strict filtering criteria used may be inadvertently excluded potential candidate variants. In the analysis of subsequent cases, the team considers revising this strategy in cases where no candidate variants are identified, such as for Case 2. The filtration strategy is also largely dependent on the availability of annotated information for the called variants (i.e., population allele frequency in databases, in-silico prediction scores, etc.), which has its own set of limitations as previously discussed. Finally, as the sample size for this pilot remains small, the diagnostic yield cannot be reliably compared to previous studies. Furthermore, the presence of repeat expansions was not

assessed in this pilot study, but it will be important to include them in future analyses to ensure comprehensive evaluation of all potential pathogenic variants. However, the results from this pilot are promising in demonstrating the utility of LR-WGS in detecting certain disease-causing variants, particularly in genes with well-established links to genetic disorders.

4.3. Significance and conclusions

In summary, this pilot study is the first study to our knowledge that utilizes the PacBio Revio LRS system for molecular diagnosis of NMDs in a small number of individuals. LR-WGS successfully re-identified most of the variants previously detected by short read methods, while also identifying novel variants missed by SRS, suggesting that LR-WGS may be a more comprehensive approach for clinical diagnostics. This supports the expansion of the pilot to include more individuals to describe the diagnostic yield for LRS methods in a diverse cohort. Of the four individuals with strong candidate variants, three were found to carry potential pathogenic variants, potentially leading to a definitive molecular diagnosis and bringing an end to their '*diagnostic odyssey*' (Bauskis et al., 2022). A molecular diagnosis will not only aid in monitoring disease progression, guiding treatment and access to relevant clinical trials, but also confer other benefits including family cascade screening, reproductive risk assessment, and access to advocacy groups and counselling services (Benarroch et al., 2024; Dratch et al., 2024). In addition, the results of this thesis will contribute to the broader scientific community by refining information in clinical genetic databases, assisting in the diagnosis of other individuals found to carry these variants. Taken together, in addition to providing answers for participants and their families, the results of this study are valuable in furthering our understanding of the

genetic etiology of NMDs, as well as asserting the value of LR-WGS for clinical diagnostics in the future.

5. LITERATURE CITED

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., & Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research*, *43*(Database issue), D789-798. https://doi.org/10.1093/nar/gku1205

Antonarakis, S.E., Krawczak, M., & Cooper, D.N. (2019). The nature and mechanisms of human gene mutation. Valle D.L., & Antonarakis S, & Ballabio A, & Beaudet A.L., & Mitchell G.A.(Eds.), The Online Metabolic and Molecular Bases of Inherited Disease. *McGraw-Hill Education.* https://ommbid-mhmedical-com.uml.idm.oclc.org/content.aspx?bookid=2709&sectionid=225072020

Bang, M. L., Centner, T., Fornoff, F., Geach, A. J., Gotthardt, M., McNabb, M., Witt, C. C., Labeit, D., Gregorio, C. C., Granzier, H., & Labeit, S. (2001). The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circulation Research*, *89*(11), 1065–1072. https://doi.org/10.1161/hh2301.100981

Bauskis, A., Strange, C., Molster, C., & Fisher, C. (2022). The diagnostic odyssey: Insights from parents of children living with an undiagnosed condition. *Orphanet Journal of Rare Diseases*, *17*(1), 233. https://doi.org/10.1186/s13023-022-02358-x

Benarroch, L., Bonne, G., Rivier, F., & Hamroun, D. (2024). The 2024 version of the gene table of neuromuscular disorders (nuclear genome). *Neuromuscular Disorders: NMD*, *34*, 126–170. https://doi.org/10.1016/j.nmd.2023.12.007

Berger, S. M., Appelbaum, P. S., Siegel, K., Wynn, J., Saami, A. M., Brokamp, E., O'Connor, B. C., Hamid, R., Martin, D. M., & Chung, W. K. (2022). Challenges of variant reinterpretation: Opinions of stakeholders and need for guidelines. *Genetics in Medicine*, *24*(9), 1878–1887. https://doi.org/10.1016/j.gim.2022.06.002

Bladen, C. L., Salgado, D., Monges, S., Foncuberta, M. E., Kekou, K., Kosma, K., Dawkins, H., Lamont, L., Roy, A. J., Chamova, T., Guergueltcheva, V., Chan, S., Korngut, L., Campbell, C., Dai, Y., Wang, J., Barišić, N., Brabec, P., Lahdetie, J., Walter, M. C., … Lochmüller, H. (2015). The TREAT-NMD DMD Global Database: analysis of more than 7,000 Duchenne muscular dystrophy mutations. *Human Mutation*, *36*(4), 395–402. https://doi.org/10.1002/humu.22758

Boycott, K. M., Hartley, T., Biesecker, L. G., Gibbs, R. A., Innes, A. M., Riess, O., Belmont, J., Dunwoodie, S. L., Jojic, N., Lassmann, T., Mackay, D., Temple, I. K., Visel, A., & Baynam, G. (2019). A Diagnosis for All Rare Genetic Diseases: The

Horizon and the Next Frontiers. *Cell*, *177*(1), 32–37. https://doi.org/10.1016/j.cell.2019.02.040

Broad Institute. (2023, November). gnomAD v4.0: A comprehensive release for understanding human genetic variation. *gnomAD.* https://gnomad.broadinstitute.org/news/2023-11-gnomad-v4-0/

Bruels, C. C., Littel, H. R., Daugherty, A. L., Stafki, S., Estrella, E. A., McGaughy, E. S., Truong, D., Badalamenti, J. P., Pais, L., Ganesh, V. S., O'Donnell-Luria, A., Stalker, H. J., Wang, Y., Collins, C., Behlmann, A., Lemmers, R. J. L. F., van der Maarel, S. M., Laine, R., Ghosh, P. S., Darras, B. T., … Kang, P. B. (2022). Diagnostic capabilities of nanopore long-read sequencing in muscular dystrophy. *Annals of Clinical and Translational Neurology*, *9*(8), 1302–1309. https://doi-org.uml.idm.oclc.org/10.1002/acn3.51612

Campuzano, O., Sanchez-Molero, O., Mademont-Soler, I., Riuró, H., Allegue, C., Coll, M., Pérez-Serra, A., Mates, J., Picó, F., Iglesias, A., & Brugada, R. (2015). Rare Titin (TTN) Variants in Diseases Associated with Sudden Cardiac Death. *International Journal of Molecular Sciences*, *16*(10), 25773–25787. https://doi.org/10.3390/ijms161025773

Cardiovascular Genetics & Genomics Group. (n.d.). Titin transcripts. *Cardiovascular Genetics & Genomics Group, Imperial College London.* https://www.cardiodb.org/titin/titin_transcripts.php

Chaisson, M. J. P., Sanders, A. D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E. J., Rodriguez, O. L., Guo, L., Collins, R. L., Fan, X., Wen, J., Handsaker, R. E., Fairley, S., Kronenberg, Z. N., Kong, X., Hormozdiari, F., Lee, D., Wenger, A. M., Hastie, A. R., … Lee, C. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nature Communications*, *10*(1), 1784. https://doi-org.uml.idm.oclc.org/10.1038/s41467-018-08148-z

Chang, W. H., Mashouri, P., Lozano, A. X., Johnstone, B., Husić, M., Olry, A., Maiella, S., Balci, T. B., Sawyer, S. L., Robinson, P. N., Rath, A., & Brudno, M. (2020). Phenotate: Crowdsourcing phenotype annotations as exercises in undergraduate classes. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *22*(8), 1391–1400. https://doi.org/10.1038/s41436-020-0812-7

Chardon, J. W., Jasmin, B. J., Kothary, R., & Parks, R. J. (2016). Report on the 3rd Ottawa International Conference on Neuromuscular Biology, Disease and Therapy—September 24-26, 2015, Ottawa, Canada. *Journal of Neuromuscular Diseases*, *3*(3), 431–442. https://doi.org/10.3233/JND-169001

Chin, H. L., O'Neill, K., Louie, K., Brown, L., Schlade-Bartusiak, K., Eydoux, P., Rupps, R., Farahani, A., Boerkoel, C. F., & Jones, S. J. M. (2021). An approach to rapid characterization of DMD copy number variants for prenatal risk assessment. *American Journal of Medical Genetics, Part A, 185*(8), 2541–2545. https://doi-org.uml.idm.oclc.org/10.1002/ajmg.a.62349

Cho, Y. N., & Choi, Y. C. (2013). Female carriers of Duchenne muscular dystrophy. *Journal of Genetic Medicine*, *10*(2), 94-98. https://doi.org/10.5734/JGM.2013.10.2.94

Choi, H., Nam, J., Yang, S., & Eyun, S. I. (2024). Highly contiguous genome assembly and gene annotation of the short-finned eel (*Anguilla bicolor pacifica*). *Scientific Data*, *11*(1), 952. https://doi-org.uml.idm.oclc.org/10.1038/s41597-024-03817-9

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Ozen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., & Lifton, R. P. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 19096–19101. https://doi.org/10.1073/pnas.0910672106

Cohen, A. S. A., Farrow, E. G., Abdelmoity, A. T., Alaimo, J. T., Amudhavalli, S. M., Anderson, J. T., Bansal, L., Bartik, L., Baybayan, P., Belden, B., Berrios, C. D., Biswell, R. L., Buczkowicz, P., Buske, O., Chakraborty, S., Cheung, W. A., Coffman, K. A., Cooper, A. M., Cross, L. A., … Pastinen, T. (2022). Genomic answers for children: Dynamic analyses of >1000 pediatric rare disease genomes. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *24*(6), 1336–1348. https://doi.org/10.1016/j.gim.2022.02.007

Cortese, A., Beecroft, S. J., Facchini, S., Curro, R., Cabrera-Serrano, M., Stevanovski, I., Chintalaphani, S. R., Gamaarachchi, H., Weisburd, B., Folland, C., Monahan, G., Scriba, C. K., Dofash, L., Johari, M., Grosz, B. R., Ellis, M., Fearnley, L. G., Tankard, R., Read, J., Merve, A., … Ravenscroft, G. (2024). A CCG expansion in ABCD3 causes oculopharyngodistal myopathy in individuals of European ancestry. *Nature Communications*, *15*(1), 6327. https://doi-org.uml.idm.oclc.org/10.1038/s41467-024-49950-2

Darras, B. T., Jones Jr, H. R., Ryan, M. M., & Darryl, C. (2014). *Neuromuscular disorders of infancy, childhood, and adolescence: A clinician's approach*. Elsevier.

Deenen, J. C. W., Horlings, C. G. C., Verschuuren, J. J. G. M., Verbeek, A. L. M., & van Engelen, B. G. M. (2015). The Epidemiology of Neuromuscular Disorders: A Comprehensive Overview of the Literature. *Journal of Neuromuscular Diseases*, *2*(1), 73–85. https://doi.org/10.3233/JND-140045

Dowling, J. J., D Gonorazky, H., Cohn, R. D., & Campbell, C. (2018). Treating pediatric neuromuscular disorders: The future is now. *American Journal of Medical Genetics. Part A*, *176*(4), 804–841. https://doi.org/10.1002/ajmg.a.38418

Dratch, L., Azage, M., Baldwin, A., Johnson, K., Paul, R. A., Bardakjian, T. M., Michon, S.-C., Amado, D. A., Baer, M., Deik, A. F., Elman, L. B., Gonzalez-Alegre, P., Guo, M. H., Hamedani, A. G., Irwin, D. J., Lasker, A., Orthmann-Murphy, J., Quinn, C., Tropea, T. F., … Ellis, C. A. (2024). Genetic testing in adults with neurologic disorders: Indications, approach, and clinical impacts. *Journal of Neurology*, *271*(2), 733–747. https://doi.org/10.1007/s00415-023-12058-6

Duan, M., Yang, S., Li, X., Tang, X., Cheng, Y., Luo, J., Wang, J., Song, H., Wang, Q., & Zhu, G. X. (2024). Chromosome-level genome assembly and annotation of the *Rhabdophis nuchalis* (Hubei keelback). *Scientific Data, 11*(1), 850. https://doi-org.uml.idm.oclc.org/10.1038/s41597-024-03708-z

Duff, R. M., Tay, V., Hackman, P., Ravenscroft, G., McLean, C., Kennedy, P., Steinbach, A., Schöffler, W., van der Ven, P. F. M., Fürst, D. O., Song, J., Djinović-Carugo, K., Penttilä, S., Raheem, O., Reardon, K., Malandrini, A., Gambelli, S., Villanova, M., Nowak, K. J., Williams, D. R., … Laing, N. G. (2011). Mutations in the N-terminal actin-binding domain of filamin C cause a distal myopathy. *American Journal of Human Genetics, 88*(6), 729–740. https://doi-org.uml.idm.oclc.org/10.1016/j.ajhg.2011.04.021

Geng, C., Zhang, C., Li, P., Tong, Y., Zhu, B., He, J., Zhao, Y., Yao, F., Cui, L. Y., Liang, F., Wang, Y., Wang, Y., Jin, H., Lang, D., Liu, S., Wang, D., Park, M. S., Chen, L., Peng, J., & Dai, Y. (2023). Identification and characterization of two DMD pedigrees with large inversion mutations based on a long-read sequencing pipeline. *European Journal of Human Genetics, 31*(5), 504–511. https://doi-org.uml.idm.oclc.org/10.1038/s41431-022-01190-y

Geng, C., Tong, Y., Zhang, S., Ling, C., Wu, X., Wang, D., & Dai, Y. (2021). Sequence and Structure Characteristics of 22 Deletion Breakpoints in Intron 44 of the *DMD* Gene Based on Long-Read Sequencing. *Frontiers in Genetics, 12*, 638220. https://doi.org/10.3389/fgene.2021.638220

Genoox. (n.d.). Franklin by Genoox. https://franklin.genoox.com/

Gohlke, J., Lindqvist, J., Hourani, Z., Heintzman, S., Tonino, P., Elsheikh, B., Morales, A., Vatta, M., Burghes, A., Granzier, H., & Roggenbuck, J. (2024). Pathomechanisms of Monoallelic variants in TTN causing skeletal muscle disease. *Human Molecular Genetics, 33*(23), 2003–2023. https://doi.org/10.1093/hmg/ddae136

Grigull, L., Lechner, W., Petri, S., Kollewe, K., Dengler, R., Mehmecke, S., Schumacher, U., Lücke, T., Schneider-Gold, C., Köhler, C., Güttsches, A.-K., Kortum, X., & Klawonn, F. (2016). Diagnostic support for selected neuromuscular diseases using answer-pattern recognition and data mining techniques: A proof of concept multicenter prospective trial. *BMC Medical Informatics and Decision Making, 16*, 31. https://doi.org/10.1186/s12911-016-0268-5

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., van de Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L., Abajian, C., Beckmann, C. F., Bernard, A., Bertagnolli, D., Boe, A. F., Cartagena, P. M., Chakravarty, M. M., Chapin, M., Chong, J., Dalley, R. A., … Jones, A. R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature, 489*(7416), 391–399. https://doi-org.uml.idm.oclc.org/10.1038/nature11405

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L. A., Teerlink, C. C., Stanford, J. L., Isaacs, W. B., Xu, J., Cooney, K. A., Lange, E. M., Schleutker, J.,

Carpten, J. D., Powell, I. J., … Sieh, W. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American Journal of Human Genetics*, *99*(4), 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S. J., & Farh, K. K.-H. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, *176*(3), 535-548.e24. https://doi.org/10.1016/j.cell.2018.12.015

Joynt, A. C. M., Axford, M. M., Chad, L., & Costain, G. (2021). Understanding genetic variants of uncertain significance. *Paediatrics & Child Health*, *27*(1), 10–11. https://doi.org/10.1093/pch/pxab070

Kane, N. J., Cohen, A. S. A., Berrios, C., Jones, B., Pastinen, T., & Hoffman, M. A. (2023). Committing to genomic answers for all kids: Evaluating inequity in genomic research enrollment. *Genetics in Medicine*, *25*(9), 100895. https://doi.org/10.1016/j.gim.2023.100895

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., … MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kleinberger, J., Maloney, K. A., Pollin, T. I., & Jeng, L. J. (2016). An openly available online tool for implementing the ACMG/AMP standards and guidelines for the interpretation of sequence variants. *Genetics in Medicine*, *18*(11), 1165. https://doi.org/10.1038/gim.2016.13

Köhler, S., Øien, N. C., Buske, O. J., Groza, T., Jacobsen, J. O. B., McNamara, C., Vasilevsky, N., Carmody, L. C., Gourdine, J. P., Gargano, M., McMurry, J. A., Danis, D., Mungall, C. J., Smedley, D., Haendel, M., & Robinson, P. N. (2019). Encoding Clinical Data with the Human Phenotype Ontology for Computational Differential Diagnostics. *Current Protocols in Human Genetics*, *103*(1), e92. https://doi.org/10.1002/cphg.92

Kolesnikov, A., Cook, D., Nattestad, M., Brambrink, L., McNulty, B., Gorzynski, J., Goenka, S., Ashley, E. A., Jain, M., Miga, K. H., Paten, B., Chang, P. C., Carroll, A., & Shafin, K. (2024). Local read haplotagging enables accurate long-read small variant calling. *Nature Communications*, *15*(1), 5907. https://doi-org.uml.idm.oclc.org/10.1038/s41467-024-50079-5

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., & Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*(1):D980-5. https://doi.org/10.1093/nar/gkt1113

Lee, H., Martinez-Agosto, J. A., Rexach, J., & Fogel, B. L. (2019). Next generation sequencing in clinical diagnosis. *The Lancet. Neurology*, *18*(5), 426. https://doi.org/10.1016/S1474-4422(19)30110-3

Liu, X., Wu, C., Li, C., & Boerwinkle, E. (2016). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Human Mutation*, *37*(3), 235–241. https://doi.org/10.1002/humu.22932

Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nature Reviews: Genetics*, *21*(10), 597–614. https://doi-org.uml.idm.oclc.org/10.1038/s41576-020-0236-x

Mantere, T., Kersten, S., & Hoischen, A. (2019). Long-Read Sequencing Emerging in Medical Genetics. *Frontiers in Genetics*, *10*, 426. https://doi.org/10.3389/fgene.2019.00426

Marwaha, S., Knowles, J. W., & Ashley, E. A. (2022). A guide for the diagnosis of rare and undiagnosed disease: Beyond the exome. *Genome Medicine*, *14*(1), 23. https://doi.org/10.1186/s13073-022-01026-w

Merker, J. D., Wenger, A. M., Sneddon, T., Grove, M., Zappala, Z., Fresard, L., Waggott, D., Utiramerur, S., Hou, Y., Smith, K. S., Montgomery, S. B., Wheeler, M., Buchan, J. G., Lambert, C. C., Eng, K. S., Hickey, L., Korlach, J., Ford, J., & Ashley, E. A. (2018). Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *20*(1), 159–163. https://doi.org/10.1038/gim.2017.86

Miao, H., Zhou, J., Yang, Q., Liang, F., Wang, D., Ma, N., Gao, B., Du, J., Lin, G., Wang, K., & Zhang, Q. (2018). Long-read sequencing identified a causal structural variant in an exome-negative case and enabled preimplantation genetic diagnosis. *Hereditas*, *155*, 32. https://doi.org/10.1186/s41065-018-0069-1

Mitsuhashi, S., & Matsumoto, N. (2020). Long-read sequencing for rare human genetic diseases. *Journal of Human Genetics*, *65*(1), 11–19. Ovid MEDLINE(R) <2020>. https://doi.org/10.1038/s10038-019-0671-8

Oehler, J. B., Wright, H., Stark, Z., Mallett, A. J., & Schmitz, U. (2023). The application of long-read sequencing in clinical settings. *Human Genomics*, *17*(1), 73. https://doi.org/10.1186/s40246-023-00522-3

Okumura, L. K., & Masuda, S. (2024). Complete genome sequence of Synechocystis sp. LKSZ1 isolated from a pond in the Suzukakedai campus of Tokyo Institute of Technology. *Microbiology Resource Announcements*, *13*(10), e0054924. https://doi-org.uml.idm.oclc.org/10.1128/mra.00549-24

Owusu, R., & Savarese, M. (2023). Long-read sequencing improves diagnostic rate in neuromuscular disorders. *Acta Myologica*, *42*(4), 123–128. https://doi-org.uml.idm.oclc.org/10.36185/2532-1900-394

Pacific Biosciences of California, Inc. (2022, October 26). *PacBio announces Revio, a revolutionary new long read sequencing system designed to provide 15 times more HIFI data and human genomes at scale for under $1,000.* PacBio. https://www.pacb.com/press_releases/pacbio-announces-revio-a-revolutionary-new-long-read-sequencing-system-designed-to-provide-15-times-more-hifi-data-and-human-genomes-at-scale-for-under-1000/

Pejaver, V., Byrne, A. B., Feng, B.-J., Pagel, K. A., Mooney, S. D., Karchin, R., O'Donnell-Luria, A., Harrison, S. M., Tavtigian, S. V., Greenblatt, M. S., Biesecker, L. G., Radivojac, P., Brenner, S. E., & ClinGen Sequence Variant Interpretation Working Group. (2022). Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *American Journal of Human Genetics, 109*(12), 2163–2177. https://doi.org/10.1016/j.ajhg.2022.10.013

Pellecchia, G. (2024a). *TCAG small variant annotation pipeline.* [Unpublished manuscript].

Pellecchia, G. (2024b). *TCAG CNV/SV annotation pipeline.* [Unpublished manuscript].

Perez, G., Barber, G. P., Benet-Pages, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J. N., Hinrichs, A. S., Lee, C. M., Nassar, L. R., Raney, B. J., Speir, M. L., van Baren, M. J., Vaske, C. J., Haussler, D., Kent, W. J., & Haeussler, M. (2025). The UCSC Genome Browser database: 2025 update. Nucleic Acids Research, 53(D1), D1243–D1249. https://doi.org/10.1093/nar/gkae974

Perrin, A., Van Goethem, C., Thèze, C., Puechberty, J., Guignard, T., Lecardonnel, B., Lacourt, D., Métay, C., Isapof, A., Whalen, S., Ferreiro, A., Arne-Bes, M. C., Quijano-Roy, S., Nectoux, J., Leturcq, F., Richard, P., Larrieux, M., Bergougnoux, A., Pellestor, F., Koenig, M., … Cossée, M. (2022). Long-Reads Sequencing Strategy to Localize Variants in *TTN* Repeated Domains. *The Journal of Molecular Diagnostics*, *24*(7), 719–726. https://doi.org/10.1016/j.jmoldx.2022.04.006

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research, 20*(1), 110–121. https://doi.org/10.1101/gr.097857.109

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, *36*(10), 983–987. https://doi.org/10.1038/nbt.4235

Previtali, S. C., Scarlato, M., Vezzulli, P., Ruggieri, A., Velardo, D., Benedetti, S., Torini, G., Colombo, B., Maggi, L., Di Bella, D., Gellera, C., D'Angelo, G., & Mora, M. (2019). Expanding the central nervous system disease spectrum associated with FLNC mutation. *Muscle & Nerve, 59*(5), E33–E37. https://doi.org/10.1002/mus.26443

R Core Team (2024). R: A Language and Environment for Statistical Computing. Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, *47*(D1), D886–D894. https://doi.org/10.1093/nar/gky1016

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., & ACMG Laboratory Quality Assurance Committee. (2015). 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. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., & Mesirov, J. P. (2017). Variant Review with the Integrative Genomics Viewer. *Cancer Research*, *77*(21), e31–e34. https://doi.org/10.1158/0008-5472.CAN-17-0337

Robinson, P. N., Köhler, S., Bauer, S., Seelow, D., Horn, D., & Mundlos, S. (2008). The Human Phenotype Ontology: A tool for annotating and analyzing human hereditary disease. *American Journal of Human Genetics*, *83*(5), 610–615. https://doi.org/10.1016/j.ajhg.2008.09.017

Samocha, K. E., Kosmicki, J. A., Karczewski, K. J., O'Donnell-Luria, A. H., Pierce-Hoffman, E., MacArthur, D. G., ... & Daly, M. J. (2017). Regional missense constraint improves variant deleteriousness prediction. *BioRxiv*, 148353. https://doi.org/10.1101/148353

Sano, Y., Koyanagi, Y., Wong, J. H., Murakami, Y., Fujiwara, K., Endo, M., Aoi, T., Hashimoto, K., Nakazawa, T., Wada, Y., Ueno, S., Gao, D., Murakami, A., Hotta, Y., Ikeda, Y., Nishiguchi, K. M., Momozawa, Y., Sonoda, K.-H., Akiyama, M., & Fujimoto, A. (2022). Likely pathogenic structural variants in genetically unsolved patients with retinitis pigmentosa revealed by long-read sequencing. *Journal of Medical Genetics*, *59*(11), 1133–1138. https://doi.org/10.1136/jmedgenet-2022-108428

Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods*, *11*(4), 361–362. https://doi-org.uml.idm.oclc.org/10.1038/nmeth.2890

Stenton, S. L., O'Leary, M. C., Lemire, G., VanNoy, G. E., DiTroia, S., Ganesh, V. S., Groopman, E., O'Heir, E., Mangilog, B., Osei-Owusu, I., Pais, L. S., Serrano, J., Singer-Berk, M., Weisburd, B., Wilson, M. W., Austin-Tse, C., Abdelhakim, M., Althagafi, A., Babbi, G., … O'Donnell-Luria, A. (2024). Critical assessment of variant prioritization methods for rare disease diagnosis within the rare genomes project. *Human Genomics*, *18*(1), 44. https://doi.org/10.1186/s40246-024-00604-w

Tong, Y. R., Geng, C., Guan, Y. Z., Zhao, Y. H., Ren, H. T., Yao, F. X., Ling, C., Wang, D. C., Chen, L., Cui, L. Y., Zhang, S. Y., & Dai, Y. (2020). A Comprehensive Analysis of 2013 Dystrophinopathies in China: A Report From National Rare Disease Center. *Frontiers in Neurology*, *11*, 572006. https://doi.org/10.3389/fneur.2020.572006

Thompson, R., Spendiff, S., Roos, A., Bourque, P. R., Warman Chardon, J., Kirschner, J., Horvath, R., & Lochmüller, H. (2020). Advances in the diagnosis of inherited neuromuscular diseases and implications for therapy development. *The Lancet. Neurology*, *19*(6), 522–532. https://doi.org/10.1016/S1474-4422(20)30028-4

Twist Bioscience. (2024). *Long Read Sequencing Panels*. https://www.twistbioscience.com/products/ngs/Long-Read-Sequencing-Panels

van den Bogaart, F. J., Claeys, K. G., Kley, R. A., Kusters, B., Schrading, S., Kamsteeg, E. J., & Voermans, N. C. (2017). Widening the spectrum of filamin-C myopathy: Predominantly proximal myopathy due to the p.A193T mutation in the actin-binding domain of FLNC. *Neuromuscular Disorders*, *27*(1), 73–77. https://doi.org/10.1016/j.nmd.2016.09.017

VanOudenhove, J., Halene, S., & Mendez, L. (2023). Is it the time to integrate novel sequencing technologies into clinical practice? *Current Opinion in Hematology*, *30*(2), 70–77. https://doi.org/10.1097/MOH.0000000000000754

Walsh, R., Thomson, K. L., Ware, J. S., Funke, B. H., Woodley, J., McGuire, K. J., Mazzarotto, F., Blair, E., Seller, A., Taylor, J. C., Minikel, E. V., Exome Aggregation Consortium, MacArthur, D. G., Farrall, M., Cook, S. A., & Watkins, H. (2017). Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genetics in Medicine*, *19*(2), 192–203. https://doi-org.uml.idm.oclc.org/10.1038/gim.2016.90

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164. https://doi.org/10.1093/nar/gkq603

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wojcik, M. H., Reuter, C. M., Marwaha, S., Mahmoud, M., Duyzend, M. H., Barseghyan, H., Yuan, B., Boone, P. M., Groopman, E. E., Delot, E. C., Jain, D., Sanchis-Juan, A., Starita, L. M., Talkowski, M., Montgomery, S. B., Bamshad, M. J., Chong, J. X., Wheeler, M. T., Berger, S. I., … Miller, D. E. (2023). Beyond the exome: What's next in diagnostic testing for Mendelian conditions. *American Journal of Human Genetics*, *110*(8), 1229–1248. https://doi.org/10.1016/j.ajhg.2023.06.009

Wright, G.E.B. (2022). *Identification of causal genetic variants and genes in neuromuscular disorders.* [Unpublished document]

Xie, Z., Sun, C., Zhang, S., Liu, Y., Yu, M., Zheng, Y., Meng, L., Acharya, A., Cornejo-Sanchez, D. M., Wang, G., Zhang, W., Schrauwen, I., Leal, S. M., Wang, Z., & Yuan, Y. (2020). Long-read whole-genome sequencing for the genetic diagnosis of

dystrophinopathies. *Annals of Clinical and Translational Neurology*, *7*(10), 2041–2046. https://doi-org.uml.idm.oclc.org/10.1002/acn3.51201

Yang, H., Robinson, P. N., & Wang, K. (2015). Phenolyzer: Phenotype-based prioritization of candidate genes for human diseases. *Nature Methods*, *12*(9), 841–843. https://doi.org/10.1038/nmeth.3484

Yeetong, P., Kulsirichawaroj, P., Kumutpongpanich, T., Srichomthong, C., Od-Ek, P., Rakwongkhachon, S., Thamcharoenvipas, T., Sanmaneechai, O., Pongpanich, M., & Shotelersuk, V. (2023). Long-read Nanopore sequencing identified D4Z4 contractions in patients with facioscapulohumeral muscular dystrophy. *Neuromuscular Disorders, 33*(7), 551–556. https://doi-org.uml.idm.oclc.org/10.1016/j.nmd.2023.05.004

Yuen, R. K. C., Thiruvahindrapuram, B., Merico, D., Walker, S., Tammimies, K., Hoang, N., Chrysler, C., Nalpathamkalam, T., Pellecchia, G., Liu, Y., Gazzellone, M. J., D'Abate, L., Deneault, E., Howe, J. L., Liu, R. S. C., Thompson, A., Zarrei, M., Uddin, M., Marshall, C. R., … Scherer, S. W. (2015). Whole-genome sequencing of quartet families with autism spectrum disorder. *Nature Medicine*, *21*(2), 185–191. https://doi.org/10.1038/nm.3792

Yu, J., Deng, J., Guo, X., Shan, J., Luan, X., Cao, L., Zhao, J., Yu, M., Zhang, W., Lv, H., Xie, Z., Meng, L., Zheng, Y., Zhao, Y., Gang, Q., Wang, Q., Liu, J., Zhu, M., Zhou, B., Li, P., … Wang, Z. (2021). The GGC repeat expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy type 3. *Brain, 144*(6), 1819–1832. https://doi-org.uml.idm.oclc.org/10.1093/brain/awab077

Yu, J., Shan, J., Yu, M., Di, L., Xie, Z., Zhang, W., ... & Deng, J. (2022). The CGG repeat expansion in RILPL1 is associated with oculopharyngodistal myopathy type 4. *The American Journal of Human Genetics*, *109*(3), 533-541. https://doi.org/10.1016/j.ajhg.2022.01.012

Zatz, M., Passos-Bueno, M. R., & Vainzof, M. (2016). Neuromuscular disorders: Genes, genetic counseling and therapeutic trials. *Genetics and Molecular Biology*, *39*(3), 339–348. https://doi.org/10.1590/1678-4685-GMB-2016-0019

6. APPENDICES

Supplementary Figure 1. Standardized data collection sheet for extraction of clinical information relevant to the genetic variant filtration process.

Data Capture Sheet

(To be used with master list of unique participant codes)

Participant Number: _____

Proband Characteristics:

- Sex: _____

- Gender: _____
- Year of birth: _____
- Age of mother: _____
- Prenatal complications (if yes, describe):
  _____
_____
- Gestation of pregnancy (in weeks): _____
- Type of delivery (Caesarean, vaginal, etc.): _____
- Weight at birth (in grams): _____
- Prematurity (Y/N): _____
- Neonatal complications (if yes, describe):
  _____
_____

Family History:

- Family history of condition/known affected (relationship to proband, sex/gender):
  _____
  _____
  _____
  _____
- Pedigree (diagram):

- Possible inheritance pattern(s) (check all that may apply):
  - De novo AD
  - Inherited AD
  - AR

- De novo XL
- Inherited XL
- Mitochondrial
- Other (describe):

  _____

- Parental consanguinity (Y/N): _____
  - If yes, degree of relatedness:

    _____
- Parental ethnicity:

  _____

Initial Presentation:

- Age at presentation: _____
- Description of presentation:

  _____

_____

- Initial presumptive diagnosis (if present):

  _____

Phenotype:

- Age: _____
- Date: _____
- *Review of Systems:*
  - Growth:

    _____
  - Neurological:
    - *Neuropsychological Symptoms (if present):*
      - Description (i.e., difficulties with memory, speech, cognition, intellectual disability, etc.):

        _____

        _____

        _____
    - *Muscle Strength:*
      - Description (normal, slight weakness, weakness):

        _____

        _____
      - Location (proximal, distal, generalized):

        _____
      - Predominance (upper, lower):

        _____
      - Bulbar involvement (Y/N/Not Reported): _____
      - Scapular winging (Y/N/Not Reported): _____

        - Anatomic distribution (symmetrical, asymmetrical): _____
        - Associated symptoms (e.g., cramping, stiffness, pain): _____
        - Symptom frequency (acute, chronic, episodic): _____
        - Symptom progression (stable, improving, worsening): _____
            - If worsening, rate of progression (e.g., days, months, years): _____
        - Contributing or aggravating factors (e.g., exercise vs. rest): _____
    - *Muscle Function:*
        - Gait characteristics (describe): _____
        - Gower manoeuvre (Y/N/Not Described): _____
        - Passive range of motion (describe, i.e., shoulder abduction, hip flexion, contractures/arthrogryposis, etc.): _____
- *Muscle Tone:*
    - Description (normal, spasticity, rigidity, not reported, etc.): _____
    - Location (proximal, distal, generalized): _____
    - Predominance (upper, lower): _____
    - Anatomic distribution (symmetrical, asymmetrical): _____
- *Muscle Mass:*
    - Description (normal, decreased, atrophy, hypertrophy): _____
    - Location (proximal, distal, generalized): _____
    - Predominance (upper, lower): _____
    - Anatomic distribution (symmetrical, asymmetrical): _____

- *Fasciculations:*
    - Description (prominent, absent): _____
    - Location (proximal, distal, generalized): _____
    - Predominance (upper, lower): _____
    - Anatomic distribution (symmetrical, asymmetrical): _____
- *Sensation:*
    - Description (normal, increased, decreased): _____
    - Location (proximal, distal, generalized): _____
    - Predominance (upper, lower): _____
    - Anatomic distribution (symmetrical, asymmetrical): _____
- *Reflexes:*
    - Deep tendon reflexes (normal, increased, decreased): _____
    - Infantile reflexes (persistent, absent): _____
    - Babinski sign (+/-, absent): _____
- *Additional notes:*

  _____

  _____

- *Skeletal:*
    - Description (e.g., scoliosis, dislocations, joint hypermobility, etc.): _____ _
    - Facial features (i.e., myopathic/dysmorphic facies, etc.): _____
    - Cardiovascular:
    - _____
    - Respiratory: _____
    - Hepatic: _____
    - Gastrointestinal: _____
    - Urogenital: _____

- Hair/skin:
  _____
- Endocrine:
  _____
- Immunological:
- _____

Laboratory Investigations:

a. ***Genetic testing findings:***

- Overview of tests performed:
  _____
  _____
- Overview of results:
  _____
  _____
- *Single-Gene Tests (if applicable):*
  - Year performed: _____
  - Test description (i.e., company, genes tested):
    _____
    _____
    - Test technology (e.g., next-generation, Sanger, etc.):
      _____
    - Findings (if applicable):
      - Gene name: _____
        - Variant: _____
        - Interpretation: _____
      - Gene name: _____
        - Variant: _____
        - Interpretation: _____
      - Gene name: _____
        - Variant: _____
        - Interpretation: _____
    - Family studies performed (Y/N): _____
      - Description (e.g., parental carrier testing, segregation analysis):
        _____
        _____
        - Findings (if applicable):
          _____
          _____
- *Gene Panel (if applicable):*
  - Year performed: _____
  - Panel description (i.e., company, what was included):
    _____

_____

- Test technology (e.g., next-generation, Sanger, etc.): _____
- Panel findings (if applicable):
    - Gene name: _____
        - Variant: _____
        - Interpretation: _____
    - Gene name: _____
        - Variant: _____
        - Interpretation: _____
    - Gene name: _____
        - Variant: _____
        - Interpretation: _____
- Family studies performed (Y/N): _____
    - Description (e.g., parental carrier testing, segregation analysis): _____

      _____
    - Findings (if applicable): _____

      _____

- *Exome Sequencing (if applicable):*
    - Year performed: _____
    - Type of exome (singleton, trio, company): _____
    - Exome findings (if applicable):
        - Gene name: _____
            - Variant: _____
            - Interpretation: _____
        - Gene name: _____
            - Variant: _____
            - Interpretation: _____
        - Gene name: _____
            - Variant: _____
            - Interpretation: _____
    - Family studies performed (Y/N): _____
        - Description (e.g., parental carrier testing, segregation analysis): _____

          _____
        - Findings (if applicable): _____

          _____

- *Other Research Genetic Testing (if applicable):*
        - Description (e.g., parental carrier testing, segregation analysis): _____

_____

- Findings (if applicable):
  _____
  _____

b. ***Diagnostic laboratory findings (to be captured at initial presentation):***

Muscle function	Date of Measurement	Result(s)	Normal Range(s)	Additional Notes
Creatine kinase				
Myoglobin				
Metabolism				
ALT, AST, GGT				
Blood glucose				
Ammonia				
Acylcarnitine				
Urine organic acids				
Plasma amino acids				

Lumbar puncture (neurotransmitter metabolites, amino acids, lactate)				
Other				

c. ***Electromyography or nerve conduction velocity findings (if applicable):***

- Age: _____
    - Study type (EMG, NCV): _____
    - Findings:
      _____
      _____
      ____
- Age: _____
    - Study type (EMG, NCV): _____
    - Findings:
      _____
      _____
      ____

d. ***Specific imaging findings (muscle or brain MRI, MRS, etc., if applicable):***

- Age: _____
    - Imaging type:
      _____
    - Tissue and location (i.e., brain or muscle, specific muscle group/brain area):
      _____
      _____
    - Findings:
      _____
      _____
      ____
- Age: _____
    - Imaging type:
      _____
    - Tissue and location (i.e., brain or muscle, specific muscle group/brain area):
      _____
      _____
    - Findings:
      _____
      _____
      ____

- Age: _____
    - Imaging type:
      _____
    - Tissue and location (i.e., brain or muscle, specific muscle group/brain area):
      _____
      _____
    - Findings:
      _____
      _____
      ____

e. ***Muscle or nerve biopsy findings (if applicable):***

- Age: _____
    - Tissue and location sampled (i.e., muscle or nerve, specific muscle group):
      _____
      _____
    - Findings:
      _____
      _____
- Age: _____
    - Tissue and location sampled (i.e., muscle or nerve, specific muscle group):
      _____
      _____
    - Findings:
      _____
      _____

f. ***Other diagnostic testing:***

_____
_____

<u>Complications:</u>

a. **Additional confounding diagnoses** (Y/N): **_____**
    - If yes, describe:
      _____
      _____

b. **Death (if patient is deceased):**
    - Age (years/months): _____
    - Primary cause:
      _____
      _____
    - Findings on autopsy (if performed):
      _____

_____

Other Relevant Medical Assessment History (e.g. physiotherapy, genetic counselling, etc.):

_____

_____

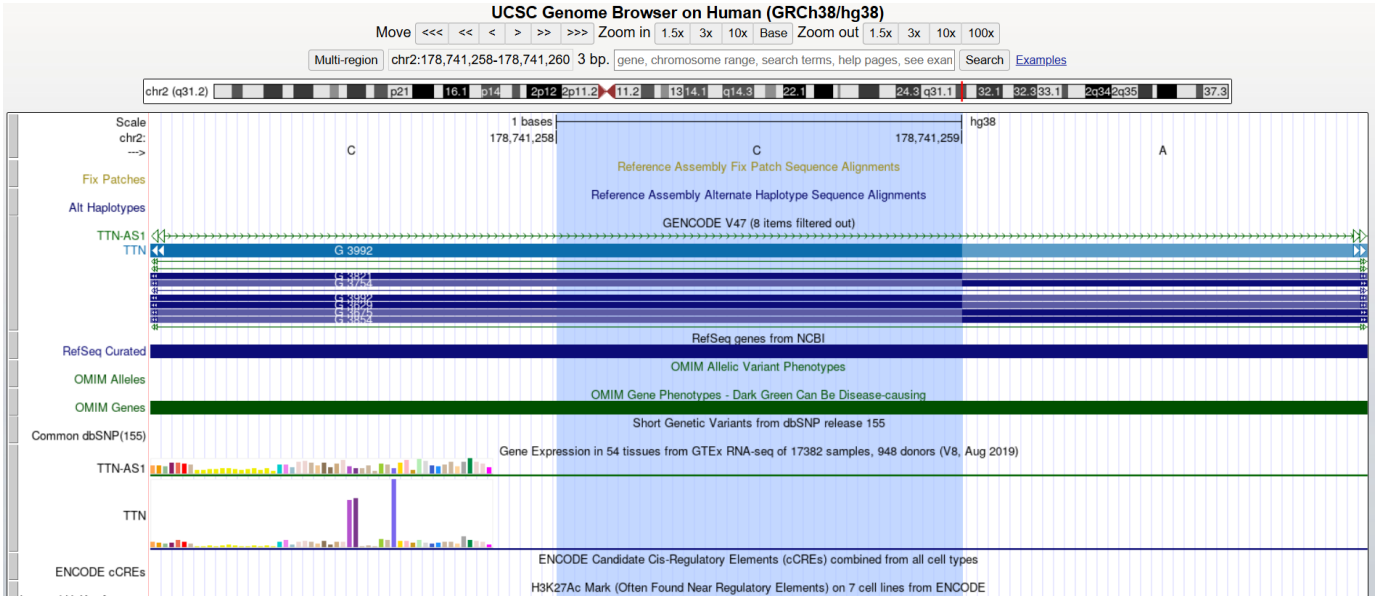
_____

Notes:

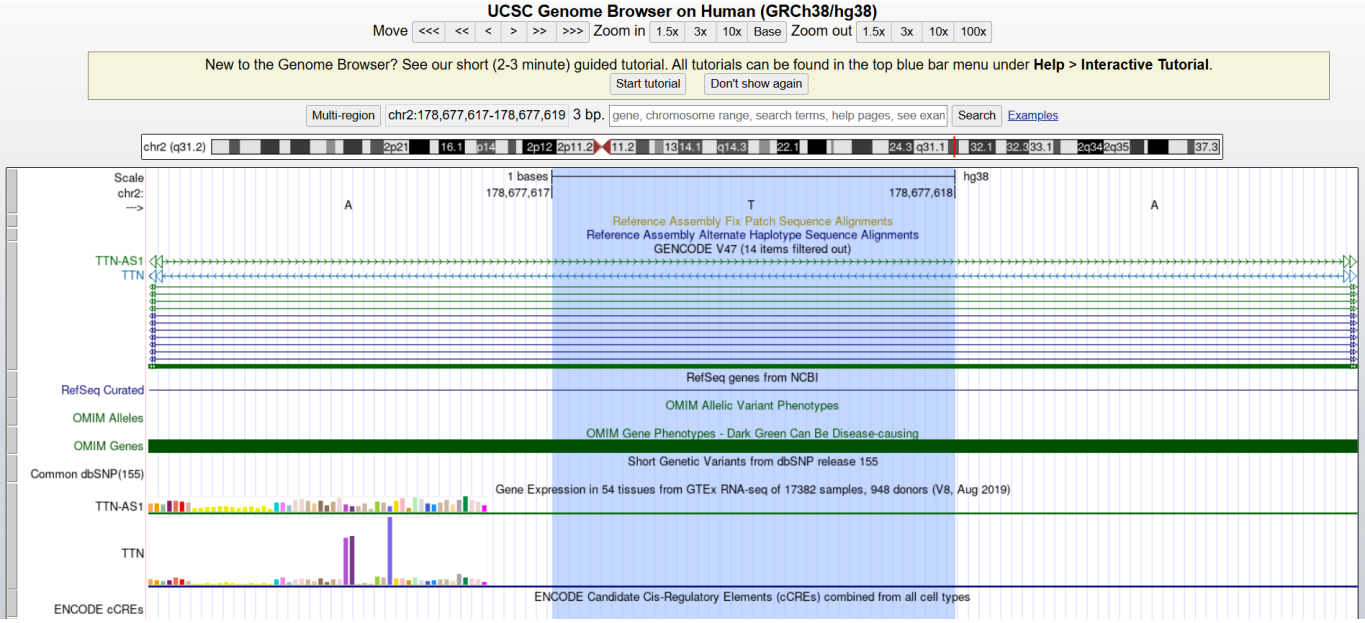
_____

_____

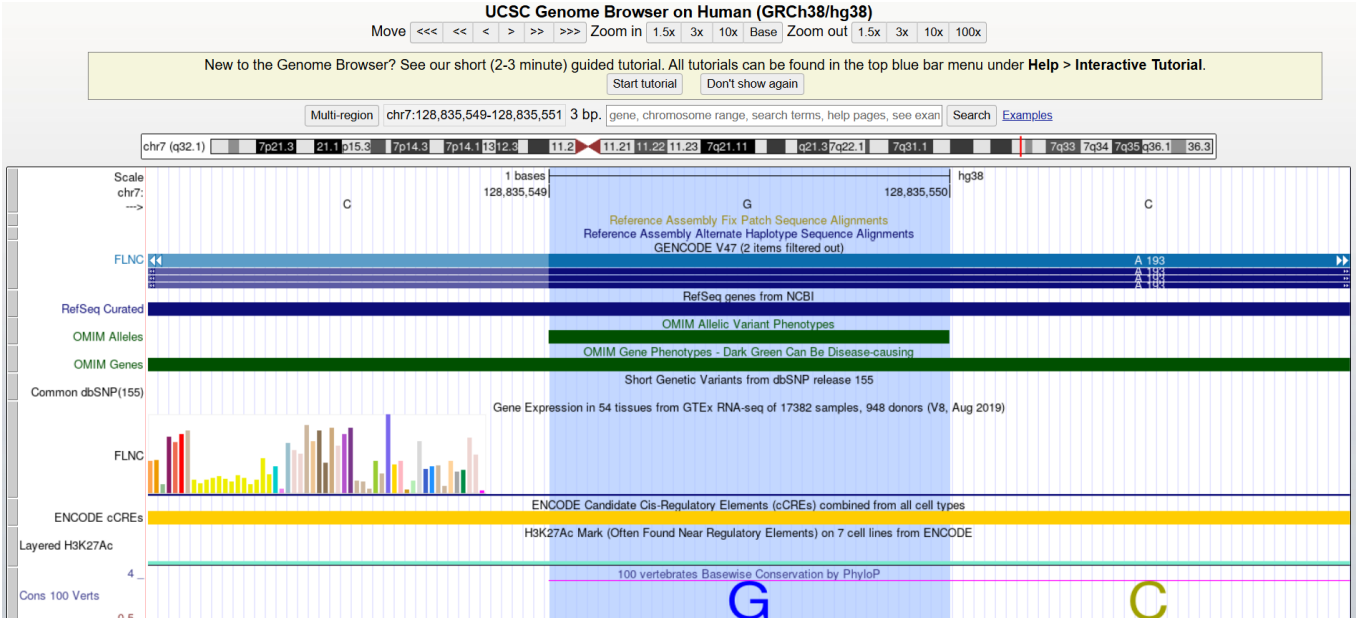
_____



Supplementary Figure 2. UCSC Genome Browser snapshot visualizing the genomic regions affected by the exonic variant in the *TTN* gene (c.11023G>A, p.Gly3675Ser) as identified in Case 1. The variant is indicated in blue.



Supplementary Figure 3. UCSC Genome Browser snapshot visualizing the genomic regions affected by the intronic variant in the *TTN* gene (c.34291+3A>G; chr2:178677618T>C) as identified in Case 1. The variant is indicated in blue.



Supplementary Figure 4. UCSC Genome Browser snapshot visualizing the genomic regions affected by the missense variant in the *FLNC* gene (c.577G>A; p.Ala193Thr) as identified in Case 3. The variant is indicated in blue.