

Solving the genetics of neuromuscular disorders using multi-omic long-read sequencing

by

Rissa Hamlin

A Thesis submitted to the Faculty of Graduate and Postdoctoral Studies of

the University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biochemistry and Medical Genetics

Genetic Counselling Program

Department of Biochemistry & Medical Genetics

Max Rady College of Medicine

University of Manitoba

Winnipeg, Manitoba, Canada

Copyright © 2026 by Rissa Hamlin

Abstract

Neuromuscular disorders (NMDs) are a heterogeneous group of diseases in which approximately half of affected individuals remain without a genetic diagnosis, highlighting the need for alternate diagnostic approaches. Long-read whole-genome sequencing (LR-WGS) improves detection of diverse variant classes and variants within complex genomic regions. Complementary long-read RNA sequencing (LR-RNA-seq) from disease-relevant tissues (e.g., skeletal muscle) can improve diagnostic outcomes by identifying gene and isoform expression outliers. This study evaluated the utility of integrated LR-WGS and LR-RNA-seq for diagnosis and functional variant interpretation in 22 individuals with genetically unresolved NMDs.

Participants were recruited from the Manitoba clinical genetics program. LR-WGS was performed on the PacBio Revio system (mean 17.4 kb read length, 34.6X depth) and candidate variants were prioritized using a phenotype-driven, NMD-centred bioinformatic pipeline. LR-RNA-seq was performed on skeletal muscle biopsies from 19 individuals and analyzed with IsoRanker to identify gene expression and isoform outliers, assess aberrant splicing and support variant interpretation.

The cohort exhibited prolonged diagnostic odysseys (mean 20.32 years) and extensive prior genetic testing (mean 3.95 tests). LR-WGS generated approximately five million variants per genome, with 113 final variants prioritized for manual review across the cohort. LR-WGS identified pathogenic variants in five individuals from four families and enabled reclassification of a variant in an additional case. Transcriptomic analyses further supported pathogenic variant identification and interpretation in two unrelated individuals, while identifying disease-relevant expression profiles in two other unrelated individuals. Altogether, integrated LR-WGS and transcriptomic analyses facilitated genetic diagnosis in 27% of the cohort, including copy-number,

structural, splicing and repeat-expansion variants not captured by previous testing. This study demonstrates the utility of integrated LR-multi-omics for NMD diagnostics, improving diagnostic yield while providing functional insights into unresolved cases and supporting broader implementation in rare disease diagnostics.

Acknowledgements

Thank you to my thesis advisor, Dr. Galen Wright, for his mentorship, guidance and invaluable expertise throughout this project. His passion for genetics, dedication to his trainees, and rigour in scientific research have shaped this project and will have a lasting impact on me as a researcher. I am deeply grateful for the opportunity to have learned from and worked alongside such an exceptional scientist and mentor.

Thank you to my thesis committee, Dr. Britt Drögemöller, Jessica Hartley, Dr. Patrick Frosk and Dr. Paul Marcogliese for your contributions to this project. Your insights broadened the scope of this work and helped ensure that it may ultimately benefit patients in more meaningful ways.

Thank you to the patients and families who contributed to this study. I am honoured to have been entrusted with the opportunity to perform this work and hope that it will have a meaningful impact for individuals living with neuromuscular disorders in the years to come

Thank you to other essential members of the research team, including Alana Lamont, John Pham, Sophia Perron, and Dr. Alejandra Llanes-Cuesta, for their collaboration. Thank you to Feryal Ladla, whose expertise in bioinformatics was essential to the development of our IsoRanker analysis and who generously supported me in developing new skills in bioinformatics. Lastly, a sincere thank you to Alexandria Martin for her work in the pilot project of this study, which set many foundational elements for the remainder of this study.

I am thankful for the community of support that I had access to as a student of the UManitoba Genetic Counselling Program. Thank you to Program Director Jessica Hartley and Assistant Program Director Angela Krutish for your encouragement and feedback; your hard work has shaped a world-class program that I was grateful to be a part of. I am thankful to my clinical rotation supervisors, whose support and knowledge have been integral to my development as a genetic counsellor. I am grateful to Emily Moss, Zoe Goldstone-Joubert and Christian LaPierre (my second-years) for their encouragement and mentorship, and to Shreya Shah, Sam Smith and Maddie Gebhardt (my first-years) for a wonderful year of learning and friendship. Moreover, I am eternally thankful to my class members Zahra Nasser-Moghaddam and David Di Iorio for these two years. They inspired me to work harder, learn more and show more kindness every day. I look forward to seeing how we all continue to shape the field of genetics and patient care in our careers moving forward.

Last but not least, I am thankful to my friends and family who supported me throughout my time in this program. I could not have succeeded without your love and support. In particular, I am thankful to my grandfather, Robert Hamlin, who passed away this year. He inspired my love for learning and for helping others, and was himself the recipient of a Master's degree.

For this project, I was grateful to have received funding from the Canada Graduate Scholarship – Master's, awarded by the Canadian Institutes of Health Research. The Wright Laboratory is also grateful to Research Manitoba for providing funding for this study.

Table of Contents

Abstract.....	2
Acknowledgements	4
List of figures.....	9
List of tables.....	11
List of abbreviations	14
Chapter 1: Background and review of the literature	17
1.1. Neuromuscular Disorders (NMDs).....	17
1.1.1. What are NMDs?	17
1.1.2. Epidemiology and impact; NMDs in a Manitoban and Canadian context	18
1.1.3. Psychosocial significance of NMDs	20
1.2. NMDs as genetic disorders.....	24
1.2.1. Genetics of NMDs	24
1.2.2. Genetic testing in NMD diagnostics	27
1.2.3. Genetic diagnosis for NMDs: Treatment and management.....	30
1.3. Current genetic testing methods for NMDs.....	31
1.3.1. Massively parallel sequencing (MPS) for rare human disease	32
1.3.2. Gene panels	33
1.3.3. Whole exome sequencing (WES)	34
1.4. Comprehensive genetic testing methods for NMDs	36
1.4.1. Whole genome sequencing (WGS).....	36
1.4.2. Long-read sequencing (LRS).....	38
1.4.3. Approaches to variant filtration and interpretation	43
1.4.4. Long-read RNA-sequencing as a complementary tool for genetic diagnosis.....	44
1.5. Thesis rationale	50
Chapter 2: Methods	51
2.1. Ethical approval and funding	51
2.2. Participant recruitment.....	51
2.3. Retrospective chart review and phenotyping	52
2.4. LR-WGS, variant annotation and data transfer.....	54
2.5. LR-WGS bioinformatics filtering pipeline	56
2.5.1. Hard filters for small variants	56
2.5.2. Hard filters for SVs and CNVs	61
2.5.3. Phenotypic filters for short variants, SVs and CNVs	62
2.6. STR genotyping of LR-WGS data.....	64
2.7. Manual genetic variant prioritization and validation	66
2.8. LR-RNA sequencing from disease-relevant skeletal muscle tissue	71

Chapter 3: Results	73
3.1. Cohort clinical and demographic results	73
3.1.1. Length of diagnostic odyssey and genetic testing.....	77
3.1.2. Genetic diagnoses established by clinical testing before completion of study analyses	80
3.2. LRS Results	82
3.2.1. LR-WGS and bioinformatic filtering.....	82
3.2.2. LR-WGS and TRGT STR genotyping.....	85
3.2.3. LR-RNA-seq and IsoRanke	87
3.2.4. Cohort diagnostic outcomes.....	91
Chapter 4: Case descriptions	94
4.1. Case descriptions: Solved cases demonstrating diagnostic resolution through LR- WGS and LR-RNA-seq	94
4.1.1. Complementary LR-WGS and LR-RNA-seq resolved a large-scale copy number variant in brother duo (NM03 and NM06)	95
4.1.2. LR-WGS identified a novel <i>ORAI</i> variant in a recently identified NMD gene (NM21)	108
4.1.3. LR-WGS facilitated reclassification of <i>FLNC</i> missense variant through expanded case-level evidence from additional affected individuals (NM09).....	119
4.1.4. LR-RNA demonstrated splice-altering effect of a previously identified <i>TTN</i> predicted splicing variant (NM12).....	128
4.1.5. Complementary LR-WGS and LR-RNA-seq approach allowed identification of biallelic <i>NUP188</i> variants with accurate direct phasing (NM05).....	141
4.2. Case Descriptions – Genomic findings of uncertain clinical significance	153
4.2.1. LR-WGS resolved a rare <i>PRX</i> structural variant of unknown significance (NM25)	153
4.2.2 LR-WGS characterized <i>ATXN8OS</i> STR expansion of uncertain clinical significance in father-son duo (NM02 and NM01)	165
4.3. Case Descriptions – Functional insights from disease-relevant tissue transcriptomic profiles	174
4.3.1. LR-RNA-seq from disease-relevant tissue identified mitochondrial transcriptomic dysregulation supporting a working diagnosis of mitochondrial myopathy (NM17).....	174
4.3.2. LR-RNA-seq identifies inflammatory and immune pathway activation (NM04)	181
Chapter 5: Discussion	189
5.1. Long-read genomic sequencing enables comprehensive variant detection and interpretation	192
5.2. WGS diagnostic power compared to WES	195
5.3. Comprehensive and reproducible methods support genetic re-analysis	196
5.4. LR-RNA-seq from disease-relevant tissue adds discovery and interpretive power to LR-WGS	197
5.5. Moving towards broader, clinical adoption of LR, multi-omics approaches	200
5.6. Limitations	202

5.7. Future Directions	205
5.8. Conclusions.....	206
References	209
Appendix.....	279

List of figures

Figure 1.1. Genetic diagnosis for individuals affected with NMDs provides numerous potential counselling, psychosocial, clinical and management benefits.....	28
Figure 3.1. NMD cohort ($n = 22$) demonstrates variety in body systems affected and results on standard clinical investigations performed	75
Figure 3.2. Rare disease NMD cohort ($n = 22$) demonstrates diversity of disease onset, features and clinical testing findings	76
Figure 3.3. NMD cohort ($n = 22$) demonstrates extended length of diagnostic odyssey and increased number of prior genetic testing	78
Figure 3.4. LR-WGS resolved GAA STR alleles in <i>FGF14</i> , with interruptions modulating pathogenicity	86
Figure 3.5. IsoRanker-guided LR-RNA-seq from skeletal muscle biopsy enables identification of genes of interest in individuals with known genetic diagnoses to be used as positive controls within the NMD cohort.....	90
Figure 3.6. Complementary LR-WGS and LR-RNA-seq improved diagnostic yield and identified functional disease-associated findings in a cohort of individuals with neuromuscular disorders.....	93
Figure 4.1. IGV snapshot of LR-WGS read alignments demonstrates 346kb <i>DMD</i> deletion in NM06 and NM03 (chrX:31,730,001-32,076,000).....	100
Figure 4.2. IGV snapshot of LR-WGS and LR-RNA-seq supports 346kb <i>DMD</i> deletion in NM03 (GRCh38 chrX:31,730,001-32,076,000).....	101
Figure 4.3. UCSC Genome Browser snapshot of <i>DMD</i> 346 kb deletion in NM03 and 06 demonstrates genomic overlap with ClinVar-reported deletions, and minimal overlap with common structural variants in healthy populations (NCBI dbVar)	105
Figure 4.4. IGV snapshot of NM21's long-read genome and RNA sequencing read alignments for NM21 demonstrate heterozygous missense <i>ORAI1</i> variant (chr12:121641154:G:A)	114
Figure 4.5. IGV snapshot of LR-WGS read alignments demonstrate a heterozygous missense <i>FLNC</i> variant in NM09 (chr7:128835550:G:A)	124
Figure 4.6. IGV snapshots of LR-WGS and LR-RNA-seq alignments in <i>TTN</i> demonstrate demonstrate homozygous splicing variant (chr2:178677618:T:C) with alterations to expression of exon 146 compared to other individuals in NMD cohort.....	136

Figure 4.7. LR-RNA-seq with IsoRanker identifies aberrant <i>TTN</i> splicing and decreased <i>TTN</i> gene expression in NM12 compared with other individuals in NMD cohort ($n = 19$).....	139
Figure 4.8. LR-RNA-seq with IsoRanker identifies reduced <i>NUP188</i> gene expression in NM05 compared with other individuals in NMD cohort ($n = 19$)	148
Figure 4.9. IGV snapshot of <i>NUP188</i> compound heterozygous variants in NM05 demonstrate ability of LR-WGS to accurately resolve variant phasing at lengths >1kb	149
Figure 4.10. Complementary LR-WGS and LR-RNA-seq identify 78bp <i>PRX</i> deletion (chr19:40,397,691-40,396,869) in NM25 which shows significant overlap with other deletions of unknown significance reported in ClinVar	160
Figure 4.11. Father (NM02) and son (NM01) duo with NMDs demonstrate heterozygosity for elevated CTG repeat number in <i>ATXN80S</i> , associated with spinocerebellar ataxia 8	172
Figure 4.12. IsoRanker-guided LR-RNA-seq identifies mitochondrial gene upregulation and increased alternative splicing in NM17 relative to NMD cohort ($n = 22$).....	179
Figure 4.13. Long-read transcriptomic analysis using IsoRanker identifies increased immune-related gene expression in NM04 relative to the neuromuscular disease (NMD) cohort ($n = 18$)	186
Figure 5.1. Integrated LR-WGS and LR-RNA-seq approach in rare disease cohort demonstrated clinical utility of this approach through improved diagnostic yield and identification and interpretation of multiple variant types, leading to a diagnostic rate of 6/22 (27%) of participants	190

List of tables

Table 1.1. LRS methods, notably LR-WGS, demonstrate significant clinical yield in select published studies of individuals and cohorts with NMDs and enable identification of variants not well captured by SRS methods	41
Table 1.2. RNA-seq methods demonstrate contributions to clinical yield in select published studies of individuals and cohorts with NMDs.....	47
Table 2.1. Inclusion and exclusion criteria for recruitment into NMD long-read multi-omic study	52
Table 2.2. Small variant predicted deleteriousness classifications and filtering thresholds, with variants predicted to have high functional impact (bolded) prioritized during LR-WGS filtering	59
Table 3.1. Select participant clinical and demographic characteristics ($n = 22$)	74
Table 3.2. Summary of prior genetic testing performed across NMD cohort ($n = 22$)	79
Table 3.3. Average number of small variants identified from LR-WGS (whole blood) in a cohort of individuals with NMDs ($n = 22$), following subsequent variant filtration steps	83
Table 3.4. Average number of SVs and CNVs identified from LR-WGS (whole blood) in a cohort of individuals with NMDs ($n = 22$), following subsequent variant filtration steps	83
Table 3.5. Average and total gene and isoform expression outliers in phenotypically relevant genes across NMD cohort using IsoRanker analysis ($n = 19$)	88
Table 4.1. Human Phenotype Ontology (HPO) terms selected for NM03 and NM06 based on their clinical presentation	96
Table 4.2. Summary of neuromuscular investigations in NM06 and NM03	97
Table 4.3. Pathogenicity and phenotype information for hemizygous <i>DMD</i> copy number deletion variant in NM03 and NM06.....	99
Table 4.4. Human Phenotype Ontology (HPO) terms selected for NM21 based on their clinical presentation	108
Table 4.5. Summary of neuromuscular investigations in NM21	110
Table 4.6. Allele frequency, pathogenicity evidence and phenotype information for heterozygous <i>ORAI1</i> variant in NM21. GRCh38 chr12:121641154:G:A, rsID not available.....	113

Table 4.7. <i>In silico</i> prediction tool values for variant <i>ORAI1</i> c.417G.A, p.(Met139Ile) (chr12:121641154:G:A) in NM21	113
Table 4.8. Human Phenotype Ontology (HPO) terms selected for NM09 based on their clinical presentation	120
Table 4.9. Summary of neuromuscular investigations in NM09.....	121
Table 4.10. Allele frequency, pathogenicity evidence and phenotype information for heterozygous <i>FLNC</i> missense variant in NM09	123
Table 4.11. <i>In silico</i> prediction tool values for variant <i>FLNC</i> c.577G>A (p. A193T) in NM09	125
Table 4.12. Human Phenotype Ontology (HPO) terms selected for NM12 based on their clinical presentation	129
Table 4.13. Summary of neuromuscular investigations in NM12.....	130
Table 4.14. Summary of genetic testing performed for NM12	131
Table 4.15. Allele frequency, pathogenicity evidence and phenotype information for homozygous <i>TTN</i> predicted splicing variant in NM12	133
Table 4.16. <i>In silico</i> prediction tool values for homozygous <i>TTN</i> predicted splicing variant in NM12	133
Table 4.17. Human Phenotype Ontology (HPO) terms selected for NM05 based on their clinical presentation	141
Table 4.18. Allele frequency, pathogenicity evidence and phenotype information for heterozygous <i>UNC45B</i> variant in NM05	144
Table 4.19. <i>In silico</i> prediction tool values for variant <i>UNC45B</i> c.712T>A, p.(S238T) in NM05	144
Table 4.20. Allele frequency, pathogenicity evidence and phenotype information for two compound-heterozygous <i>NUP188</i> variants in NM05	151
Table 4.21. Human Phenotype Ontology (HPO) terms selected for NM25 based on their clinical presentation	153
Table 4.22. Summary of genetic testing performed for NM25. Tests are ordered chronologically	158

Table 4.23. Pathogenicity and phenotype information for heterozygous <i>PRX</i> structural variant (loss) in NM25	159
Table 4.24. Human Phenotype Ontology (HPO) terms selected for NM01 and NM02 based on their clinical presentation.....	166
Table 4.25. Summary of neuromuscular investigations in NM02 and NM01	168
Table 4.26. Human Phenotype Ontology (HPO) terms selected for NM17 based on their clinical presentation	175
Table 4.27. Summary of neuromuscular investigations in NM17.....	176
Table 4.28. Human Phenotype Ontology (HPO) terms selected for NM04 based on their clinical presentation	181
Table 4.29. Summary of neuromuscular investigations in NM04.....	182
Table A1. VCF annotations and annotation descriptions for SNVs, as per SickKids TCAG	286
Table A2. VCF annotations and annotation descriptions for CNVs and SVs, as per SickKids TCAG.....	296
Table A3. List of 34 STR loci assessed using TRGT repeat analysis with LR-WGS data	304
Table A4. Summary of participant phenotypes, and LR-WGS and LR-RNA-seq outcomes in a cohort of 22 individuals with NMDs	306

List of abbreviations

ACMG	American College of Medical Genetics and Genomics
AD	Autosomal dominant
ALS	Amyotrophic lateral sclerosis
AR	Autosomal recessive
BMD	Becker muscular dystrophy
CADD	Combined Annotation-Dependent Depletion
CK	Creatine kinase
CM	Cardiomyopathy
CMA	Chromosomal microarray
CMT	Charcot–Marie–Tooth
CN	Copy number
CNV	Copy number variant
CRAC	Calcium-release activated calcium
DCM	Dilated cardiomyopathy
MT1/2	Myotonic dystrophy type 1/type 2
DMD	Duchenne muscular dystrophy
DSS	Dejerine-Sottas syndrome
EMG	Electromyography
ExAC	Exome Aggregation Consortium
FDA	Food and Drug Administration
FSHD	Facioscapulohumeral muscular dystrophy
GA4K	Genome Answers for Kids
gDNA	Genomic DNA
gnomAD	Genome Aggregation Database
GRCh	Genome Reference Consortium Human
GTE _x	Genotype-Tissue Expression
HMM	Hidden Markov Model
HPO	Human Phenotype Ontology
IGV	Integrative Genomics Viewer
KSS	Kearns–Sayre syndrome
LGMD	Limb-girdle muscular dystrophy
LR-RNA-seq	Long-read RNA sequencing
LRS	Long-read sequencing
LR-WGS	Long-read whole genome sequencing
MAF	Minor allele frequency
MD	Myotonic dystrophy

MLPA	Multiplex ligation-dependent probe amplification
MPC	Missense Badness, PolyPhen-2, and Constraint
MPS	Massively parallel sequencing
MRI	Magnetic resonance imaging
MT	MutationTaster
mtDNA	Mitochondrial DNA
NCBI dbVar	National Center for Biotechnology Information Database of Genomic Structural Variation
NCS	Nerve conduction studies
NEM	Nemaline myopathy
NMD	Neuromuscular disorder
OMIM	Online Mendelian Inheritance in Man
ONT	Oxford Nanopore Technologies
OPDM	Oculopharyngodistal myopathy
OPMD	Oculopharyngeal muscular dystrophy
PacBio	Pacific Biosciences
PAF	Population allele frequency
PCR	Polymerase chain reaction
REVEL	Rare Exome Variant Ensemble Learner
RIN	RNA integrity number
RNAi	RNA interference
RNA-seq	RNA sequencing
ROH	Runs of homozygosity
SANDSTEF	Sandestig-Stefanova syndrome
SBMA	Spinobulbar muscular atrophy
SCA	Spinocerebellar ataxia
SIFT	Sorting Intolerant From Tolerant
SMA	Spinal muscular atrophy
SMRT	Single-molecule real-time
SNV	Single nucleotide variant
SRS	Short-read sequencing
SR-RNA-seq	Short-read RNA sequencing
SR-WES	Short-read whole exome sequencing
SR-WGS	Short-read whole genome sequencing
STIR	Short tau inversion recovery (sequence)
STR	Short tandem repeat
SV	Structural variant
TAM	Tubular aggregate myopathy

TCAG	The Centre for Applied Genomics
TPM	Transcripts per million
TRGT	Tandem Repeat Genotyping Tool
TRVZ	Tandem Repeat Visualization
UCSC	University of California, Santa Cruz
UTR	Untranslated region
VCF	Variant Call Format
VUS	Variant of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing

Chapter 1: Background and review of the literature

1.1. Neuromuscular Disorders (NMDs)

1.1.1. *What are NMDs?*

NMDs are a heterogeneous group of disorders characterized by muscle impairment due to abnormalities in the peripheral nervous system, skeletal muscle or neuromuscular junction (Bhatt, 2016; Haskell et al., 2018; Mary et al., 2018). The unifying clinical characteristic of NMDs is abnormal muscle function, with the most common symptom being muscle weakness (Dowling et al., 2018; Khadilkar et al., 2024). Muscle weakness in NMDs can be chronic, resulting in ambulation difficulties, joint contractures, skeletal deformities, and respiratory failure, or episodic, manifesting as exercise intolerance, myalgia, rhabdomyolysis and fatigable weakness (Dowling et al., 2018). Additional features of NMDs include muscle cramps, atrophy or hypertrophy, myotonia, diminished tendon reflexes, ataxia and sensory features, such as paresthesia or sensory loss (Khadilkar et al., 2024). Respiratory failure and difficulty swallowing (dysphagia) are also common concerns in NMDs (Dziewas et al., 2021; Khan et al., 2023). Cardiac involvement is common in NMDs, with cardiomyopathies, conduction defects and arrhythmias occurring at varying frequencies in both pediatric and adult populations (Finsterer et al., 2013). Finally, extra-muscular or syndromic features may occur, including cognitive impairment, behavioral phenotypes, hearing loss and neurologic symptoms (Ebbink et al., 2012; Kieny et al., 2013; Orsini et al., 2018; Reimann & Kornblum, 2020).

As of May 2026, the GeneTable of Neuromuscular Disorders, an updated and curated list of known NMDs and associated genes, had described 1,261 distinct NMDs (Bonne et al., 2026). This elevated number of NMDs reflects the vast heterogeneity in symptom onset, progression and severity (Bonne et al., 2026). Depending on the NMD at hand, onset can occur in the neonatal,

childhood or adult periods (Gonzalez-Quereda et al., 2020). Disease progression may be slow, rapid or static, and physical disability is highly prevalent in NMDs (Barp et al., 2021; Bos et al., 2019). The clinical impact of NMDs is significant; while NMDs range in severity, prognosis is progressive and life-limiting for numerous NMDs (Chang & Wong, 2018). Fatality in NMDs may occur in infancy or childhood in severe early-onset disorders, or later in life as progressive neuromuscular impairment leads to respiratory or cardiac complications (Perrin et al., 2004; Rose et al., 2019; Voulgaris et al., 2019). Mortality in NMDs is largely owing to respiratory failure, resulting from weakness of the respiratory muscles and chest wall compliance, as well as cardiac failure (Perrin et al., 2004; Voulgaris et al., 2019). Although certain NMDs can be acquired through autoimmune reactions, viral infections and other origins, the vast majority of NMDs are genetic conditions (Zatz et al., 2016).

1.1.2. Epidemiology and impact; NMDs in a Manitoban and Canadian context

Individually, NMDs are rare conditions (Deenen et al., 2015); for example, Duchenne muscular dystrophy (DMD), an X-linked recessive condition, has a pooled global prevalence of 20 per 100,000 live male births, while spinal muscular atrophy (SMA) has a Canadian prevalence of approximately 8 per 100,000 live births (Crisafulli et al., 2020; Price et al., 2024). However, the collective prevalence of inherited NMDs is substantial (Deenen et al., 2025; Toscano et al., 2017). The collective prevalence across 24 NMDs was previously estimated at 160 per 100,000 individuals, comparable to the prevalence of better-known neurological disorders such as Parkinson's disease (Deenen et al., 2015). The true prevalence of NMDs is likely higher than this number due to ongoing discovery of disease-causing genes, improving access to genetic testing, and increased recognition and diagnosis of previously undiagnosed individuals (Deenen et al., 2015, 2025; Rose et al., 2019). Additionally, the prevalence of NMDs is thought to increase in the

coming years due to an aging international population, which is at risk of developing late-onset NMDs (Rose et al., 2019).

Founder variants are an important factor influencing the prevalence of genetic conditions, including NMDs. A founder variant is a pathogenic variant that occurs at an elevated frequency within a specific population as a result of geographic or cultural isolation, where one or more founding ancestors carried the variant (Wallace & Bean, 2023a). Multiple founder variants contribute to the prevalence of NMDs in Canada. For example, French-Canadian populations have elevated frequencies of pathogenic variants in *SACS* and *DMPK*, resulting in increased prevalence of autosomal recessive spastic ataxia of Charlevoix-Saguenay and myotonic dystrophy (MT) type 1, respectively (Bchetnia et al., 2021; Mathieu et al., 1990). In Manitoba, individuals in the Hutterite populations exhibit increased prevalence of limb-girdle muscular dystrophy (LGMD) due to pathogenic variants in several genes (Frosk et al., 2005; Frosk et al., 2005). Lastly, founder variants have also been identified across several Inuit, Cree, Ojibwe, Salteaux and Métis Nations (Leckie et al., 2021; Wallace & Bean, 2023a, 2023b; Wallace & Mirzaa, 2022). Founder effects have been reported for NMDs in First Nations populations in Manitoba, including *DYSF*-associated dysferlinopathies in Saulteaux and Chipewyan communities (Weiler, 2001; Weiler et al., 1996). Further, NMDs have shown significance in neighbouring regions, where fatal infantile myofibrillar myopathy is of elevated prevalence among First Nations, notably Cree, individuals across central and western Canada (Del Bigio et al., 2011). Meanwhile, spinobulbar muscular atrophy (SBMA) is of elevated prevalence among persons of Indigenous descent in Saskatchewan, though data about specific populations affected is not available (Leckie et al., 2021).

Diagnosis and clinical management of NMDs are significant to the Canadian healthcare system, including the Manitoban healthcare system. With approximately 70% of direct healthcare

expenditures, including diagnostic tests, covered by government funding, NMD care is economically relevant to both provincial and personal spending (Canada, 2018; McMillan et al., 2021). In a retrospective cohort study in Ontario from 2003 to 2014, Rose et al. (2019) found that NMD prevalence had increased in both adults and children by 8 and 10%, respectively, but that mortality decreased in both of these groups. Use of palliative care was increased throughout the years studied for both children and adults with NMDs; overall, 62% of adults and 50% of children who died due to their NMD had received palliative care, representing in total 21,743 individuals (Rose et al., 2019). The group also found that the number of children and adults with NMDs requiring hospital admission increased throughout 2003-2014 (Rose et al., 2019). Many individuals with NMDs require complex, long-term clinical support, which can include multidisciplinary care, rehabilitative and assistive technologies, and, in some cases, long-term mechanical ventilation, which can place substantial demands on acute care resources (Dale et al., 2018; Paganoni et al., 2017; Rose et al., 2015; Skalsky & Scott-Wyard, 2023). The collective impact of NMDs underscores the importance of continued investment in specialized diagnostic and clinical services to meet the complex and evolving needs of individuals with NMDs within the Canadian healthcare system.

1.1.3. Psychosocial significance of NMDs

Living with an NMD can have profound psychosocial impacts for affected individuals; among them, the occupational impact of disability is significant (Smith et al., 2025). Firstly, chronic fatigue and pain frequently decrease the endurance of sustained physical and mental activities, increasing the effort needed to participate in activities of daily living (Bos et al., 2015; Chaudhuri & Behan, 2004; Graham et al., 2011; Peric' et al., 2010; Torri et al., 2023). Secondly, physical and mobility limitations lead to loss of independence and accessibility challenges when

navigating physical spaces (Osman et al., 2025). Individuals with NMDs commonly experience barriers to employment, including inaccessible environments, lack of supports and resources, limited career options, and workplace discrimination (Lexell et al., 2017; Manzo et al., 2025). In total, resulting consequences of accessibility issues for individuals with NMDs include financial insecurity, social isolation, anxiety and depression and decreased quality of life (Landfeldt et al., 2014; Osman et al., 2025; Paradis et al., 1993; Winblad et al., 2010). In particular, children with physical disabilities demonstrate reduced access to social activities and are thus at increased risk of social isolation and withdrawal (Bushby et al., 2010; Law et al., 2006). Finally, individuals with NMDs may experience anticipatory anxiety and emotional distress regarding the future concept of disability, due to foreseen progressive loss of autonomy and dependency on medical devices (Manzo et al., 2025). Psychosocial significance of NMDs is also significant to caregivers of individuals with NMDs (Mekhuri et al., 2025). While caregiving on behalf of individuals with NMDs is often rewarding and positive, substantial caregiver burden also arises due to social, psychological, and financial factors (Magliano et al., 2014; Pangalila et al., 2012). For example, contributing factors to caregiver stress include continuous attendance at medical appointments and the need to advocate for appropriate health services for the individual with an NMD, especially before a diagnosis is achieved (de Moura et al., 2015). Further, caregiver awareness of disease trajectory for NMDs, which are incurable, progressive and often terminal, can have significant adverse effects on caregiver well-being (Landfeldt et al., 2016). Caregiver burden tends to increase with older age of the individual with NMD, due to the progressive nature of these conditions (de Moura et al., 2015).

Another important factor contributing to both patient and caregiver well-being in the context of NMDs is the time required to achieve a diagnosis. This concept, referred to as the

“diagnostic odyssey”, describes the period between an individual’s symptom onset and the time at which a definitive diagnosis is reached (Carmichael et al., 2015). Some individuals affected by NMDs, especially in the context of congenital conditions that can be easily recognized or detected through newborn screening programs, can obtain a diagnosis within several days or months of birth, such as SMA and congenital MT (Hilbert et al., 2013; Pera et al., 2020). For individuals with other NMD conditions, especially very rare diseases or those with more subtle, non-specific features, the diagnostic odyssey can last years (Hilbert et al., 2013). For example, individuals with myotonic dystrophy types 1 and 2 (MT1 and MT2) receive a diagnosis on average 7 and 14 years after symptom onset, respectively (Hilbert et al., 2013). The psychological impact of the diagnostic odyssey is important, as this time can be fraught with uncertainty, anxiety surrounding an unknown prognosis and lack of perceived legitimacy for the individual’s condition (Benito-Lozano et al., 2022; Nettleton, 2006; Ng et al., 2022). For Canadians affected with LGMD, their diagnostic odyssey was marked by misdiagnoses (e.g., inflammatory myopathy, polymyositis) and feeling dismissed by doctors or told that their symptoms were psychological or behavioural in nature (Osman et al., 2025). Their diagnostic journeys were lengthy and marked by emotional distress and feelings of hopelessness, depression, fear and anxiety (Osman et al., 2025). However, 92% (80/87) of participants in the study expressed acceptance of their diagnosis following genetic confirmation (Osman et al., 2025). Receiving a genetic diagnosis can invoke feelings of validation for affected individuals and families, providing support by identifying a label for their child’s condition (Carmichael et al., 2015). Having a genetic diagnosis can also help individuals or families to advocate for their needs and repair past feelings of dismissal by medical professionals (Carmichael et al., 2015). In sum, the diagnostic odyssey in NMDs represents a major source of

clinical and psychosocial burden, emphasising the critical need to streamline genetic diagnostic pathways and shorten time to definitive molecular diagnosis.

1.2. NMDs as genetic disorders

1.2.1. Genetics of NMDs

Most NMDs are inherited disorders caused by genetic variants in individual genes, both occurring *de novo* and through Mendelian patterns of inheritance, although some polygenic or multifactorial NMDs exist (Luglio et al., 2024; Zatz et al., 2016). NMD genes are often large and structurally complex (Martin et al., 2026). For instance, the Duchenne muscular dystrophy (*DMD*) gene spans more than 2.3 Mb and contains 79 exons, while titin (*TTN*) encompasses 363 exons with an open reading frame exceeding 100 kb (Chae et al., 2015). In addition, these genes harbor a high degree of polymorphism, complicating the discrimination of benign from pathogenic variants (Efthymiou et al., 2016). Compared to non-NMD genes, those implicated in NMDs often exhibit greater structural and transcriptional complexity, with longer transcripts and more isoforms (Martin et al., 2026). Frequently altered genes implicated in NMDs include *TTN*, *ANO5* and *RYR1*; in a recent large-scale NMD gene panel study, variants in these genes accounted for nearly 30% of diagnosed cases (Barbosa-Gouveia et al., 2022). In another study, variants were also commonly found in the *COL6* (i.e., *-A1*, *-A2*, *-A3*), *LMNA*, *SCN4A* and *SELENON* genes (Westra et al., 2019).

Intronic variants are an increasingly recognized cause of NMDs, which can cause disease by altering the abundance of normally or aberrantly spliced gene products in affected tissues (Anna & Monika, 2018; Cummings et al., 2017). Further, non-coding variants are an important, albeit infrequent, cause of genetic pathology in inherited NMDs (Baskin et al., 2009; Kevelam et al., 2015; Ruggieri et al., 2015). This includes variants that lie outside the intron/exon boundary, such as deep intronic or regulatory region variants. Interpretation of non-coding variants remains challenging due to limited understanding of intronic and regulatory sequences (Marwaha et al., 2022). Thus, non-coding variants are often difficult to prioritise in genetic testing, requiring

additional functional or transcriptomic evidence to establish pathogenicity (Marwaha et al., 2022; Wright et al., 2018).

Complex genetic alterations are also significant contributors to NMD etiology. Among these, structural variants (SVs) are gene variations of 50 base pairs (bp) or more, including deletions, insertions, duplications, translocations and more complex rearrangements (Mahmoud et al., 2019). Copy number variants (CNVs) are a subset of SVs that involve only gains or losses in DNA copy number (i.e., deletions and duplications) (Carvalho & Lupski, 2016). Both SVs and CNVs can range in size from a single exon to several megabases, resulting in gene dosage changes (Feuk et al., 2006; Sudmant et al., 2015). The identification of SVs is particularly helpful for accurately detecting and diagnosing AR conditions, which can result from compound heterozygous alleles of different variant types (Winder et al., 2020). In a gene panel study, detected SVs accounted for 39% of all clinically significant variants (Winder et al., 2020); this finding highlights the need for routine SV analysis in NMD genetic testing. While 93% of disease-causing SVs occurred in either *SMN1*, *PMP22* or *DMD*, the remaining 7% of clinically significant SVs were found across 77 genes, contributing to genetic diagnosis for 113 participants (Winder et al., 2020). Further, over 80% of the unique SVs detected in this study included only a few exons, emphasizing the need for high-resolution testing methods (Winder et al., 2020). Overall, there is substantial contribution of SVs to NMD pathology and a need to incorporate reliable SV detection into routine genetic diagnostic workflows (Stavropoulos et al., 2016).

Another key variant type contributing to NMDs is short tandem repeats (STRs). STRs are repetitive sequences of 2-6 bp (“repeats”) in succession, which make up a significant (~7%) of the human genome (Shortt et al., 2020). STRs are highly dynamic and polymorphic, and individuals will have variations in the number of repeats (“repeat length”) at a given locus (Chintalaphani et

al., 2021). STRs are prone to alteration during DNA replication and are consequently unstable, with higher genetic variation rates compared to single-nucleotide variants (SNVs) (Hannan, 2012). This instability arises from strand misalignment in STR regions during replication or transcription, including strand slippage and formation of secondary structures, resulting in the gain or loss of repeat units (Verbiest et al., 2023). Longer repeat lengths are more unstable, with an increased propensity to expand over successive generations and a greater risk of becoming pathogenic (Chintalaphani et al., 2021). STRs play significant roles in pathogenic variation, in which expansions can alter gene expression or splicing or generate toxic protein products (Chintalaphani et al., 2021). Indeed, STR expansions are enriched in NMD-related genes, pathogenic variation of which are associated with over 60 neurological disorders and NMDs (Martin et al., 2026; Stevanovski et al., 2022). For example, MT1 and MT2 are caused by STR expansions in *DMPK* and *CNBP*, while spinocerebellar ataxias (SCAs) are frequently caused by STR expansions, including in *FGF14*, *ATXN8OS*, *RFC1*, *FXN* and other genes (Müller, 2021).

Variants of uncertain significance (VUS) are a persistent and significant challenge in interpreting clinical genetic test results for inherited disorders, including NMDs (Jones & Moore, 2020). VUS arise when available evidence is insufficient to classify a variant as either pathogenic or benign under established interpretation frameworks (Richards et al., 2015). Importantly, the likelihood of VUS detection in a given genetic analysis increases with the number and length of genes tested; given that some of the largest human genes are associated with NMDs (e.g., *DMD*, *TTN*, *NEB*, *RYR1*), VUS are a frequent challenge in NMD diagnostics (Martin et al., 2026; Perez, 2024). VUS are observed in an estimated 42-53% of comprehensive sequencing strategies for NMDs (i.e. large gene panel and whole exome sequencing (WES)), highlighting their elevated occurrence in clinical diagnostics for NMDs (Johnson et al., 2024; Winder et al., 2020). Among

individuals with one or more VUS, 16% also had a co-occurring definitive diagnostic result, indicating that the VUS was likely unrelated to their clinical phenotype and underscoring the difficulty of distinguishing causative from background genetic variation (Winder et al., 2020).

1.2.2. Genetic testing in NMD diagnostics

Genetic diagnosis can provide a variety of potential benefits to individuals affected with NMDs (**Figure 1.1**). Firstly, genetic testing can provide a clear diagnosis for an individual's condition, when clinical diagnosis may otherwise be challenging. Clinical diagnosis is complicated by heterogeneity among individuals with the same NMD, as well as by clinical similarities among individuals with different NMDs (Rosenberg et al., 2023). Further, clinical diagnosis in the pediatric population can be challenging due to limited information based on age and ongoing development of the clinical presentation (Tian et al., 2015). Additionally, the rarity of some NMDs can limit clinician recognition of the disorder (Mansfield et al., 2024). Altogether, affected individuals can often present with nonspecific features, symptoms and laboratory findings (Haskell et al., 2018). In this context, genetic testing can be a valuable first-line test for NMDs (Córdoba et al., 2018). A genetic diagnosis can shorten the diagnostic odyssey and limit clinical testing that might otherwise be used to reach a diagnosis (Córdoba et al., 2018). Importantly, genetic diagnosis can sometimes remove need for invasive diagnostic muscle or nerve biopsies, thereby preventing both unnecessary procedures to patients and reducing healthcare costs (Schofield et al., 2017). For example, when studying a small cohort of individuals with NMDs who had received a genetic diagnosis, 6/16 (37%) had undergone prior diagnostic procedures which were uninformative in pointing to a presumed diagnosis (Haskell et al., 2018). In the remaining 10/16 (63%) individuals, while prior testing had been informative, it had not always succeeded in establishing a suspected diagnosis prior to WES (Haskell et al., 2018). Considering that medical procedures can be

distressing for children and that many NMDs are childhood onset, the use of genetic tests to bypass certain clinical testing is significant (Carmichael et al., 2015). In summary, pathogenic variant identification can supply valuable diagnostic information for NMDs, either informing or substituting clinical testing.

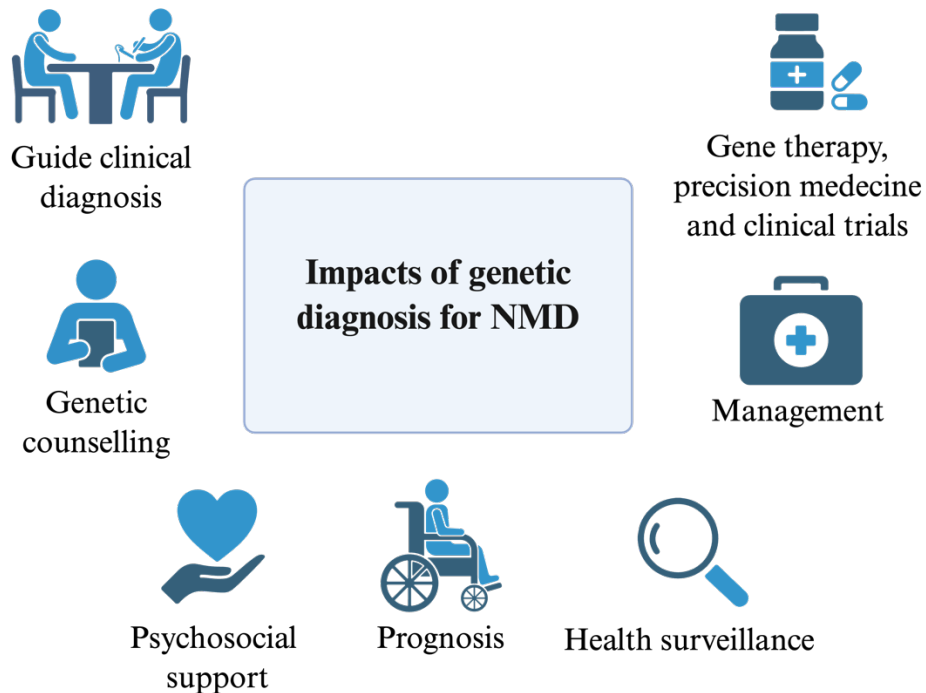


Figure 1.1. Genetic diagnosis for individuals affected with NMDs provides numerous potential counselling, psychosocial, clinical and management benefits.

1.2.3. Genetic diagnosis for NMDs: counselling and risk assessment

Genetic testing and diagnosis are central to the efficacy of genetic counselling for NMDs. With a genetic diagnosis, the counsellor can communicate prognostic information and facilitate understanding of the disease in the individual's specific personal and family circumstances (Roggenbuck et al., 2023). As NMDs are associated with all forms of inheritance patterns, a genetic diagnosis allows clear identification of at-risk relatives and organization of cascade testing (Zatz et al., 2016). Diagnosis can connect patients and caregivers to relevant support groups, which can reduce isolation and stress and increase self-efficacy (Meade et al., 2018; Mekhuri et al., 2025). Especially in the context of rare disease, a genetic diagnosis may facilitate connections with other individuals and families affected by the same condition (Esquivel-Sada & Nguyen, 2018)

Genetic diagnosis is important for identifying appropriate long-term surveillance for affected individuals. For example, given common cardiac involvement in many NMDs, genetic diagnosis can support implementation of a cardiac surveillance or management strategy specific to that condition (Finsterer & Stöllberger, 2016). Among such NMDs, notably DMD, appropriate cardiac treatment significantly improves long-term health outcomes (Feingold et al., 2017; Finsterer & Stöllberger, 2016). On the contrary, cardiac involvement is not a recognized component of some other NMDs; for example, individuals with oculopharyngeal muscular dystrophy (OPMD) or facioscapulohumeral muscular dystrophy (FSHD) are not recommended asymptomatic cardiac surveillance and could avoid ongoing cardiac evaluation by accurate genetic diagnosis (Preston & Wang, 1993; Trollet et al., 1993). Further, while LGMD comprises over 30 diverse conditions affecting the proximal musculature, certain subtypes (e.g. LGMD D1 DNAJB6-related) have cardiac involvement and recommended screening, while other subtypes (e.g., LGMD R1 Calpain3-related) do not (Narayanaswami et al., 2014). Finally, as genetic conditions can have

multi-systemic involvement, NMDs which are deemed syndromic by genetic diagnosis, can enable surveillance of those other systems, including ophthalmic, audiology and immunological health management (Dziewas et al., 2021; Karnebeek et al., 2022; Moshirfar et al., 2022).

1.2.3. Genetic diagnosis for NMDs: Treatment and management

Genetic diagnosis is valuable for establishing the most effective available treatment for an NMD. In one sense, clarifying that the etiology of the condition is in fact genetic can prevent unnecessary or harmful medical treatment for alternative causes. For example, in one WES study, a patient thought to have a chronic inflammatory demyelinating polyneuropathy was in fact diagnosed with spastic ataxia of Charlevoix-Saguenay and thus not started on immunotherapy (Haskell et al., 2018). For NMDs with effective treatments, early clinical intervention guided by genetic diagnosis is key to improving outcomes. For example, newborn screening programs for SMA, implemented in Canada and across other countries, have allowed for early or presymptomatic clinical intervention; long-term outcomes in treated individuals reflect the importance of early intervention in maximizing motor and non-motor outcomes, and mitigating life-threatening events (Karachunski et al., 2022; Wang et al., 2025; Wong et al., 2024). In Canada, early identification and treatment of SMA are expected to reduce public healthcare costs and improve quality of life and lifespan for affected individuals (Pace et al., 2025).

In addition to access to established treatments, a genetic diagnosis can enable participation in clinical trials and access to emerging treatment options (Ng et al., 2022). For example, eligibility for many FDA-approved personalised therapies requires confirmation of genetic diagnosis. In fact, given the access to targeted drugs in the cases of NMDs such as amyotrophic lateral sclerosis (ALS), genetic testing is recommended early in disease onset (Müller et al., 2022; Salmon et al., 2022). Timely genetic diagnosis is especially important in the pediatric context, in which

progression of disease can be rapid, and treatment timing is therefore critical (Bick et al., 2019). In contrast, there remain many NMDs for which there is no cure: by identifying such a genetic diagnosis, medical professionals can provide a prognosis and support affected individuals with the reality of living with an incurable disease (Chu & Moran, 2018).

Further, genetic diagnosis of an NMD is especially critical given the increasing availability of gene therapies. Gene therapy aims at treating or preventing a genetic disease by modifying the genetic material within an individual's cells, delivered by viral vectors, antisense oligonucleotides or small molecule drugs. These therapeutics have revolutionized the treatment of several NMDs, including DMD and SMA (Wu et al., 2025). For example, in the context of Charcot-Marie-Tooth (CMT) disease subtypes, knowledge of the distinct genetic etiology has supported the ongoing research and development of targeted therapies, largely in murine models to date (Stavrou et al., 2023). While loss-of-function mechanisms (e.g., *GJB1* variants in CMT X-linked dominant 1) may be effectively treated by gene replacement, dominant toxic gain-of-function mechanisms (e.g., *PMP22* overexpression in CMT1A) are treated with gene silencing or editing approaches (Gautier et al., 2021; Kagiava et al., 2021; Lee et al., 2020; Stavrou et al., 2022). Even for more complex CMT etiology, such as the dominant negative *MNF2*-variant-associated condition, combined RNA interference and gene therapy approaches have been proposed to restore functional gene expression (Rizzo et al., 2023). As gene therapy continues to gain traction across rare disease, early and effective genetic diagnosis will be critical for identifying individuals for which a gene therapy is available.

1.3. Current genetic testing methods for NMDs

The traditional approach to diagnosing NMDs includes clinical phenotyping, biochemical tests and tissue phenotyping, via muscle magnetic resonance imaging (MRI), electromyography

(EMG), or muscle biopsy; these analyses typically inform type of genetic testing performed afterwards (Aleem et al., 2025). An early model of genetic testing for NMDs included sequential single-gene testing using low throughput Sanger sequencing (Sanger et al., 1977). However, as the number of NMD genes to sequence grew and the challenge of phenotypic heterogeneity in NMDs remained, Sanger sequencing of individual genes tended to be time-consuming and cost-prohibitive (Chae et al., 2015). Technological advances in genetics, allowing for a more comprehensive testing approach, would revolutionize testing in the following years.

1.3.1. Massively parallel sequencing (MPS) for rare human disease

Massively parallel sequencing (MPS) is a high-throughput approach which allows for the rapid sequencing of large amounts of nucleic acids (Chae et al., 2015). In this approach, DNA is fragmented into pieces, and millions of these fragments are sequenced simultaneously to generate sequence reads (Knierim et al., 2011; Qin, 2019). Base calling occurs, translating raw instrument signals into a DNA or RNA nucleotide sequence (Ledergerber & Dessimoz, 2011). Next, bioinformatic analysis allows for the reads to be aligned to a human reference genome sequence, a process called read alignment (Qin, 2019). Lastly, by identifying differences between the reference and tested sequences, genetic variants are detected (Qin, 2019). Altogether, MPS is crucial in the context of genetic testing for NMDs, enabling the simultaneous analysis of a much larger number of genes relevant to the individual's phenotype. MPS using a short-read sequencing (SRS) approach generates millions of short sequence reads (50-300 bp) simultaneously, whereas Sanger sequencing produces a single read of up to 800 bp per reaction. MPS can also detect somatic mosaicism, that is, the presence of two or more cell lines in an individual that are genotypically distinct (Cao et al., 2019; Qin et al., 2016), while Sanger is unable to detect alleles below 15-20% mosaicism (Rohlin et al., 2009). MPS can be separated into two formats, distinguished by the

length of the DNA fragments generated and sequenced. SRS generates reads typically under 300 bp, while long-read sequencing (LRS) is an emerging technology that produces reads over 10 kilobases (kb) in length (Logsdon et al., 2020). SRS is the main strategy currently employed in both clinical and research settings, owing to its lower cost per data unit (Eisenhofer et al., 2024). As such, most gene panels, WES and whole genome sequencing (WGS) are currently performed using SRS (Eisenhofer et al., 2024; Merker et al., 2018).

1.3.2. Gene panels

Gene panels are laboratory tests that assess a subset of genes associated with a particular phenotype (Bean et al., 2020). By investigating multiple genes associated with NMDs, gene panels allow for faster and more cost-efficient turnaround times compared to sequential Sanger sequencing of individual genes (Lek & MacArthur, 2014; Schofield et al., 2017). Large gene panels ($n > 50$) also reduce clinician bias in testing and improve diagnosis for atypical presentations of a given condition (Antoniadi et al., 2015; Hu et al., 2018). Indeed, gene panels are often applied as a first-tier genetic test when assessing patients with suspected NMD (Kitamura et al., 2016; Kuhn et al., 2016). In a large study analyzing over 25,000 unrelated patients with a suspected genetic NMD, a 266-gene panel approach yielded a diagnosis in 20% (5055) of the individuals (Winder et al., 2020). In another study, use of a 324-gene panel to analyze 268 Spanish and Portuguese individuals with a suspected inherited NMD yielded a diagnostic rate of 40% (Barbosa-Gouveia et al., 2022).

Altogether, the diagnostic yield of SRS-based gene panels ranges between 15-49%, depending on the size of the panel and the phenotypic heterogeneity of the recruited participants (Beecroft et al., 2020; Chae et al., 2015; Gonzalez-Quereda, et al., 2020; Park et al., 2019). Conversely, more than half of individuals with NMDs remain undiagnosed after gene panel testing,

underscoring the limitations of panels to fully capture the genetic complexity of these disorders. VUS are a potential source of limitations in gene panels. However, reanalysis of previously identified VUS only shows a partially increased diagnostic rate for gene panels, thereby indicating that number of genes analyzed and technical aspects of the testing method may limit the detection of causative gene variants (Schobers et al., 2022).

1.3.3. Whole exome sequencing (WES)

WES examines the protein-coding genome, which accounts for approximately 1-2% of the whole genome and up to 60% of disease-causing genomic variation, reflecting the central role of coding sequences in gene function (Marian, 2014; Ross et al., 2020). WES typically achieves sensitive detection of SNVs and small insertions/deletions, due to targeted enrichment of coding regions and high sequencing depth across exons (Meienberg et al., 2016). Further, WES can detect variants across 99% of the exome, thereby analyzing approximately 20,000 protein-coding genes rather than a predefined subset selected for a gene panel (Kreiman & Boles, 2020). As such, WES represents an important step moving beyond phenotype-based selection towards comprehensive genetic testing approaches, such as WGS (**Section 1.3.2**). Such exome- or genome-wide approaches perform three important functions. Firstly, comprehensive methods help to unveil variants in genes not yet associated with NMDs, thereby expanding knowledge of genetic variation (Lee et al., 2021; Taylor et al., 2015). Second, standardized and unbiased approaches can diagnose atypical presentations of previously recognized NMDs, thereby expanding knowledge of phenotypic variation (Lee et al., 2021; Posey et al., 2016). Third, exome- and genome-wide approaches can facilitate genetic testing for individuals with highly non-specific clinical presentations (Lee et al., 2021; Srivastava et al., 2019).

Altogether, the comprehensive nature of WES and WGS positions them as powerful tools to potentially replace time- and cost-consuming stepwise genetic testing approaches (Lee et al., 2021). While individual targeted gene panels generally have lower upfront costs than exome- or genome-wide testing, serial testing following non-diagnostic gene panel results can increase cumulative costs, ultimately meeting or exceeding those of comprehensive sequencing approaches (Li et al., 2021; McElwee et al., 2026; Tagimacruz et al., 2026). For example, at the British Columbia Children's Hospital in 2024, the cost of individual gene panel testing for developmental and seizure disorders could exceed C\$1,300, while the cost of WES and WGS ranged from C\$2,000-4,600 and C\$3,100-C\$7,700, respectively (Regier et al., 2024). Due to the relatively affordable price of WES, estimated at C\$2,888.79 for laboratory workflow and sequencing for trio-WES in 2026, it is favoured in some clinical settings for use either when a diagnosis has not been obtained through other testing methods or as a first-tier test, especially for complex or atypical phenotypic presentations (Piñeros-Fernández et al., 2024; Ungar et al., 2026). Reported diagnostic yields for suspected Mendelian disorders using WES vary with patient selection, age and presence of consanguinity, ranging from 12-45% (Ghaoui et al., 2015; Haskell et al., 2018; Herman et al., 2021; Waldrop et al., 2019; Westra et al., 2019). Trio WES can further increase diagnostic rates for individuals with rare inherited diseases, that is, using WES data from an affected child's parents to identify and interpret *de novo* or compound heterozygous variants (Ghaoui et al., 2015; Lee et al., 2014). For some patients, re-analysis of WES data is performed two to three years after a negative or uncertain result and provides an additional 10-15% diagnostic yield compared to the initial WES (Fung et al., 2020; Gibson et al., 2018; Wenger et al., 2017).

However, important limitations of WES remain. Firstly, while WES can capture some information outside of the coding regions, such as some untranslated regions and intron-exon

boundaries, it cannot capture all deep intronic variants, which have been implicated in NMD pathology (Guo et al., 2012; Marwaha et al., 2022). Further, due to its capture-based methodology, WES demonstrates uneven coverage across various coding regions, particularly GC-rich first exons, GC/AT-rich regions and low-complexity sequences, including STRs (Meienberg et al., 2016; Wang et al., 2017). The uniformity of coverage distribution decreases for longer exons; this is an important consideration for long genes, including several NMD-related genes, such as *DMD*, which spans over 2.2 Mb across 79 exons (Wang et al., 2017). In conclusion, while WES shows an increase in diagnostic yield compared to gene panels, across these two traditional sequencing methods, over 50% of individuals with NMDs have not received a genetic diagnosis, despite family inheritance patterns suggesting a pathogenic variant in a single gene (Efthymiou et al., 2016; Thompson et al., 2020). Research into alternative methods, including WGS, LRS and complementary transcriptomics, may allow for improved diagnostic rates.

1.4. Comprehensive genetic testing methods for NMDs

1.4.1. Whole genome sequencing (WGS)

WGS is a high-throughput genetic sequencing method that interrogates up to 98% of the human genome (Alfares et al., 2018). As with WES and gene panels, most WGS is performed using SRS. Similar to WES, the comprehensive, unbiased nature of WGS enables diagnosis for patients with atypical or unspecific NMD presentations, or for those who harbor variants in novel disease genes (**Section 1.2.4**). However, by investigating across all coding and non-coding regions, WGS can comprehensively capture a broader spectrum of genetic variation. This includes variants not assessed by WES (variants in introns, intergenic regions and regulatory elements) and those not reliably assessed by WES (complex alterations spanning several genes) (Mitsuhashi & Matsumoto, 2020). Further, WGS shows more uniform coverage compared to WES across exons,

especially in the first exons (due to high GC content), other GC/AT-rich regions, and low-complexity sequences (Lee et al., 2021; Marwaha et al., 2022). WGS enables more reliable detection of CNVs, as their content and breakpoints often extend beyond coding regions (Belkadi et al., 2015; Fromer et al., 2012). Finally, WGS data include both nuclear and mitochondrial genome data, allowing for the detection of mitochondrial disorders (Meienberg et al., 2016). Altogether, WGS can surmount capabilities of both phenotype-directed gene panels and exonic analyses. Indeed, clinical studies have demonstrated the diagnostic value of WGS in rare disease populations, including NMDs. For example, short-read whole genome sequencing (SR-WGS) demonstrated a diagnostic rate of 29.3% (218/744 families) in a recent study (Wojcik et al., 2024). 8.2% of families (28% of those with a positive diagnosis) had variants that required WGS for identification, including coding variants, intronic variants, small SVs, copy-neutral inversions, complex rearrangements and tandem repeat (Wojcik et al., 2024). In a cohort of 247 families with suspected genetic NMDs, SR-WGS following negative short-read whole exome sequencing (SR-WES) increased the diagnostic rate from 55% (83/152) to 62% (94/152) (Marchant et al., 2024).

Previously, the higher cost of WGS has been a barrier to clinical implementation, exacerbated by increased demand for computational variant calling and interpretation (Tsiplova et al., 2017). Recently, the decreased cost and turnaround time of the technology, as well as increased resources in data analysis, have allowed WGS to be a possibility for use in the clinical setting (Alfares et al., 2018; Bagger et al., 2024; Clark et al., 2018; Saunders et al., 2012). Clinical indications for WGS include clinical features known to be associated with multiple conditions or genes, lack of correspondence to a specific clinical disorder, and atypical or complex clinical course (ACMG Board of Directors, 2015; Bick et al., 2019). WGS may also be important in pediatric cases, wherein neonates and children may present fewer specific clinical features,

allowing for specific genetic testing (Bick et al., 2019). Many argue for the use of WGS as a first-line strategy to capture the majority of clinical indications in genetic diagnostics (Schobers et al., 2024). However, although SR-WGS offers broader and more uniform coverage than SR-WES, its reliance on SRS limits the ability to resolve complex genomic regions and classes of genetic variation, thereby constraining diagnostic yield (Hayeems et al., 2026). Alternatively, long-read WGS (LR-WGS) has emerged as an approach with the potential to overcome these technical limitations of SR-WGS and improve diagnostic outcomes (Marwaha et al., 2022).

1.4.2. Long-read sequencing (LRS)

The limitations of SRS-based approaches arise from the use of short read lengths (50-300bp), which make it difficult to accurately map sequencing reads to the reference genome, especially within complex and repetitive genomic regions (Logsdon et al., 2020). Consequently, areas containing segmental duplications, high GC content and repetitive elements are not sequenced reliably using SRS (Marwaha et al., 2022). This limitation includes, importantly, the inability of SRS to reliably identify STR expansions, especially those which are undiscovered and/or in long and complex GC-rich regions (Marwaha et al., 2022). Furthermore, SRS cannot reliably determine repeat sequence composition (motif type) or identify internal repeat interruptions, both of which may be critical for establishing pathogenicity (Ebbert et al., 2018; Stevanovski et al., 2022). Additionally, while SRS is adept at capturing small variants (e.g., SNVs), SVs and CNVs are not reliably captured when they are larger than the fragment size itself (Marwaha et al., 2022; Moustakli et al., 2025). While improvements in computational tools, such as CNV callers and repeat expansion detection algorithms, have expanded the utility of SRS, a sequencing approach that can comprehensively resolve more diverse variant types and genomic areas is needed (Dolzhenko et al., 2024; Fearnley et al., 2022; Zhu et al., 2012).

LRS generates longer read lengths of approximately 10 to over 100 kb depending on the sequencing platform, allowing it to overcome many limitations encountered with SRS (Logsdon et al., 2020). The increased read length improves alignment of sequencing reads to the reference genome and enables more complete genome assembly, particularly across repetitive and structurally complex regions (Chaisson et al., 2015; Logsdon et al., 2020). Notably, long reads can span across or even encompass SVs, low-complexity repetitive loci, duplicated regions and STRs, allowing these regions and variant classes to be more reliably characterized compared with SRS (Warburton & Sebra, 2023). The long read length can also help to distinguish between highly homologous genomic regions, such as the highly homologous NMD-associated genes *SMN1* and *SMN2* (Borràs et al., 2017; Long et al., 2024; Zeng et al., 2025).

Beyond improving resolution of complex genomic regions, LRS provides additional advantages through its ability to sequence long, intact DNA molecules (Marwaha et al., 2022). The extended read length of LRS enables improved variant phasing, that is the determination of whether two variants are on the same allele (in cis) or on opposite alleles (in trans) (Maestri et al., 2020; Tewhey et al., 2011). SRS rarely enables direct phasing because individual reads (fewer than 300 bp in length) are typically too short to span more than one variant (Mantere et al., 2019). In contrast, LRS improves phasing resolution by enabling direct read-based phasing of variants within the span of a single long read (approximately 10-20 kb in this study), as well as indirect phasing across haplotype blocks formed by overlapping reads (Maestri et al., 2020). Improved phasing resolution is vital for the interpretation of compound heterozygous variants underlying recessive disease mechanisms, which have been implicated in NMDs (Zhao et al., 2025). Moreover, unlike SRS, which typically sequences fragmented DNA molecules, LRS directly sequences intact native molecules (Logsdon et al., 2020). In doing so, LRS preserves genomic

context and enables additional information to be derived, including epigenetic information (e.g., methylated nucleotides) (Gigante et al., 2019; Ivashchenko et al., 2026). Direct sequencing of native molecules also eliminates amplification bias and errors which can occur in some PCR-based SRS methods (Amarasinghe et al., 2020).

Altogether, LRS is a powerful emerging technology that enables comprehensive characterization of genomic regions and variant types that are difficult to resolve using SRS, while also providing additional information through improved phasing and detection of epigenetic modifications (Marwaha et al., 2022). Indeed, LRS has been successful in identifying pathogenic variants in cases where SRS was unrevealing (Ishiura et al., 2018; Sone et al., 2019), including studies for NMDs (Bruels et al., 2022; Xie et al., 2020) (**Table 1.1**). There are currently two principal LRS platforms used for genomic applications. The first, PacBio single-molecule real-time (SMRT) sequencing, generates highly accurate (99.8%) high-fidelity (HiFi) reads of 10-30 kb in length through circular consensus sequencing of individual DNA molecules (Logsdon et al., 2020; Wenger et al., 2019). Conversely, Oxford Nanopore Technology sequences native DNA molecules by measuring changes in electrical current as DNA molecules pass through protein nanopores (Jain et al., 2018). ONT can generate substantially longer reads than PacBio, including ultra-long reads exceeding 100kb, although historically with lower per-read accuracy compared with PacBio HiFi sequencing (Logsdon et al., 2020).

Table 1.1. LRS methods, notably LR-WGS, demonstrate significant clinical yield in select published studies of individuals and cohorts with NMDs and enable identification of variants not well captured by SRS methods. Ordered by year of publication.

Study	Condition(s) investigated	Cohort size	Sequencing type (platform)	Variants identified	Clinical impact of LRS
SVs and splice variants					
Xie et al., (2020)	Dystrophinopathy	1 individual	LRS of <i>DMD</i> (ONT)	Complex inversion/deletion-insertion rearrangement (<i>DMD</i>)	Resolved pathogenic variant missed by routine clinical testing ¹
Bruels et al., (2022)	Muscular dystrophy	12 individuals from 10 families	LR-WGS (ONT)	Duplication (<i>LAMA2</i>) Inversion (<i>DMD</i>) Intronic splice variants (<i>DMD</i>) Small indel (<i>SMCHD1</i>)	Identified pathogenic variants missed by previous SRS in 50% (5/10) of probands Confirmed all SNVs identified in previous SRS
Geng et al., (2023)	DMD	2 individuals	LRS of <i>DMD</i> (ONT)	Large inversions (<i>DMD</i>)	Identified SVs not captured in previous testing ² Precisely mapped breakpoints
Li et al., (2023)	SMA	79 subjects from 24 families ³	LRS of <i>SMN1</i> & <i>SMN2</i> (PacBio)	<i>SMN1</i> copy number variants with haplotype phasing	Accurate haplotype phasing of <i>SMN1</i> copy number in affected individuals and carrier parents
Repeat expansions					
Cumming et al., (2018)	MT1	3 unrelated individuals	LR-WGS (PacBio SMRT)	CTG repeat expansion with interruptions (3' UTR of <i>DMPK</i>)	Enabled complete allele characterization, including repeat interruptions Identified repeat interruptions as modifiers of disease
Sone et al., (2019)	NIID	9 families & 40 sporadic cases	LR-WGS (PacBio and ONT)	GGC repeat expansion (<i>NOTCH2NLC</i>)	Identified novel disease-causing repeat expansion

Yu et al., (2021)	OPDM3	4 families & 7 sporadic cases	LR-WGS (ONT)	GGC repeat expansion (<i>NOTCH2NLC</i>)	Identified pathogenic repeat expansion in 16.7% of 24 unrelated individuals with OPDM
Mangin et al., (2021)	MT1	2 families	LRS (PacBio)	Large CTG repeat expansions ($\geq 2,000$ repeats)	Resolved complete sequence, including repeat numbers and interruptions
Yu et al., (2022)	OPDM4	2 individuals	LR-WGS (ONT)	CGG repeat expansion (5' UTR of <i>RILPL1</i>)	Improved diagnosis compared with SR-WES
Yeetong et al., (2023)	FSHD	7 unrelated individuals	LR-WGS (ONT)	Repeat contraction (D4Z4 region)	Identified pathogenic repeat contractions in all affected individuals
Cortese et al., (2024)	OPDM	19 individuals from 6 families	LRS of <i>ABCD3</i> 5' UTR (ONT)	CCG repeat expansion	Identified pathogenic repeat range and established novel repeat expansion disorder
Repetitive or difficult-to-sequence loci					
Perrin et al., (2022)	<i>TTN</i> -related myopathies	6 unrelated individuals	LRS of <i>TTN</i> (ONT)	Small variants: frameshift, non-sense, splicing	Accurately localized known <i>TTN</i> variants in complex repeat domains
Huang et al., (2024)	FSHD	12 unrelated individuals	LR-WGS (ONT)	Repeat contraction (D4Z4 region), including low-level mosaic variant	Effectively differentiated homologous regions Enabled allele-specific methylation analyses

¹MLPA-based deletion-duplication analysis and SRS for *DMD*; ²MLPA-based deletion-duplication analysis for *DMD*, SR-WES, SR-WGS, karyotyping, chromosomal microarray; ³Groups of trios or more. *DMD*, Duchenne muscular dystrophy; *FSHD*, facioscapulohumeral muscular dystrophy; LRS, long-read sequencing; MT, myotonic dystrophy; NIID, neuronal intranuclear inclusion disease; NMD, neuromuscular disorder; ONT, Oxford Nanopore Technology; OPDM, oculopharyngodistal myopathy; PacBio, Pacific Biosciences; SMA, spinal muscular atrophy; SMRT, single-molecule real-time; SRS, short-read sequencing; SR-WES, short-read whole exome sequencing; SR-WGS, short-read whole genome sequencing; SV, structural variant; UTR, untranslated region.

1.4.3. Approaches to variant filtration and interpretation

A central challenge in the application of comprehensive genome diagnostics is the vast amount of genomic information generated; for example, WGS typically yields approximately five million variants per genome (Auton et al., 2015; Berg et al., 2011). Consequently, bioinformatic filtering is essential to reduce the initial set of variants to a small number of candidate variants suitable for manual interpretation and pathogenicity classification. One fundamental and highly standardised step in variant prioritisation is population frequency filtering of a given variant. This step is based on the expectation that highly penetrant pathogenic variants will be rare or absent in large and/or healthy populations. As such, variants are commonly filtered according to allele frequency in large population reference datasets, such as the Genome Aggregation Database (gnomAD), which provide variant frequency information in large cohorts that are not enriched for individuals with severe pediatric genetic disorders (Gudmundsson et al., 2022). Only variants with allele frequencies below defined thresholds are retained for further evaluation, with thresholds ranging from 0% for some AD diseases to 1% for AR diseases (Richards et al., 2015; Whiffin et al., 2017; Wright et al., 2018). Finally, assessment of variant information in correlation to the affected individual's phenotype and expected inheritance pattern can be used to further refine genomic variants identified in large-scale sequencing (Smedley et al., 2021).

Another vital step in variant filtering is the assessment of predicted functional consequence. In its simplest form, this step provides an initial stratification of variants based on variant class and likely molecular impact (Wright et al., 2018). Nonsense and frameshift variants are generally prioritised due to their high likelihood of disrupting gene function (Richards et al., 2015). In contrast, the functional impact of other variant types (missense, non-coding and synonymous) requires additional supporting evidence. As such, a wide range of *in silico* tools for predicting

variant deleteriousness are applied to estimate the likelihood that a variant is deleterious (Kircher et al., 2014). These methods have historically functioned by integrating diverse sources of biological information, including evolutionary conservation, biochemical substitution properties, sequence context and protein structural features (Adzhubei et al., 2010, 2013; Sim et al., 2012). In more recent approaches, machine-learning frameworks trained on known pathogenic and benign variants have been used to generate predicted deleteriousness scores (Ioannidis et al., 2016; Rentzsch et al., 2019). In addition, deep learning-based approaches, such as SpliceAI, have further improved the prediction of splice-altering variants (Jaganathan et al., 2019). Collectively, the scores generated by *in silico* variant-predicted deleteriousness methods are vital tools for filtering variants in large genomic datasets (Wright et al., 2018). Thus, *in silico* tools continue to advance and have substantial power to reduce the interpretive burden associated with large-scale genomic sequencing. While bioinformatic tools are vital in supporting pathogenic variant identification and can provide supporting pathogenicity evidence for classification, additional information is still required to reach a pathogenic classification (Richards et al., 2015).

1.4.4. Long-read RNA-sequencing as a complementary tool for genetic diagnosis

While LR-WGS improves capture of genomic sequencing, genetic diagnostics remain limited by challenges in classifying variant pathogenicity, notably for elucidating the functional impact of rare missense, non-coding or predicted splicing variants (Lewis et al., 2025; Yépez et al., 2026). As such, RNA-sequencing (RNA-seq) presents an important tool for a complementary approach, with the potential to provide variant-level functional insight and improve diagnostic rates of genomic sequencing (Lewis et al., 2025). RNA-seq is a high-throughput technique that quantifies and characterises the complete set of RNA molecules expressed in a biological sample at a specific point in time (Kukurba & Montgomery, 2015). In the context of bulk RNA-seq, RNA

is extracted from a heterogeneous tissue sample containing a mixture of cell types; the resulting data represents an averaged gene expression and splicing signal across all cells within that tissue (Hrdlickova et al., 2017). RNA is first converted into complementary DNA (cDNA), which is then sequenced using MPS (Hrdlickova et al., 2017). Subsequently, the reads are aligned to a reference transcriptome, thus allowing reconstruction of transcript isoforms and quantification of gene expression levels in transcripts per million (TPM) (Conesa et al., 2016; Wagner et al., 2012).

Altogether, by profiling the complete set of RNA transcripts in a given tissue, RNA-seq assists in elucidating transcriptional changes caused by gene variations (Ku et al., 2012; Xuan et al., 2013). Short-read RNA-seq (SR-RNA-seq) is a well-established technique for identifying gene expression and/or splicing outliers by comparing transcriptomic data to comprehensive reference cohorts, notably the widely used Genotype-Tissue Expression project (GTEx Consortium, 2020; Marwaha et al., 2022). By identifying a gene for aberrant expression or splicing in an individual, the gene can be prioritized for identification of pathogenic variant in genomic analyses (Kremer et al., 2017; Marwaha et al., 2022). Additionally, RNA-seq can support interpretation of known genomic variants by visualizing impact on transcript expression and splicing (Marwaha et al., 2022). Summary of selected relevant RNA-seq studies for genetic diagnosis of NMDs is found in **Table 1.2**. A recent study implemented parallel whole blood SR-WGS and SR-RNA-seq testing for 33 NMD individuals with negative WES results (Yang et al., 2023). An additional diagnostic rate of 42% (15/33) was yielded, with SR-WGS and SR-RNA-seq contributing to 36% and 6% of the diagnoses, respectively (Yang et al., 2023). Truncating *RBCK1* variants detected via SR-WGS were identified in two unrelated families with children with NMDs; using SR-RNA-seq, aberrant splicing and truncated protein products were confirmed (Wang et al., 2013). Lastly, SR-RNA-seq

has allowed characterization of several deep intronic *DMD* variants in unrelated pediatric patients (Gonorazky et al., 2019; Zhao et al., 2024).

However, SR-RNA-seq has technical limitations, which may be overcome by long-read RNA-seq (LR-RNA-seq). Notably, SR-RNA-seq requires computational reconstruction of transcripts, which is made challenging by the difficulty of the short read length to accurately resolve transcripts in the context of complex genes or genes with many isoforms (Marwaha et al., 2022). This limitation is especially relevant in the context of NMDs, where many disease-associated genes are large, structurally complex and enriched for alternative splicing events, making accurate isoform resolution difficult using SR-RNA-seq (Gonorazky et al., 2019; W. F. Lim & Rinaldi, 2023; Zatz et al., 2016). Conversely, LR-RNA-seq uses longer reads to span the entire transcript, thereby resolving entire isoforms without computational reassembly (Sedlazeck et al., 2018). Accordingly, LR-RNA-seq may provide a valuable complement to genomic sequencing by enabling direct assessment of transcript-level consequences of genetic variation. Despite these advantages over SR-RNA-seq, the use of complementary LR-RNA-seq has not been studied yet in the genetic diagnostics of NMDs to date.

Table 1.2. RNA-seq methods demonstrate contributions to clinical yield in select published studies of individuals and cohorts with NMDs. Ordered by year of publication.

Study	Conditions investigated	Cohort size	Technology and source	RNA-seq contributions and benefits
Cummings et al. (2017)	Primary muscle disorders	63 individuals ¹	SR-RNA-seq (skeletal muscle)	• Aberrant splice events led to diagnosis of 21% of patients with no strong candidates from SR-WGS or SR-WES • Enabled detection of aberrant splicing and gene expression signatures consistent with molecular diagnoses
Gonorazky et al. (2019)	NMDs, with negative previous SR-WES or SRS-based gene panels	70 individuals from 29 families (4 families were positive controls with known genetic diagnosis)	SR-RNA-seq (skeletal muscle, primary skin fibroblasts or myotubes created by transdifferentiating fibroblasts)	• 9/25 (36%) diagnostic rate • Identified variants in regulatory upstream regions, deep intronic variants and variants that impact allele-specific expression
Hong et al. (2022)	NMDs	117 individuals	SR-RNA-seq (skeletal muscle), compared to SR-WES	• 38% from SR-RNA-seq analysis alone, compared with 34.9% from SR-WES alone
<u>Yang et al. (2023)</u>	NMDs with negative SR-WES	33 individuals	SR-WGS; SR-RNA-seq (whole blood)	• 2/33 (6%) diagnostic rate • Integration of alternative splicing result into WGS filtering supported filtering of rare intronic variants
Marchant et al. (2024)	NMDs	47 individuals	SR-WGS, SR-RNA-seq (skeletal muscle).	• 45% of diagnoses required genome sequencing, RNA and/or protein studies for variant interpretation • Improved classification of VUS through functional transcriptomic evidence

Segarra-Casas et al. (2024)	Muscular dystrophy	1 individual (female)	SR-WGS; SR-RNA-seq (skeletal muscle)	• 1/1 (100%) diagnostic rate: de novo balanced translocation • RNA-seq rapidly identified <i>DMD</i> as expression outlier, after <i>DMD</i> analysis negative following MLPA, SRS and trio SR-WES
------------------------------------	--------------------	-----------------------	--------------------------------------	---

MLPA, multi-plex ligation-dependent probe amplification; NMD, neuromuscular disorder; SR-RNA-seq, short-read RNA sequencing; SRS, short-read sequencing; SR-WES, short-read whole exome sequencing; SR-WGS, short-read whole genome sequencing; VUS, variant of uncertain significance.

Additionally, the diagnostic utility of RNA-seq is highly dependent on the tissue analysed, as gene expression and transcript isoform usage vary substantially between tissues (Cummings et al., 2017; GTEx Consortium, 2020). Consequently, the transcriptional effects of a variant are most informatively assessed in the tissue most relevant to that disease (Cummings et al., 2017; Hekselman & Yeger-Lotem, 2020). This consideration is particularly important in NMDs; many NMD genes display enriched expression and alternative splicing profiles in disease-relevant tissues (i.e. skeletal muscle and heart), which is poorly assessed using other tissues (e.g., blood and skin-derived fibroblasts (Gonorazky et al., 2019; Martin et al., 2026). Indeed, among the pathogenic variants identified through transcriptomic analysis by Cummings et al. (2017), the affected exons exhibited low coverage in GTEx whole-blood and fibroblast samples. This finding suggests that many of the diagnoses identified through transcriptomic analysis would not have been achieved with RNA-seq from tissues other than disease-relevant skeletal muscle (Cummings et al., 2017). As such, the availability of skeletal muscle biopsies from individuals undergoing neuromuscular investigation presents a unique opportunity to evaluate whether disease-relevant transcriptomic data can improve genetic diagnosis and variant interpretation in NMDs (Hong et al., 2022).

1.5. Thesis rationale

NMDs pose a significant healthcare burden and are clinically relevant to the Manitoban and Canadian healthcare systems. Many affected individuals do not receive a molecular diagnosis through current sequencing approaches, nor the related psychological or clinical benefits. The combined use of LR-WGS and LR-RNA sequencing presents a promising avenue for improving diagnostic outcomes in NMDs. This paired approach can improve resolution of complex structural variants, repeat expansions and deep intronic variants, which are often missed in the short-read WES method used in clinic currently. While a multi-omics approach can improve diagnostic rates for individuals with NMDs (Gonorazky et al., 2019), LR-RNA-seq has yet to be pursued using LR-WGS for NMD diagnostics. Further research is needed to understand this approach and to develop tools, such as filtering pipelines, that enable clinical implementation.

We hypothesize that a multi-omics approach, including paired long-read WGS sequencing and transcriptomic data from a disease-relevant tissue, will increase diagnostic yield in individuals with genetically undiagnosed NMDs. By improving diagnostic resolution, this approach has the potential to end the diagnostic odyssey for affected individuals and families. Based on recent literature using SRS-based expression profiles, we estimate that 36% of study participants will receive a diagnosis (Gonorazky et al., 2019).

Chapter 2: Methods

2.1. Ethical approval and funding

This study was approved by the University of Manitoba's Bannatyne Campus Health Research Ethics Board (REB approval number: HS19854/H2016:247). The student principal investigator (PI) was added to this ethics approval. This study is funded by a Research Manitoba New Investigator Operating Grant (grant number: 5381, 2022-2025) awarded to the supervisor, Dr. Galen Wright. This work is also supported by a Canada Graduate Scholarship – Master's, awarded by the Canadian Institutes of Health Research from May 2025 to May 2026.

2.2. Participant recruitment

Participants ($n = 22$) with suspected genetic NMDs were recruited through the Shared Health Genetics and Metabolism Program in Winnipeg, Manitoba, from 2017 to 2018. Inclusion and exclusion criteria are listed in **Table 2.1**. All study participants (or their legal guardians) met with the study genetic counsellor to obtain information about the study and provide informed consent, including for the release of relevant clinical specimens, as well as for genomic and transcriptomic investigations. Participants consented to be re-contacted to receive study results or to obtain further information as required by the study team. Participants were informed of the potential for incidental findings, including clinically relevant genetic variants unrelated to the individual's NMD presentation. Participants chose whether to receive incidental findings as part of the consent process. Unaffected family members were not recruited in this study; however, affected family members were eligible for recruitment.

Table 2.1. Inclusion and exclusion criteria for recruitment into NMD long-read multi-omic 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	
NMD, neuromuscular disorder.	

For each participant, genomic DNA (gDNA) was extracted from whole blood samples. Fresh-frozen disease-relevant tissue from skeletal muscle biopsies was also procured from 19/22 individuals; these were released from the clinical laboratory at Shared Health (RNA extractions described in **Section 2.8**).

2.3. Retrospective chart review and phenotyping

Each participant's medical record in Genetics & Metabolism was reviewed for relevant clinical information. This program within Shared Health provides public healthcare through assessment, diagnosis and management for individuals with genetic indications, including NMDs

(*Genetics and Metabolism*, n.d.). To promote consistency and help reduce error in data collection, a data capture form was used to collect information in a similar manner across the participants (**Appendix A**) (Vassar & Holzmann, 2013). This study used a data collection form developed in a pilot feasibility study within this cohort, in collaboration with a clinical geneticist, P.F., genetic counsellor, J.N.H., and genetics researchers, B.D. and G.W. This pilot study, conducted by an Honours Thesis student, A.M., provided initial phenotyping, bioinformatic pipeline development and variant filtration of LR-WGS data for five of the 22 participants (NMIDs 03, 06, 07, 09 and 12) (Martin, 2025). The adapted data collection form included 81 fields selected to comprehensively capture information relevant to NMDs and genetic conditions. This information included: family structure, family history, ethnicity, age at first and last presentation in genetics, NMD phenotype, other medical history, and laboratory, imaging, and genetic investigations.

Moreover, chart review data were used to generate a list of standardized phenotypic terms using the Human Phenotype Ontology (HPO) system (Köhler et al., 2019; Robinson et al., 2008). HPO is a standardized system used to describe phenotypic abnormalities in human disease (Gargano et al., 2024). Each HPO term has a unique identifying code that is linked to associated genes based on disease descriptions in the Online Mendelian Inheritance in Man database (OMIM; <https://omim.org/>) (Amberger et al., 2015). By assigning HPO terms to participants, variants in genes associated with the phenotype could be prioritized in LR-WGS filtering. The HPO database uses a hierarchical system, wherein the most specific HPO terms are categorized under increasingly broad HPO terms. For example, the term “*Proximal muscle weakness*” (HP:0003701) is categorized directly underneath “*Muscle weakness*” (HP:0001324), while the broadest phenotypic abnormality term it falls under is “*Abnormality of the musculoskeletal system*” (HP:0033127). For each participant, the most specific HPO terms were assigned for phenotype

description with the research team, while the broadest phenotypic abnormality terms were selected for use in variant filtration. Further, only terms most likely related to an underlying NMD were selected for filtering. HPO terms less likely to be related to an underlying NMD, or which could be explained by another health condition, were excluded. All assigned HPO terms were reviewed and approved by the clinical geneticist, P.F., prior to use in variant filtration.

2.4. LR-WGS, variant annotation and data transfer

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 (Agilent Technologies, 2022). Samples were verified to have a concentration between 20-200 ng/μl with a total concentration of 5-7 ng/μl of DNA, peak sizes >60 kb with DNA integrity numbers ≥ 9 , and an A260/A280 of 1.8 or greater prior to downstream analysis (Pellecchia, 2024a, 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, 2024b). Sequencing data for each sample was aligned to the most recent human reference genome (GRCh38/hg38) using the pbmm2 aligner specifically curated for PacBio sequencing, which interfaces with the miimap2 alignment engine (Li, 2017/2026; PacBio, 2017/2026).

In addition to library preparation and sequencing, variant calling and annotation were performed at the TCAG facility using a customized PacBio workflow (Pellecchia, 2024a, 2024b). Small variants (e.g., SNVs) were called using the machine-learning-based tool DeepVariant, developed by Google (Pellecchia, 2024aa). Subsequently, small variants were annotated for 171

features using a custom small-variant bioinformatic annotation pipeline (rev28.2) developed at TCAG using the ANNOVAR tool (Pellecchia, 2024a; Wang et al., 2010; Yuen et al., 2015). Small variants were annotated with predicted deleteriousness scores from several *in silico* prediction tools (i.e., CADD, MPC, MutationTaster, SpliceAI, etc.) (Pellecchia, 2024aa). CNVs were called by HiFiCNV, and SVs were called by pbsv; these two variant types were then annotated using a custom bioinformatic annotation pipeline developed at TCAG (Pellecchia, 2024ba). CNVs were annotated for 99 features, while SVs were annotated for 110 features. Annotated information for SNVs, CNVs and SVs includes gene and protein information, disease and phenotype associations (e.g., HPO terms, OMIM, ClinVar), and population allele frequencies obtained from data repositories and independent cohorts, including both SRS- and LRS-derived data (Pellecchia, 2024a, 2024b). These population allele frequencies were obtained from both SRS- and LRS-derived data and included ancestral make-up as described in **Sections 2.5.2 and 2.5.3**. Annotated variants for each participant were returned as three variant call format files (VCFs), one for each variant type of SNV, CNV and SV. Annotation terms and descriptions for SNV, CNV and SV VCFs are described in **Appendix B**.

Annotated files and raw sequencing data for small variants and CNVs/SVs were transferred via a secure portal established by TCAG (<https://data-portal.tcag.ca/>) for input into a local secure computing infrastructure at the University of Manitoba (Pellecchia, 2024a, 2024b). The secure computational infrastructure includes a Dell PowerEdge R740 Server housed in the John Buhler Research Centre, configured to meet privacy and security standards for handling data from human research participants. Access to the server and participant-annotated files, including raw sequencing data, is restricted exclusively to the student PI and authorized members of the research team.

2.5. LR-WGS bioinformatics filtering pipeline

As LR-WGS can generate over five million variants per genome; a bioinformatics pipeline is required to filter raw LR-WGS data into a focused set of candidate variants for downstream analysis (**Figure 2.1**). The pipeline was adapted from the one developed by A.M. and Dr. G.W. in the pilot feasibility study (**Section 2.3**), which focused on optimizing the number of candidate variants that could be reasonably assessed through manual interpretation (A. B. Martin, 2025). Ultimately, the bioinformatic pipeline was consistent with those used in the field, including the use of allele frequencies, computational, predictive data, and quality control parameters (Bruels et al., 2022; Pedersen et al., 2021; Pejaver et al., 2022; Steyaert et al., 2025). Filtering encompassed small variants, CNVs and SVs, while analysis of STRs is described in **Section 2.6**. Bioinformatic filtering and analyses were performed in R v.4.3.3 using the tidyverse v2.0.0 package for data handling and visualization (R Core Team, 2024; Wickham et al., 2019). A set of “hard” filters was applied uniformly across all samples to retain only rare variants with high predicted deleteriousness, followed by optional “soft” filters tailored to individual clinical features. Altogether, this filtering strategy aimed to prioritize the most relevant candidate variants to the individual’s phenotype, while limiting incidental findings.

2.5.1. *Hard filters for small variants*

The first hard filter for small variants retained only rare variants based on minor allele frequency $\leq 0.1\%$ (Stenton et al., 2024). Population allele frequencies were obtained from the gnomAD v4.0 short-read genome dataset using the maximum frequency in any subpopulation (Karczewski et al., 2020). gnomAD v4.0 is a diverse dataset of >807,000 genomic samples (730,947 from SR-WES and 76,215 from WGS), including individuals of European (77.07%), South Asian (5.64%), African (4.65%), East Asian (2.78%) and other ancestral backgrounds.

Secondly, population allele frequencies were derived from LR-WGS data from a small, unrelated cohort of individuals ($n \sim 500$) of 75% European ancestry (Cohen et al., 2022; Pellicchia, 2024a). This latter cohort was from the Genomic Answers for Kids (GA4K) initiative at the Children’s Mercy Hospital Research Institute and serves as a technology-matched allele frequency dataset (Cohen et al., 2022). Small variants were first filtered based on their frequency in the larger gnomAD dataset, then in the smaller GA4K dataset; variants with minor allele frequency $\leq 0.1\%$ in both datasets were retained.

The second hard filter for small variants retained only variants with high predicted deleteriousness. Small variant VCFs were annotated by SickKids TCAG with variant effect impact classifications (e.g., “*Frameshift*”, “*Missense-High*”, “*Missense-Low*”). These classifications were subsequently used in the filtering pipeline. Small variants were retained if variant effect impact consisted of frameshift, stop-gain, splice site or loss-of-function classifications. High deleteriousness was defined for missense, predicted splicing, and non-coding variants using *in silico* tools (**Table 2.2**). The description of the *in silico* prediction tools used in the current investigation is provided below.

Firstly, missense variants were filtered based on rare exome variant ensemble (REVEL) scores ≥ 0.45 (Ioannidis et al., 2016). This tool aggregates results of multiple individual pathogenicity predictors for missense variants (MutPred, FATHMM, VEST, PolyPhen, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP, SiPhy, phyloP, and phastCons), and was trained using both pathogenic and rare neutral missense variants. As such, REVEL demonstrates improved variant prediction performance compared to independent test sets and other ensemble methods, and helps to distinguish pathogenic from rare neutral variants (Ioannidis et al., 2016).

Secondly, small splice variants were retained if located within 100 bp of a splice site junction or if they had high predicted impact based on dbNSFP v3.0 and SpliceAI scores (**Table 2.2**). dbNSFP derives impact scores by combining outputs from multiple *in silico* prediction tools that assess alterations to donor and acceptor splice sites and overall splicing patterns (Liu et al., 2016). By contrast, SpliceAI is a deep learning algorithm designed to predict variant impact on pre-mRNA splicing, including both the nearby splice junction (destruction or creation), and global splicing outcomes (e.g., exon skipping) (de Sainte Agathe et al., 2023; Jaganathan et al., 2019).

Lastly, non-coding variants were retained based on nucleotide-level conservation inferred from either placental mammals or the 100 vertebrate group (Pollard et al., 2010). If conservation information was not available for that nucleotide, the variant was retained based on an elevated combined annotation-dependent deletion (CADD) PHRED score (Rentzsch et al., 2019). CADD scores are aggregate scores of 60 genomic feature annotations (including evolutionary constraint, functional predictions, and surrounding sequence context) derived from machine learning models which used a large training set of variants fixed in human populations (compared to chimpanzees) and simulated *de novo* variants (both neutral and deleterious) (Rentzsch et al., 2019). CADD PHRED scores, as used in the filtering pipeline, are $\log_{10}$ -derived rank scores based on the genome-wide distribution of scores for all ~9 billion potential SNVs. For example, a score of 20 corresponds approximately to the top 1% of most deleterious variants in the genome, while scores of 30 and 40 correspond to the top 0.1% and 0.01%.

Table 2.2. Small variant predicted deleteriousness classifications and filtering thresholds, with variants predicted to have high functional impact (bolded) prioritized during LR-WGS filtering.

Variant type	Predicted functional impact	Threshold(s)
Missense (non-synonymous single nucleotide variants)	High	REVEL_score ≥ 0.45
	Medium	REVEL_score ≥ 0.25 <u>OR</u> phylopMam_avg ≥ 1.30 <u>OR</u> phylopVert100_avg ≥ 3.90 <u>OR</u> MPC_score ≥ 2.00 <u>OR</u> 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 <u>OR</u> phylopVert100_avg ≥ 3.90 <u>OR</u> CADD_phred ≥ 21.1) <u>AND</u> (dbSNP_common != "NA")) <u>OR</u> ((phylopMam_avg ≥ 1.10 <u>OR</u> phylopVert100_avg ≥ 1.60 <u>OR</u> CADD_phred ≥ 13.7) <u>AND</u> (dbSNP_region != "NA"))
	Low	'Other' variants that do not meet the above criteria.
Predicted splicing	High	dbSNV_ADA_SCORE > 0.6 <u>OR</u> dbSNV_RF_SCORE > 0.6 <u>OR</u> (spliceAI_DS_AG > 0.5 <u>AND</u> spliceAI_DP_AG ≤ 50) <u>OR</u> (spliceAI_DS_AL > 0.5 <u>AND</u> spliceAI_DP_AL ≤ 50) <u>OR</u> (spliceAI_DS_DG > 0.5 <u>AND</u> spliceAI_DP_DG ≤ 50) <u>OR</u> (spliceAI_DS_DL > 0.5 <u>AND</u> spliceAI_DP_DL ≤ 50)

	Medium	dbscSNV_ADA_SCORE > 0.6 <u>OR</u> dbscSNV_RF_SCORE > 0.6 <u>OR</u> (spliceAI_DS_AG > 0.2 <u>AND</u> spliceAI_DP_AG ≤ 50) <u>OR</u> (spliceAI_DS_AL > 0.2 <u>AND</u> spliceAI_DP_AL ≤ 50) <u>OR</u> (spliceAI_DS_DG > 0.2 <u>AND</u> spliceAI_DP_DG ≤ 50) <u>OR</u> (spliceAI_DS_DL > 0.2 <u>AND</u> spliceAI_DP_DL ≤ 50)
	Low	Predicted splicing variants that do not meet the above two criteria.
Non-coding RNA	High	((phylopMam_avg ≥ 1.30 <u>OR</u> phylopVert100_avg ≥ 3.90 <u>OR</u> CADD_phred ≥ 21.1) <u>AND</u> (phastCons_placental != "NA"))
	Medium	((phylopMam_avg ≥ 1.10 <u>OR</u> phylopVert100_avg ≥ 1.60 <u>OR</u> CADD_phred ≥ 13.7) <u>AND</u> (phastCons_placental != "NA"))
	Low	Non-coding RNA variants that do not meet the above two criteria.

Deleteriousness prediction thresholds identified based on The Centre for Applied Genomics annotation and filtering guidelines (Pellecchia, 2024ab). CADD_phred = PHRED-like c-score ranking from CADD (Combined Annotation Dependent Depletion); dbscSNV_ADA_SCORE/dbscSNV_RF_SCORE = splice site prediction scores from dbscSNV; dbsnp_common = exact match (position, allele) to common dbSNP track USCS; dbsnp_region = overlap-based match for dbSNP; LR-WGS = long-read whole genome sequencing; MPC_score = MPC (Missense badness, PolyPhen-2, and Constraint) score based on the Exome Aggregation Consortium (ExAC) database; 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).

2.5.2. Hard filters for SVs and CNVs

SVs and CNVs followed a similar filtration pipeline, firstly to retain variants with low population allele frequencies. SVs/CNVs 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 with a population allele frequency <1% based on internal cohorts available to SickKids TCAG (Pellecchia, 2024b). One such internal cohort was a database of unrelated samples sequenced using the Illumina NovaSeq 6000 SRS system at an average coverage of 35X, called by Dragen 3.8.4, with at least 90% reciprocal overlap (Pellecchia, 2024b). Further allele frequencies for CNVs and SVs were obtained from a heterogeneous dataset from NCBI known as NTSD186. Lastly, allele frequencies were derived from a LRS analysis of SVs called from over 500 unrelated samples sequenced by the GA4K at Children’s Mercy Hospital, with >90% overlap for SVs and >50% reciprocal overlap for CNVs (Pellecchia, 2024b). SVs/CNVs had to have a population allele frequency <1% across each of these internal cohorts to be retained. A less stringent population frequency was used to filter SVs and CNVs (<1%) compared to small variants (<0.1%) due to the limited size of available LRS-matched reference datasets for structural variation compared to the larger reference datasets available for small variant filtering.

Secondly, SVs and CNVs underwent filtering to retain variants with **high predicted deleteriousness**. Given that most genome sequencing has used short reads and lacked analysis of SVs, there is a paucity of tools to predict deleteriousness of such variants as exists for small variants. We applied the following criterion to limit the number of SVs/CNVs returned; SVs/CNVs were retained if at least one gene was encompassed, and SVs were retained if they impacted coding (exonic) regions.

2.5.3. Phenotypic filters for short variants, SVs and CNVs

For both short variants and SVs/CNVs, additional clinical filters were applied after the hard filters were completed, based on chart review data. Firstly, variants were retained if they matched at least one of the selected HPO terms for the participant. This filter prioritized variants in genes associated with the affected individual's observed phenotype, thereby limiting incidental findings (Schön et al., 2021). Secondly, variants were retained if located in a gene known to be implicated in an NMD, as obtained from the most recent version of the GeneTable of Neuromuscular Disorders at the time filtration began ($n = 686$) (Benarroch et al., 2024; downloaded November 2024). The GeneTable compiles genes with established evidence for association with monogenic NMDs due to a primary defect in the nuclear genome; the list is curated through expert clinical interpretation of ongoing findings from the scientific community. By prioritizing variants in genes with established NMD associations, this filter limited findings related to participant health concerns beyond NMDs.

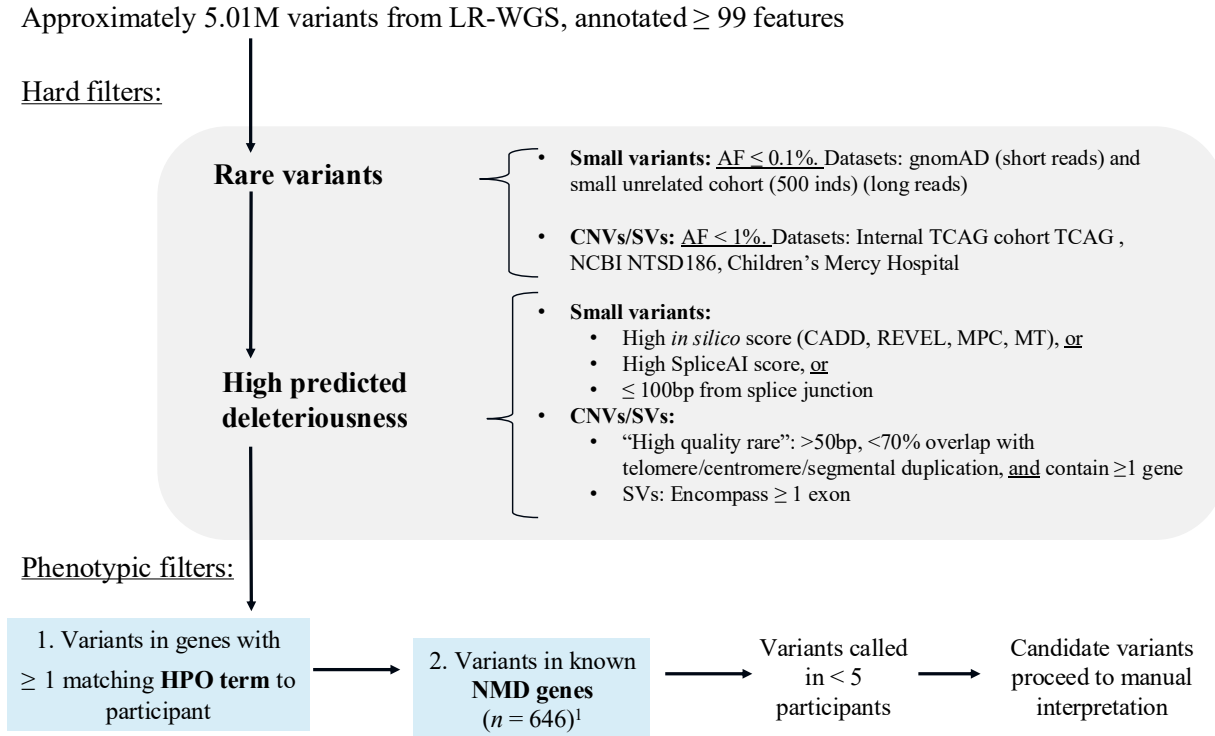


Figure 2.1. NMD-focused variant filtration and prioritization pipeline applied to LR-WGS data from a cohort of individuals with NMDs ($n = 22$), with incorporated allele frequency, predicted deleteriousness and phenotypic (HPO) terms. The grey box indicates “hard” filters with defined thresholds in allele frequency and measures of predicted deleteriousness applied equally to all variants in all individuals. The blue boxes indicate phenotype-specific filters, the first of which uses a curated list of HPO terms for each participant to support filtering. The second phenotype-specific filter retains variants located in known NMD-associated genes. ¹NMD-associated genes were derived from the GeneTable of Neuromuscular Disorders downloaded November 2024 (Benarroch et al., 2024). CADD, combined annotation dependent depletion; CNV, copy number variant; gnomAD, genome aggregation database; HPO, Human Phenotype Ontology; inds = individuals; LR-WGS, long-read whole genome sequencing; MAF, minor allele frequency; MPC, Missense Badness, Polyphen-2 and Constraint; MT, MutationTaster; NMD, neuromuscular disorder; PAF, population allele frequency; REVEL, Rare Exome Variant Ensemble Learner.

A for loop, which automates repeated operations across multiple files, was used to process and filter each participant’s variant files concurrently, then combine the results into a single master dataset containing all variant types across individuals. At this stage, variants which appeared five

or more times across the cohort could be excluded. This filtering step allowed the removal of genetic variants likely to result from sequencing artifacts or insufficient annotation (e.g., missing accurate population allele frequencies). Finally, slivar software was applied to small variants to enable rapid annotation of variants where additional variants were present in the same gene, even if they did not meet the pathogenicity or allele frequency thresholds required for retention in our primary filtering pipeline (Pedersen et al., 2021b). This step facilitated assessment of potential biallelic variation in genes with a single likely pathogenic variant in the context of recessive disorders. Following this final step, a shortlist of variants was available for manual interpretation; the variants in this list were identified as “*candidate variants*”.

2.6. STR genotyping of LR-WGS data

LR-WGS data were genotyped to assess the lengths and sequence composition of known NMD-associated STRs. To achieve this, we used the Tandem Repeat Genotyping Tool (TRGT), a PacBio-developed software package designed to genotype structurally complex tandem repeats from LRS data (Dolzhenko et al., 2024). We input aligned HiFi (i.e., PacBio LRS) reads and a list of repeat loci and definitions to assess; subsequently, TRGT identifies reads spanning the given repeat loci and reconstructs the full allele sequence, allowing accurate estimation of both repeat length and sequence composition, including repeat motifs and sequence interruptions (**Fig. 2.2**). Although pathogenic STR expansions have been implicated in more than 60 monogenic disorders, only loci relevant to NMDs were investigated in this study (Chen et al., 2025). To curate these loci, we used the STRchive catalogue (v2.1.0; released 2024-12-12), a comprehensive resource that consolidates information on disease-associated tandem repeat loci through automated literature curation and integration of genomic database information (Hiatt et al., 2025). The STRchive catalogue was subset to loci associated with NMD-related genes, generating a list of repeat loci,

motifs and repeat structures for genotyping in our cohort using TRGT (**Appendix C**) (Hiatt et al., 2023/2025).

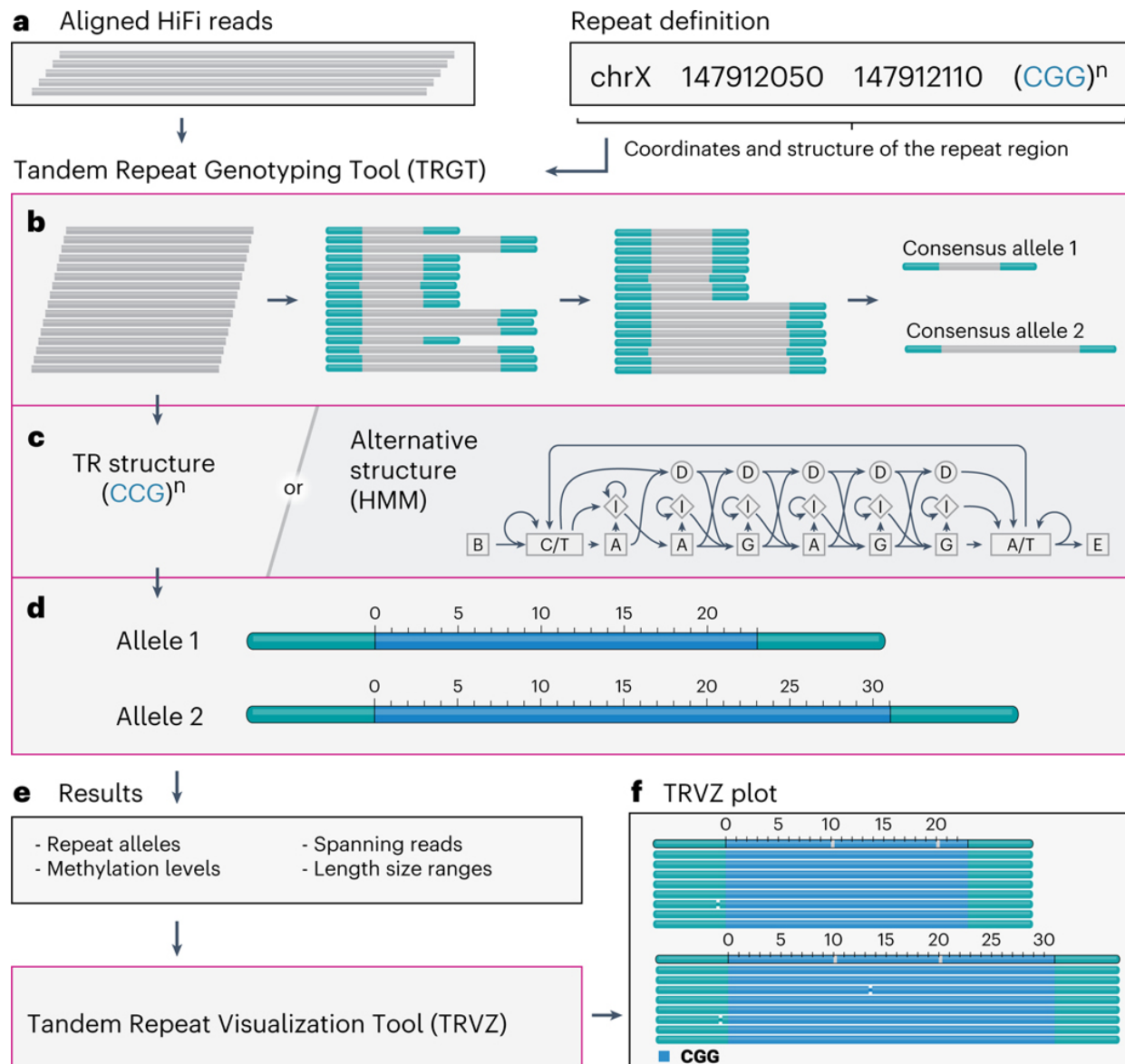


Figure 2.2. Visual overview of TRGT and TRVZ tools for genotyping and visualizing tandem repeat expansions in LRS data. **a**, Input to TRGT consists of HiFi reads and a list of repeat definitions. **b**, TRGT determines consensus repeat alleles. **c**, TRGT uses the pre-specified structure of the TR region to locate individual motif copies in each repeat allele. **d**, More complex repeat regions are specified with HMMs. **e**, Overview of key fields in TRGT’s output. **f**, TRVZ generates plots that display repeat alleles and reads aligning to them, with optional methylation.” Reproduced from

Dolzhenko et al., 2025, Figure 1, with permission from Springer Nature, license number 6292760379518. HMM, Hidden Markov Models; LRS, long-read sequencing. TRGT, Tandem Repeat Genotyping Tool; TRVZ, Tandem Repeat Visualization.

2.7. Manual genetic variant prioritization and validation

Each variant remaining in the filtered shortlist underwent detailed analysis via a standardized checklist to evaluate its potential for pathogenicity (**Figure 2.3**). Firstly, inheritance patterns and zygosity status were assessed for consistency with the suspected mode of inheritance (i.e., confirming the presence of two rare deleterious alleles for an autosomal recessive pattern of inheritance). Secondly, OMIM phenotypes for genes linked to variants were assessed for correlation with the individual's NMD phenotype. Variants which did not correspond with the individual's phenotype, but which did demonstrate elevated pathogenicity, were presented to the rare disease team for consultation regarding the investigation of potential incidental findings. Variants plausible in both inheritance pattern and phenotype were assessed for pathogenicity according to ACMG classification guidelines (**Fig. 2.4**) (Richards et al., 2015). gnomAD and ClinVar databases were also consulted for evidence of pathogenicity (Chen et al., 2024; Landrum et al., 2014). Supplementary tools, such as Franklin by Genoox and the UCSC Genome Browser, were used for additional variant visualization and information as required (*Franklin*, n.d.; Perez et al., 2025). All variants of interest were visualized using the Integrative Genomics Viewer (IGV) for manual review, quality control and confirmation (Robinson et al., 2017).

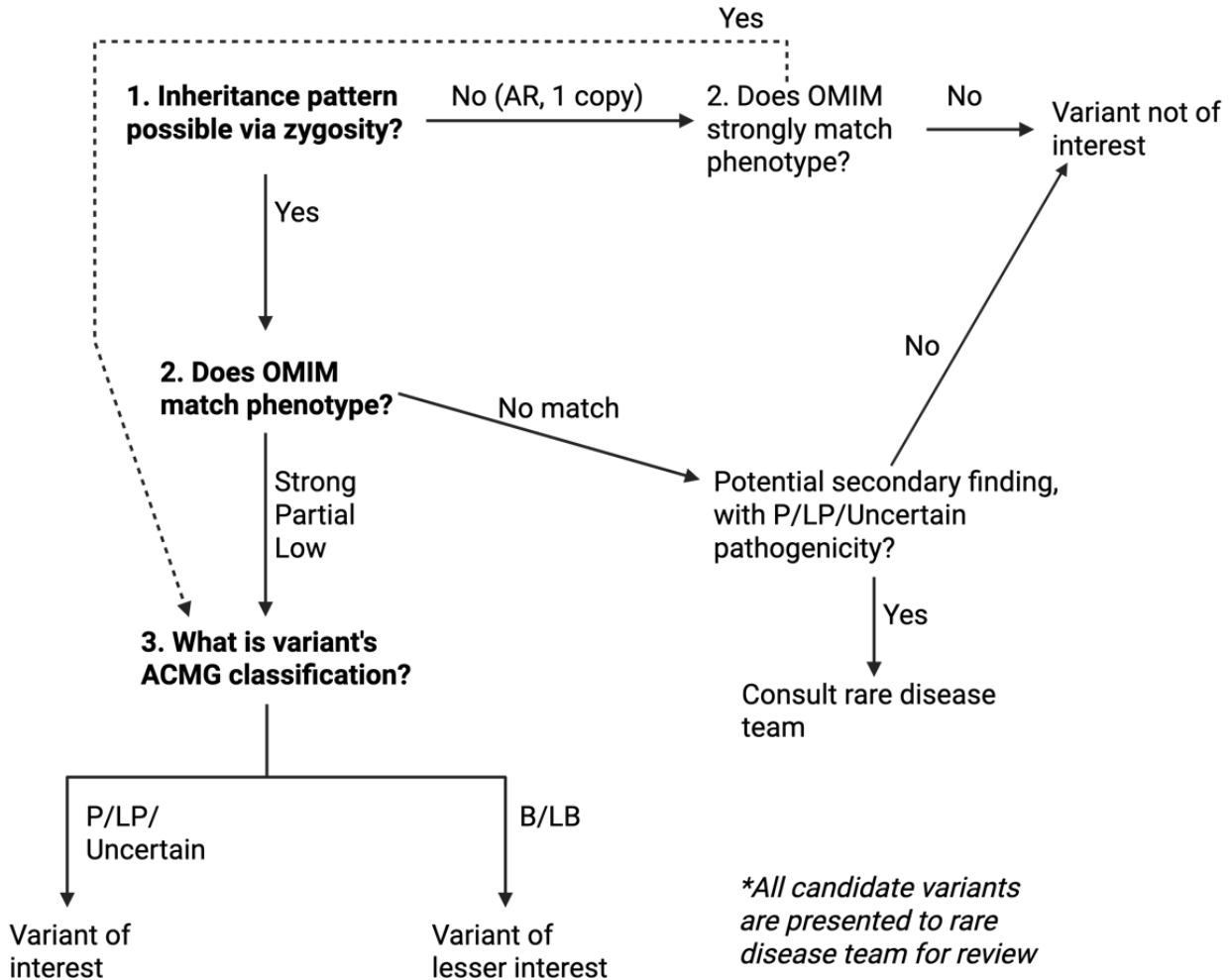


Figure 2.3. Framework for standardized manual variant interpretation of all candidate variants isolated by bioinformatic filtering process (see Figure 1). AR, autosomal recessive; B, benign; LB, likely benign; LP, likely pathogenic; P, pathogenic.

Following manual interpretation of variants, the student PI presented the analyzed variants to the research team. The group comprised a clinician-scientist (P.F., MD, FRCPC, FCCMG), a genetic counsellor (J.N.H., MS, CGC), and human genetics researchers with expertise in neurogenetics and large-scale computational analysis of rare pathogenic variants (G.W., PhD and B.D., PhD). The strength of evidence for each variant was discussed based on mode of inheritance, clinical relevance, pathogenicity and predicted deleteriousness, considering factors such as the

phenotype and family history of the affected individual. The team decided on plans for follow-up with affected individuals, as necessary, to confirm results (e.g., Sanger sequencing, model organism validation), or to request samples for familial studies as required and available. Genomic findings from this study remain of research grade and must be confirmed in clinically validated laboratory settings. Finally, findings from the manual variant analysis were compared to the individual's prior genetic testing results, allowing to assess LR-WGS performance in comparison to standard clinical approaches

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population Data	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>			Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
Computational And Predictive Data		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>
Functional Data	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>	
Segregation Data	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
De novo Data				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
Allelic Data		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>		
Other Database		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
Other Data		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>			

Figure 2.4. The American College of Medical Genetics and Genomics (ACMG) criteria evidence framework. “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” Reproduced from Richards et al. (2015), Figure 1, with permission from Springer Nature, license no. 6292730351329.

2.8. LR-RNA sequencing from disease-relevant skeletal muscle tissue

RNA was extracted from frozen skeletal muscle tissue using the RNeasy Fibrous Tissue Mini Kit and TissueLyser from Qiagen by laboratory technician Alana Lamont; a minimum of 300ng of RNA was obtained for each sample at an RNA Integrity Number (RIN) ≥ 7 . RNA was subsequently processed using the Kinnex full-length RNA kit, thereby generating long-read libraries for sequencing full-length cDNA molecules. This method concatenates several transcripts onto a single HiFi read, increasing sequencing data yield with lower RNA input. Samples were subjected to LR-RNA-seq on the PacBio Revio system at the SickKids TCAG Sequencing Facility at 15M reads/sample. The SPRQ chemistry ($>15\text{kb}$ reads) was employed, a novel technology that uses optimized reagents to increase data yield and sequencing efficiency. LR-RNA-seq data were transferred via a secure portall as described above (Pellecchia, 2024a, 2024b).

First, LR-RNA-seq was also used to assess the functional consequences of variants identified through LR-WGS, thus improving variant interpretation. For prioritized variants identified by the LR-WGS filtering pipeline, LR-RNA-seq coverage, reads, and splicing were visualized and qualitatively assessed in IGV to understand the functional impacts of genomic variants. Secondly, LR-RNA-seq was used to identify candidate genes of interest by detecting transcriptomic abnormalities. This was performed with IsoRanker, a computational framework designed to prioritize candidate genes and isoforms based on aberrant gene expression and isoform usage in long-read transcriptomic datasets (Cheng et al., 2025). IsoRanker was implemented in our cohort by bioinformatics analyst, Feryal Ladha, while plotting code was developed by Dr. Galen Wright. IsoRanker quantifies gene- and isoform-level expression from full-length transcript models using TPM, and compares expression levels between individuals within a cohort. For each gene and isoform in a participant, expression is evaluated relative to the distribution of expression

observed across all other participants, allowing the identification of expression patterns that deviate substantially in a participant. In this study, gene- and isoform-level expression were compared across all samples within the cohort ($n = 19$); there were fewer individuals in this portion of the study, as muscle biopsy was not performed in two participants (NM02 and NM06) and was unavailable for analysis in a third (NM22), as the sample was sent to a different institution. For each participant who underwent LR-RNA-seq, the 25 most upregulated and 25 most downregulated genes and isoforms were selected for manual review and interpretation by the student PI. This threshold was chosen empirically to prioritize the strongest expression outliers for downstream interpretation. Of all expression outliers per individual, genes associated with NMDs or other OMIM phenotypes were prioritized for assessment. For each candidate gene, known gene-disease associations were reviewed to determine relevance to the participant's phenotype, and genes with plausible phenotypic associations were retained. LR-WGS data were then interrogated for rare variants ($MAF < 1\%$) using the reference cohorts described in **Section 2.5**. Identified variants were classified according to ACMG guidelines (Richards et al., 2015). Genes were considered high-priority candidates when there was a strong correlation with the participant's phenotype, as well as a rare genomic variant with some evidence of pathogenicity identified by LR-WGS which also showed concordance with the observed outlier transcriptomic results. For example, splice-altering variants associated with aberrant transcript isoforms, or predicted loss-of-function variants associated with reduced gene expression. High-priority candidate genes and variants were subsequently presented to the research team for collaborative analysis and variant interpretation. Altogether, LR-RNA-seq using IsoRanker aimed to identify candidate genes of interest for LR-WGS filtering, thereby improving the discovery and interpretive power of LR-WGS.

Chapter 3: Results

3.1. Cohort clinical and demographic results

A total of 22 individuals from 20 families were analyzed in this study, comprising 18 singletons and two affected duos (i.e., father-son and brother-brother). Phenotypic information collected via retrospective chart review included 81 standardized NMD-centred fields, including medical and family history, physical examination, genetic and other clinical testing (biochemical, imaging, biopsy, etc.), and the length of the diagnostic odyssey. For select cases with informative results, narrative summaries of clinical information, LR-WGS and LR-RNA-seq findings are presented in **Sections 3.3-3.5**.

The demographics of the cohort are shown in **Table 3.1**. The study consisted of 14 males (64%) and 8 females (36%). Four of the participants had died since study recruitment and consenting occurred; of the remaining participants, the median age was 46.8 years and ranged from 13-84 years. Symptom onset was congenital in 9 participants (41%), pediatric in two participants (9%) and adult in the remaining 11 participants (50%). Self-reported ancestry was available for 21/22 participants. Of 21 individuals, 19 reported European ancestry, comprising the majority (90%) of the cohort. The cohort demonstrated broad clinical heterogeneity, with the involvement of multiple organ systems (**Figure 3.1**). There was a mean of 18.2 HPO terms applied per participant, with a range of 10-38. Three individuals in the study (13.6%) demonstrated a complex multisystemic phenotype. All participants had at least one clinical test result supportive of a myopathy on standardized clinical testing, including ten individuals (46%) with elevated serum creatine kinase (CK), ten individuals (46%) with myopathic electromyography (EMG), 13 individuals (59%) with fibrofatty muscle replacement on magnetic resonance imaging (MRI) and 17 individuals (77%) with myopathic muscle biopsy (**Figure 3.1 B**). Collectively, the findings

highlight the clinical heterogeneity of the cohort while underscoring evidence of the underlying myopathic disease process across participants. More detailed participant information can be found in **Figure 3.2**.

Table 3.1. Select participant clinical and demographic characteristics (*n* = 22).

	Number of participants (%)
Sex	
Male	14 (64)
Female	8 (36)
Onset	
Congenital	9 (41)
Pediatric	2 (9)
Adult	11 (50)
Age in years²	
13-26	7 (32)
50-65	6 (27)
66-84	5 (23)
Average age = 46.8	
Deceased ¹	4 (18)
Population group³	
White ⁴	19 (90)
First Nations	1 (5)
Métis	1 (5)
Arab	1 (5)

¹Deceased at 4, 14, 76 and 76 years of age. ²Age at thesis submission (August 2026).

³Definitions for population groups as derived from Government of Canada (2021). More than one population may be selected for each participant; totals may exceed N. Self-reported ancestry only available for 21 of 22 individuals; percentage calculated over 21 participants. ⁴Including relevant subcategories of French-Canadian, Hutterite and Mennonite.

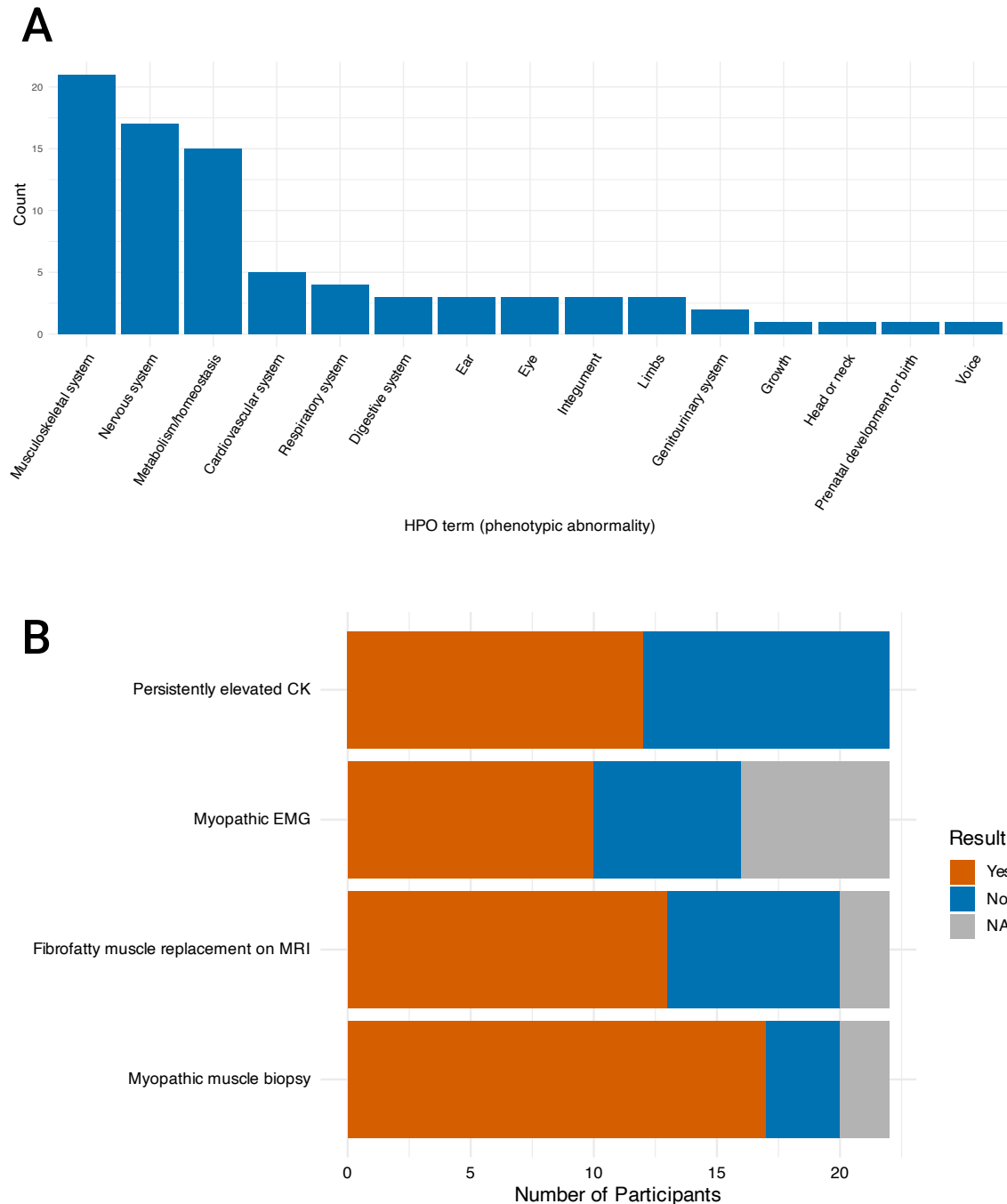


Figure 3.1. NMD cohort ($n = 22$) demonstrates variety in body systems affected and results on standard clinical investigations performed. A) Summary of how frequently various standard phenotypic HPO term counts appear across the NMD cohort to highlight body systems affected. **B)** Cohort distribution of clinical testing findings. CK, creatine kinase; EMG, electromyography; HPO, Human Phenotype Ontology; NMD, neuromuscular disease.

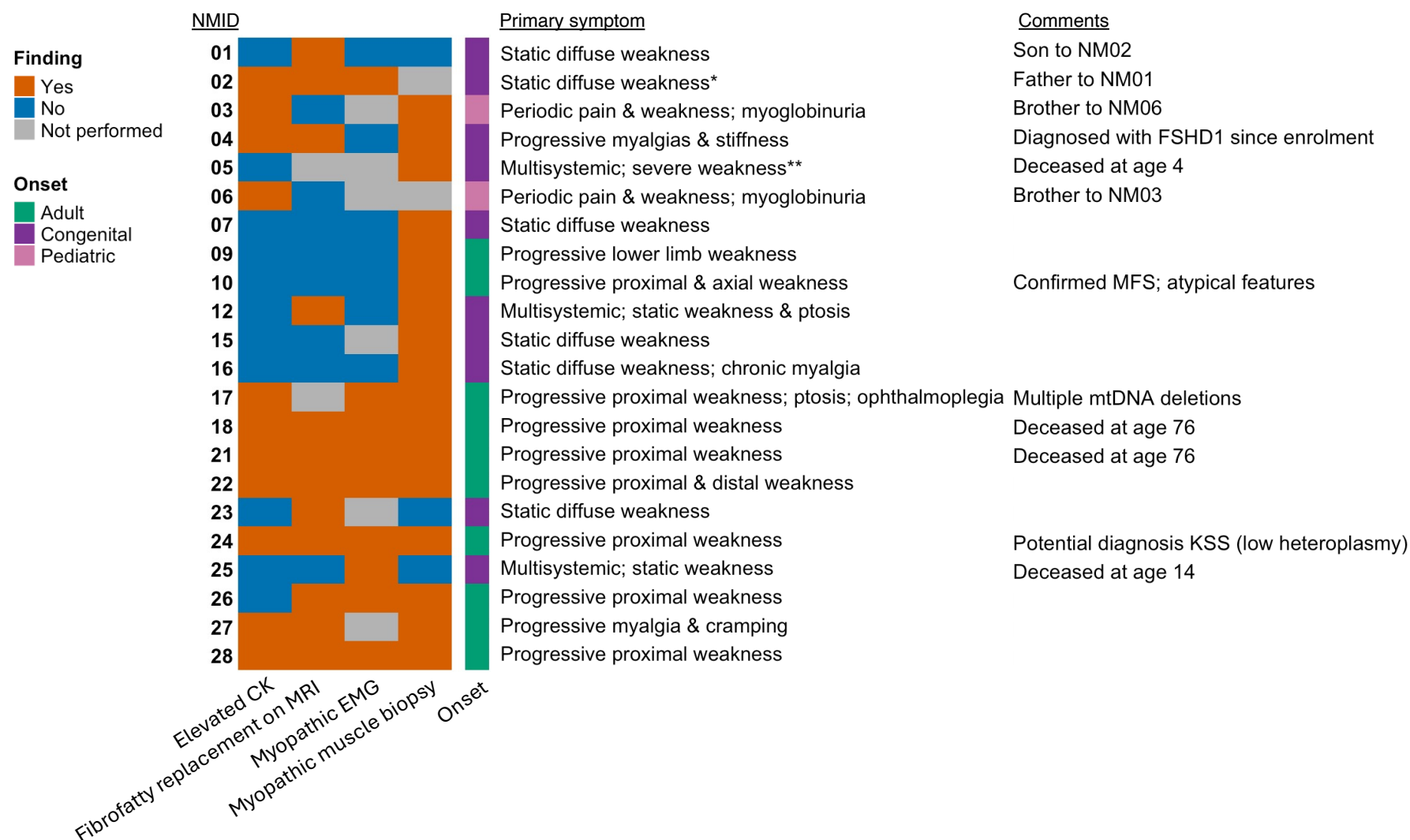


Figure 3.2. Rare disease NMD cohort ($n = 22$) demonstrates diversity of disease onset, features and clinical testing findings. *Mixed myopathy and neuropathy on EMG/NCS; **Abn alpha-dystroglycan stain on muscle biopsy. CK, creatine kinase; EMG, electromyography; FSHD1, Facioscapulohumeral muscular dystrophy; KSS, Kearns-Sayre syndrome; MFS, Marfan's syndrome; MRI, magnetic resonance imaging; mtDNA, mitochondrial DNA.

3.1.1. Length of diagnostic odyssey and genetic testing

The cohort was characterized by a prolonged diagnostic odyssey, despite extensive prior genetic investigations (**Figure 3.3**). On average, participants had experienced symptoms for 20.32 years and had been engaged with genetics services for 12.82 years. Participants had undergone extensive, yet fragmented, genetic investigations, with a mean of 3.95 different genetic tests prior to this study (range: 1-10).

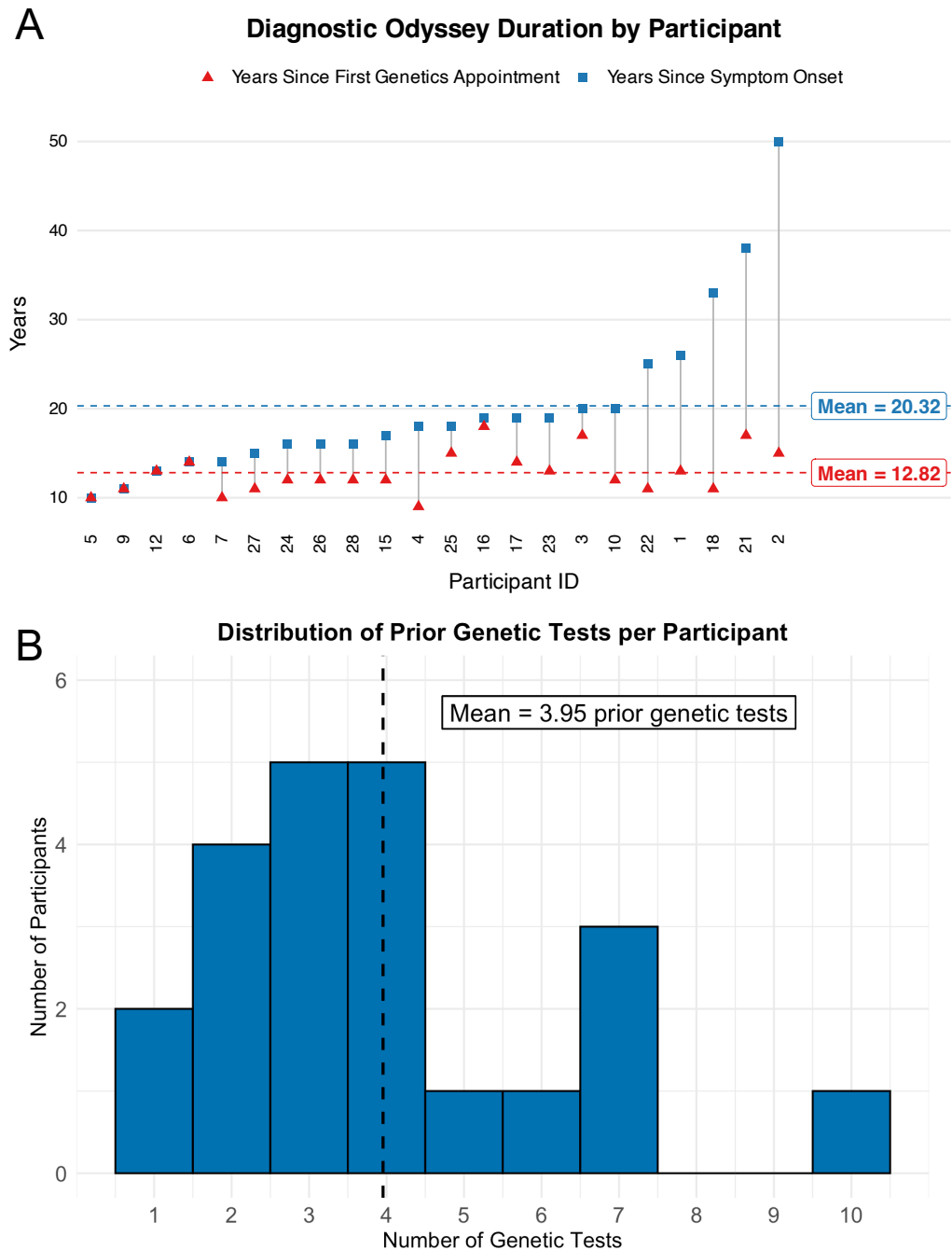


Figure 3.3. NMD cohort ($n = 22$) demonstrates extended length of diagnostic odyssey and increased number of prior genetic testing. (A) Length of diagnostic odyssey duration by participant in years. Participants are ordered on the x-axis from the fewest to the most years since symptom onset. Blue squares indicate years since symptom onset according to participant description. Red triangles indicate years since first genetics appointment. (B) Distribution of prior genetic tests undergone per participant.

Across the cohort, prior genetic investigations included both targeted and broader testing strategies (**Table 3.2**). Prior testing included single-gene sequencing (10/22), repeat expansion analysis (8/22), multi-gene panel testing (17/22) and SR-WES (6/22). In addition, seven participants had undergone dedicated mitochondrial genetic investigations. These analyses included sequencing of nuclear- and mitochondrial-encoded genes from whole blood. Skeletal muscle tissue was used in several participants to perform genotyping of common mitochondrial genetic variants and deletions, as well as mitochondrial depletion analysis. Collectively, these findings demonstrate the diversity of prior genetic testing strategies applied across a clinically heterogeneous neuromuscular cohort.

Table 3.2. Summary of prior genetic testing performed across NMD cohort ($n = 22$).

	Number of participants
Karyotype	5
Chromosomal microarray	5
Exome sequencing	6
Mitochondrial analyses ¹	7
Single-gene testing	
Sequencing	10
Repeat analysis	8
Gene panel	
1-24 gene panel (average = 12.6)	6
43-131 gene panel (average = 58.4)	14
Del/dup gene analysis following panel	4

¹Including mitochondrial genes/deletions, nuclear mitochondrial genes and mitochondrial depletion analysis. NMD, neuromuscular disorder.

Notably, participant NM27 underwent the most prior investigations, with a total of ten genetic assays performed over nine years. Initial testing consisted of several sequential single-gene analyses, including *FKRP* and *CNBP* sequencing, illustrating how targeted approaches can limit

diagnostic yield and prolong the diagnostic odyssey. Broader genetic testing strategies were subsequently applied, including a 46-gene structural myopathy and singleton exome sequencing. However, the limitations of these SRS approaches required three additional targeted assays to evaluate repeat expansion disorders included in the differential diagnosis (via D4Z4 and *DMPK* repeat analyses), as well as investigations into potential structural variation. Lastly, the participant underwent multiple mitochondrial genetic investigations using separate assays on skeletal muscle and whole-blood samples. This illustrates how conventional testing often requires several methods to comprehensively assess different classes of mitochondrial variation. In contrast, long-read genome sequencing has demonstrated the ability to simultaneously detect mitochondrial variants, their breakpoints, and heteroplasmy levels in a single assay (Jadhav et al., 2026). Altogether, these findings demonstrate how a comprehensive, LRS-based approach could reduce the need for multiple sequential genetic investigations by enabling the simultaneous assessment of diverse disease-causing variant types within a single assay.

3.1.2. Genetic diagnoses established by clinical testing before completion of study analyses

By the time of LR-WGS and LR-RNA-seq analysis, four of the participants (18%) had established genetic diagnoses, with pathogenic variants in genes associated with specific genetic conditions identified by clinical testing. Of these individuals, one had a genetic condition completely unrelated to the neuromuscular disease presentation (*MLH1*-associated Lynch syndrome). The other three participants had genetic diagnoses with neuromuscular features. The first (NM10) had a long-standing genetic diagnosis of *FBNI*-associated Marfan syndrome; however, her clinical presentation was atypical for that diagnosis and thus an additional genetic etiology was suspected to capture the full neuromuscular picture (**Section 3.5.3**). The second participant (NM24) received a genetic diagnosis of Kearns-Sayre syndrome (KSS) since study

enrollment; this was made through sequencing and deletion analysis of mitochondrial genes using skeletal muscle tissue. The identified mitochondrial variant was present at <15% heteroplasmy; as most individuals with a mitochondrial disorder have associated mtDNA deletions with heteroplasmy >15%, it was unclear whether this deletion was expected to be associated with a disease phenotype. Further, the patient's clinical presentation was atypical for this diagnosis, given his lack of ptosis and ophthalmoplegia. As such, he was provided with a likely diagnosis of a mild KSS phenotype, but the possibility of an alternate genetic etiology for his neuromuscular symptoms remains. Finally, one participant (NM04) received a likely diagnosis of facioscapulohumeral muscular dystrophy (FSHD) type 1 since initial enrolment in the study. Genetic testing from a blood sample identified a mosaic 31 kb deletion encompassing the entire D4Z4 repeat area on the permissive 4qa haplotype, the significance of which is discussed in **Section 4.3.2.**

3.2. LRS Results

3.2.1. LR-WGS and bioinformatic filtering

LR-WGS was performed for all study participants ($n = 22$) using the PacBio Revio LRS platform, yielding a mean sequencing depth of 34.6X (95% CI: 33.1X-36.0X) and a mean read length of 17.4 kb (95% CI: 16.9-17.9 kb). Variants were annotated across more than 100 genomic features using ANNOVAR (Pellecchia, 2024a, 2024b). VCF files containing >100 variant features were generated and returned by SickKids TCAG. VCF files for each participant included distinct files for each variant category: small variants, SVs, CNVs and STRs.

On average, 5.01 million variants were identified per genome following initial variant calling. **Table 3.3** summarizes the average number of variants remaining after each filtration step for small variants, while **Table 3.4** summarizes the average number of variants remaining after each filtration step for CNVs and SVs. Following filtration, a total of 113 candidate variants remained across the cohort. This comprised 82 small variants, 12 CNVs, and 19 SVs, all retained for downstream clinical interpretation and ACMG-based pathogenicity assessment. Per participant, there was an average of five SNVs (range: 1-9), one CNV (range: 0-3) and two SVs (range: 0-4) following all filtration steps. While all 119 candidate variants were presented to the multidisciplinary review team, 85 small variants, 10 CNVs and 19 SVs were not considered of interest due to factors such as lack of phenotype correlation with the participant, lack of biallelic variants in the case of AR or X-linked recessive conditions and lack of variant pathogenicity evidence.

Table 3.3. Average number of small variants identified from LR-WGS (whole blood) in a cohort of individuals with NMDs ($n = 22$), following subsequent variant filtration steps.

Variant type	Total variants before filtration	gnomAD and GA4K frequency $\leq 0.1\%$	High predicted pathogenicity ³	Gene matches ≥ 1 HPO term	Variant in NMD gene
Small variants	4.96M	107,222 [92,233-133,544]	53 [35-88]	19 [11-23]	4 [1-9]

Variants reported in this table are those identified in less than five participants. Average number of variants in each step was rounded to the closest whole number, with range of variants identified in brackets. ¹gnomAD frequency = <0.001 population allele frequency in gnomAD v4.0 database based on data from short reads; ²GA4K frequency = <0.001 population allele frequency in Genomic Answers for Kids (GA4K) database based on technology-matched data from long reads; ³ variant effect impact = frameshift, stop-gain, splice site or loss of function, OR missense with Revel score ≥ 0.45 , OR predicted splicing variant with relevant dbSNV and SpliceAI scores, OR non-coding RNA with average value of nucleotide-level conservation inferred from Placental Mammalian ≥ 1.3 , average value of nucleotide-level conservation inferred from 100 Vertebratae genome group ≥ 3.9 , or CADD PHRED ≥ 21.1 . CADD, Combined Annotation-Dependent Depletion; GA4K, Genome Answers for Kids; gnomAD, Genome Aggregation Database; HPO, Human Phenotype Ontology; NMD, neuromuscular disorder.

Table 3.4. Average number of SVs and CNVs identified from LR-WGS (whole blood) in a cohort of individuals with NMDs ($n = 22$), following subsequent variant filtration steps.

Variant type	Total variants before filtration	“high quality rare” ¹	In coding region	Gene matches ≥ 1 HPO term	Variant in NMD gene
CNVs	62 [39-83]	5 [1-16]	Filter not applied	1 [0-10]	1 [0-2]
SVs	51,316 [50,572-52,363]	941 [825-1,141]	72 [48-98]	14 [7-21]	1 [0-4]

Columns with bold text indicate the variant filtration steps ultimately used in the bioinformatic filtering pipeline. Variants reported in this table are those identified in less than five participants. Average number of variants in each step was rounded to the closest whole

number. ¹high quality rare = <1% population allele frequency in internal PacBio datasets, >50 bp, <70% overlap with segmental duplications/telomere/centromere. CNV, copy number variant; HPO, Human Phenotype Ontology; NMD, neuromuscular disorder; SV, structural variant.

3.2.2. LR-WGS and TRGT STR genotyping

LR-WGS data were also used to assess for STR pathogenic variation using TRGT analysis, across 34 NMD-associated loci with known pathogenic STRs (**Appendix B**). This analysis identified expanded-repeat alleles in four of 22 participants (18%). Of these expanded repeats, two occurred in *ATXN8OS* in a father-son duo. The expanded alleles consisted of 74 and 72 CTG uninterrupted repeats in NM02 and NM01 (**Section 4.2.2**), respectively. As unaffected alleles have 50 or less repeats, while repeats greater than 54 show pathogenicity with reduced penetrance, these expanded alleles were classified as VUS (Cleary et al., 2021). The remaining two alleles identified by TRGT consisted of heterozygous *FGF14* GAA repeat expansions with internal interruptions. *FGF14* alleles with uninterrupted GAA tracts of 250 repeats or more are associated with AD GAA-*FGF14* ataxia, also known as spinocerebellar ataxia 27B (SCA27B) (OMIM #620174) (Pellerin et al., 2024). However, expanded repeats with interrupted expansions are not considered pathogenic (Pellerin et al., 2024). In the two unrelated individuals highlighted by TRGT, the longest alleles measured were 297 in participant NM09 (19 reads) and 285 repeats in participant NM24 (18 reads) (**Figure 3.4**). GAA purity for both individuals was approximately 84% at this locus, while the threshold defining a pure GAA tract is at 95% (Pellerin et al., 2023). Accordingly, the observed repeat expansions were interpreted as non-pathogenic, without concern that the individuals would develop GAA-*FGF14* ataxia. As such, detection of these two individuals' repeat sizes, with resolution of interruptions, allowed identification of two non-pathogenic alleles, even though one was expanded. Importantly, while SRS can screen for *FGF14* repeat expansions using bioinformatic tools such as ExpansionHunter, LRS is required for accurate allele characterization, including precise repeat sizing and resolution of internal repeat motif configuration (Dolzhenko et al., 2017; Mohren et al., 2024; Rudaks et al., 2025; Stevanovski et al., 2022). Lastly, one individual

of the cohort showed *RFC1* dropout or failure during STR genotyping; although this finding has been identified in individuals with biallelic pathogenic expansion, the participant's lack of phenotypic similarity with the associated condition, Cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS, OMIM #614575) made this possibility unlikely (Höps et al., 2025).

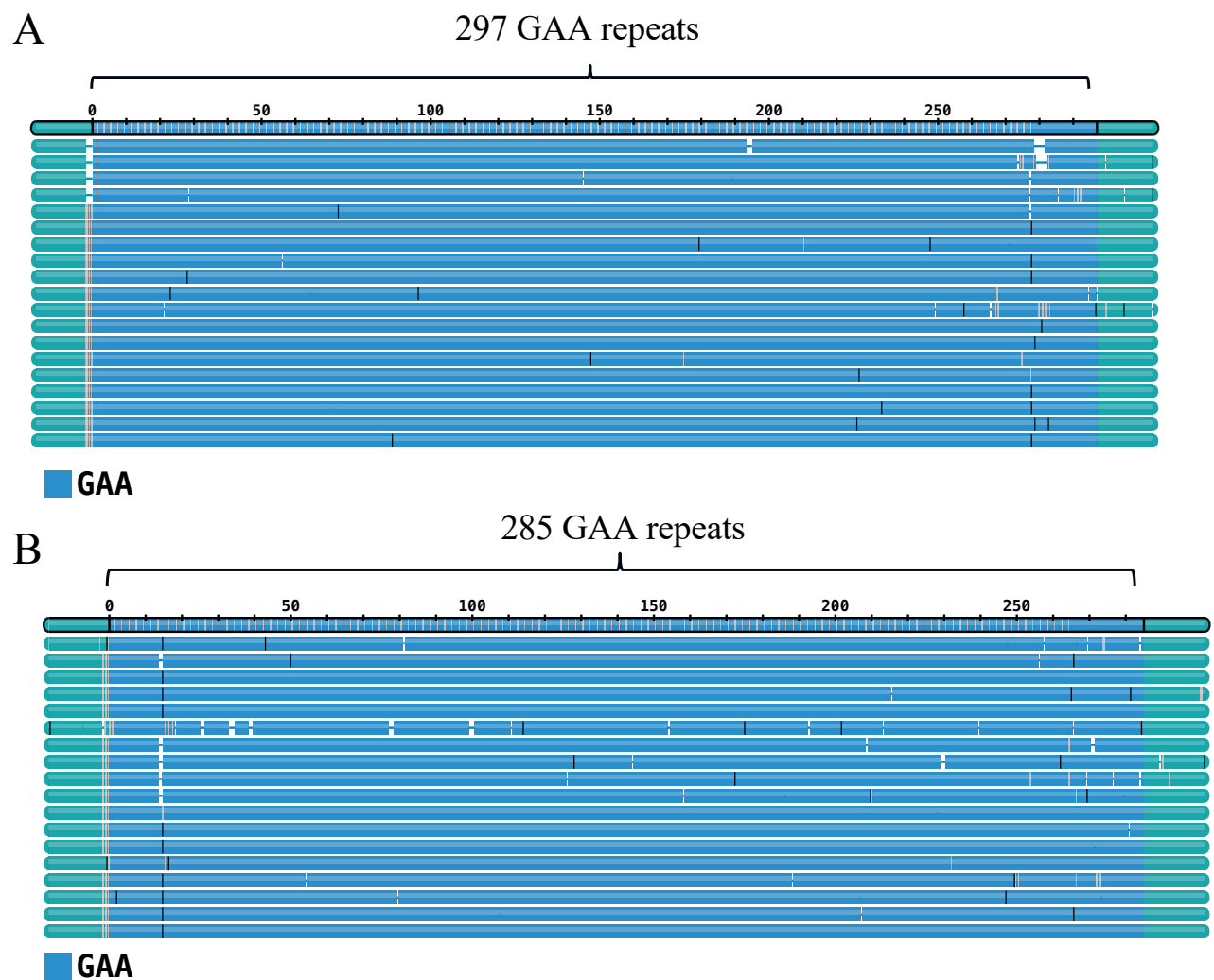


Figure 3.4. LR-WGS resolved GAA STR alleles in *FGF14*, with interruptions modulating pathogenicity. Analysis performed was performed using TRGT with STRchive v2.1.0. TRGT visualizations of repeat expansions are shown for participants (A) NM09 and (B) NM24. Blue segments indicate GAA repeat units, with total repeat length indicated in bracket above each allele. Grey lines denote interruptions within the repeat tract. LR-WGS, long-read whole genome sequencing; STR, short tandem repeat; TRGT, Tandem Repeat Genotyping Tool.

3.2.3. LR-RNA-seq and IsoRanker

LR-RNA-seq was performed on skeletal muscle biopsy samples from 19 participants using the PacBio Revio platform with the Kinnex workflow. Sequencing generated a mean of 15 million reads per sample, with a mean read length of 15 kb. Transcript classification and filtering were performed using Pigeon software, resulting in a mean of 422,003 $\pm$ 65,559 input isoforms per sample and ultimately yielding 110,400 $\pm$ 13,887 high-quality transcripts spanning 16,032 $\pm$ 1,171 genes.

IsoRanker, a framework for detecting gene expression and isoform-level outliers in LR-RNA-seq data, was applied to all participants with data from muscle biopsies ($n = 19$). Manual review of IsoRanker results was performed for all participants who had not yet yielded a high-interest candidate variant through LR-WGS analysis ($n = 15$). Through this analysis, the 25-most upregulated and 25-most down-regulated genes and isoforms were identified for each individual compared to the remainder of the cohort. As only outlier calls which occurred in NMD or other OMIM genes were analyzed manually for phenotypic relevance, the total number of outliers in these two categories, as well as the average per participant, are summarized in **Table 3.5**. At the gene expression level, IsoRanker identified 945 genes across 950 outlier calls among the 19 individuals. An average of two NMD and seven other OMIM gene outliers were identified per participant in this analysis, with ranges of 1-3 and 3-11, respectively. Across the cohort, there were a total of 26 gene expression outliers in NMD genes and 129 gene expression outliers in other OMIM genes. The vast majority of gene expression outliers were unique to individual participants, with only two genes (*AKAP17A* and *SGTA*) identified as outliers in more than one individual; this finding is consistent with IsoRanker's assumption that individuals in rare disease cohorts will have distinct underlying genetic alterations.

At the isoform level, a total of 900 isoforms were identified across the cohort, mapping to 583 unique genes. An average of 10 isoform expression outliers were identified in NMD genes (range: 3-22), while an average of eight isoform expression outliers were identified other OMIM genes (range: 1-15). Across the cohort, there was a total of 177 isoform expression outliers identified in NMD genes, which occurred across 53 unique genes. There was also a total of 158 isoform expression outliers identified across the cohort in other OMIM genes, which mapped to 115 unique genes. The genes with the highest numbers of outlier isoforms were *TTN* ($n = 67$ isoforms), *NEB* ($n = 23$), *TNNT3* ($n = 9$), *LSM14A* ($n = 8$), and *CAST* ($n = 7$).

Table 3.5. Average and total gene and isoform expression outliers in phenotypically relevant genes across NMD cohort using IsoRanker analysis ($n = 19$).

	Gene expression outliers		Isoform expression outliers	
	NMD genes	Other OMIM genes	NMD genes	Other OMIM genes
Average per participant [Range]	2 [1-3]	7 [3-11]	10 [3-22]	8 [1-15]
Total across cohort	26 ¹	129 ²	177 ³	158 ³

Average values represent the mean number of outlier calls per participant, with the range shown in brackets. Total values represent the total number of outlier calls across all 19 participants, with the number of unique genes represented indicated in the corresponding footnote. NMD genes refer to genes previously associated with NMDs, as per the GeneTable of Neuromuscular Disorders downloaded in November 2024. Other OMIM genes which are not included in the NMD GeneTable list, but which are associated with other phenotypes or conditions in humans according to the OMIM database. ¹Across 26 unique genes. ²Across 129 unique genes. ³Across 53 unique genes. ⁴Across 115 unique genes. NMD, neuromuscular disorder; OMIM, Online Mendelian Inheritance in Man.

IsoRanker was validated in this cohort by its ability to detect known pathogenic transcriptional signatures in two individuals, correctly identifying aberrant gene expression consistent with previously identified genetic diagnoses. For NM18, IsoRanker correctly identified *MLH1* as a down-regulated gene expression outlier (Z-score -2.5) (**Figure 3.5 A**), consistent with the identified heterozygous nonsense *MLH1* variant identified in NM18 which, through introducing a premature termination codon, is expected to result in loss of function through degradation of the variant transcript. Meanwhile, for NM10, IsoRanker identified *FBNI* as a splicing outlier (Z-score 4.2) (**Figure 3.5 B**). This finding is concordant with the individual's clinical diagnosis of Marfan syndrome and identified known pathogenic splicing variant in *FBNI*. Together, these findings demonstrate that IsoRanker can identify transcriptomic abnormalities that are concordant with established genetic or clinical evidence, supporting its utility for identifying disease-relevant molecular abnormalities.

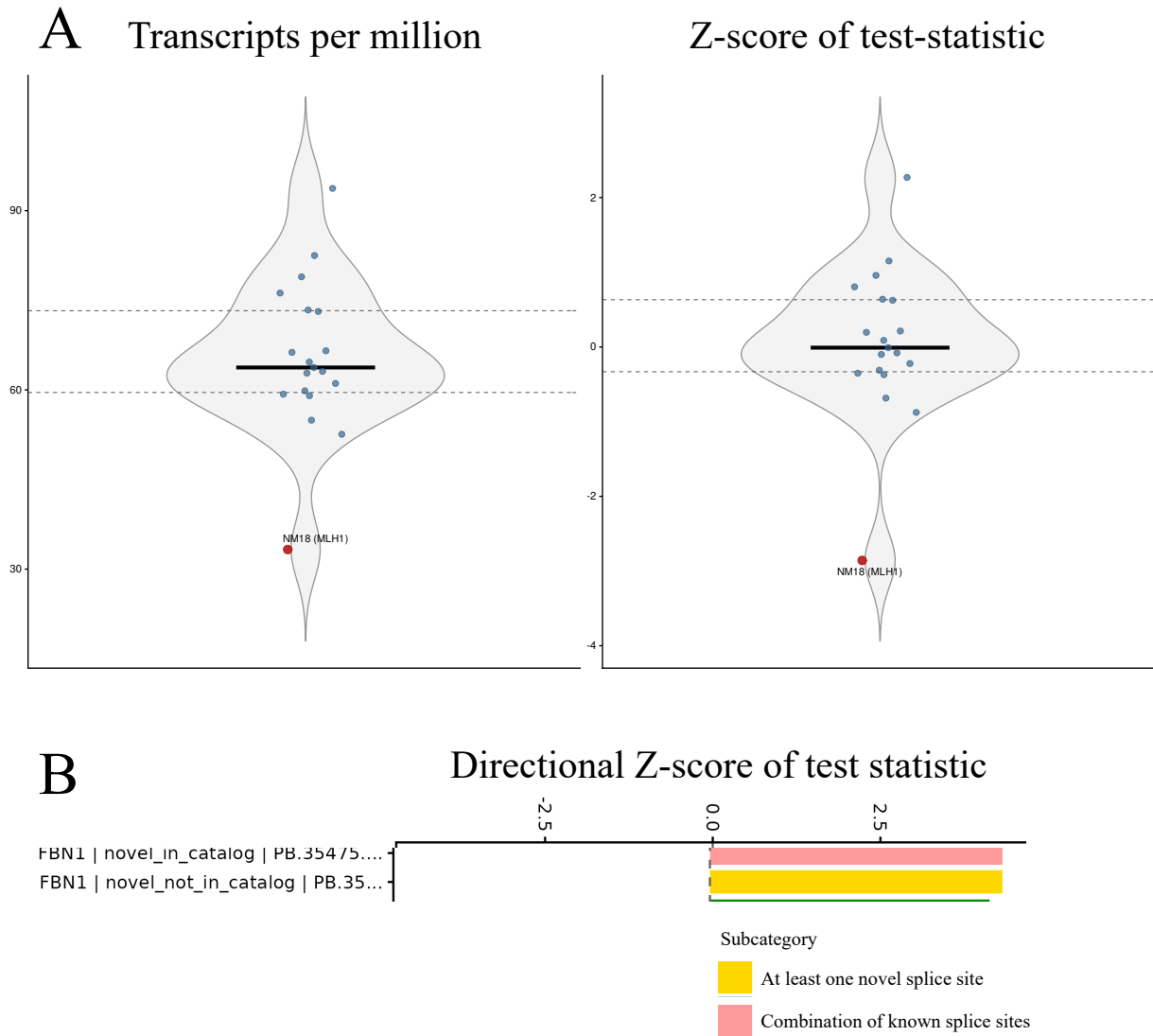


Figure 3.5. IsoRanker-guided LR-RNA-seq from skeletal muscle biopsy enables identification of genes of interest in individuals with known genetic diagnoses to be used as positive controls within the NMD cohort. A) Gene-level expression outliers in *MLH1*; *MLH1* showed reduced gene expression in NM18 (33.3 TPM compared to the rest of the cohort (mean 65.5 TPM), with a directional Z-score of -2.5, test statistic of 3.9 and z-score of test statistic of 3.7. B) Known *FBN1* splicing variant is identified as isoform expression outlier by IsoRanker in NM10 using a directional Z score $\geq [2.5]$. The aberrant splicing products are expressed at 6.8 and 5.5 TPM in NM10's sample, compared to an average of 0.4 and 0.3 mean TPM across the other individuals in the NMD cohort. Splicing products demonstrate test statistic scores of 3.0 and 2.7, and Z-scores of test statistics of 4.2 and 4.2, respectively. Bar colors indicate isoform subcategory according to PacBio Iso-Seq annotation (Guizard et al., 2023). LR-RNA-seq, long-read RNA sequencing; TPM, transcripts per million.

3.2.4. Cohort diagnostic outcomes

The overall outcomes of the study are summarized in **Figure 3.6**, while findings per study participant are summarized in **Table A4**. LR-WGS alone identified high-interest variants in four of 22 participants (18%) (**Section 4.1**). This term refers to variants with sufficient variant-level evidence, gene-disease association, and phenotypic correlation to be considered likely explanatory for the participant's phenotype. These cases can be considered “solved”, pending confirmation of variants in an accredited clinical-grade laboratory. Of the four solved cases, three associated variants were classified as “Pathogenic/Likely Pathogenic” and consisted of two CNVs and one missense variant (**Sections 4.1.1** and **4.1.3**, respectively). The remaining solved case was a small missense variant classified as a “*VUS*” (**Section 4.1.2**). Subsequently, LR-RNA-seq and IsoRanker analysis supported the identification of biallelic high-interest variants in a fifth participant (**Section 4.1.5**); these were classified as “Pathogenic/Likely Pathogenic”.

LR-WGS identified uncertain candidate variants in four additional participants (**Section 4.2** and **4.1.4**); these refer to cases with insufficient evidence to confirm the variant as causal for the participant's phenotype. Of these, one variant was an SV classified as a “*VUS*” (**Section 4.2.1**), one variant was a predicted splicing variant (**Section 4.1.4**), and two variants consisted of an expanded repeat allele seen in the father-son duo (**Section 4.2.2**). The remaining variant was a predicted splicing variant; subsequent visualization and IsoRanker analysis of LR-RNA-seq data provided functional evidence of splicing, thereby upgrading this variant to a high-interest variant (**Section 4.1.4**).

Lastly, in addition to the solved cases, multi-omic LR-WGS and LR-RNA-seq also provided functional information to support a working diagnosis without an established gene variant, and an established mosaic diagnosis, in two further participants (**Section 3.5**).

In the present study, eleven individuals had no variant of interest identified and were therefore considered unsolved. Of these, ten also underwent IsoRanker analysis, which did not identify a gene of interest with a corresponding pathogenic LR-WGS variant. The unsolved group differed descriptively from the group of individuals with solved or uncertain candidate variants. Unsolved individuals were more likely to have adult-onset phenotypes (64%, 7/11), whereas individuals with solved or uncertain candidate variants were more likely to have congenital- or pediatric-onset phenotypes (64%, 7/11). The solved or uncertain candidate group also included three individuals (27%) with complex syndromic phenotypes, whereas none in the unsolved group had complex syndromic phenotypes. Additionally, four individuals (36%) in the solved or uncertain candidate group belonged to familial duos (brother-brother or father-son), compared with none in the unsolved group. No notable differences in ancestry were observed between groups; all individuals in the unsolved group were of self-reported European ancestry.

In sum, the study identified high-interest variants in six of 22 participants (27%), uncertain candidate variants in three participants (14%), and generated disease-related transcriptomic profiles for two participants (9%). For the remaining 11 participants (50%), no gene or variant of interest was identified (**Figure 3.6**). Ultimately, the combined multi-omic approach in this study allowed enhanced identification and interpretation of multiple potentially disease-causing variant types, while also providing transcriptomic evidence to support disease characterization.

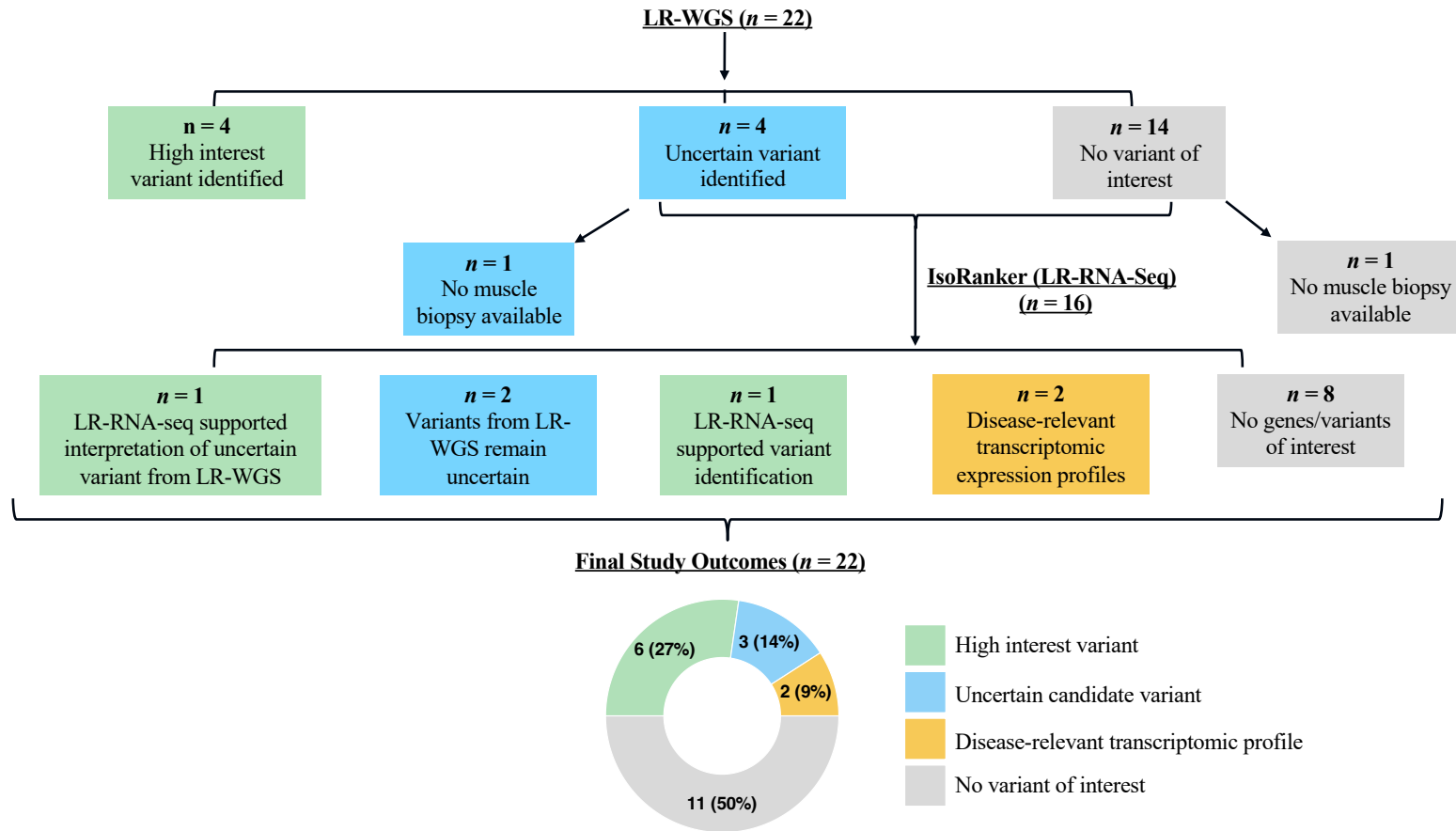


Figure 3.6. Complementary LR-WGS and LR-RNA-seq improved diagnostic yield and identified functional disease-associated findings in a cohort of individuals with neuromuscular disorders. The upper panel is an overview of participant outcomes throughout the combinatorial sequencing and analysis strategy employed (LR-WGS and LR-RNA-seq). The lower panel summarizes study outcomes for discovery of variants and functional information. LR-WGS, Long-read whole genome sequencing; LR-RNA-seq, Long-read RNA sequencing.

Chapter 4: Case descriptions

The following section will report cases with significant findings, including variants strongly considered causal for the participant's phenotype (i.e. solved) (**Section 4.1**), uncertain candidate variants (**Section 4.2**) and disease-relevant functional transcriptomic information (**Section 4.3**). Each case in this chapter represents an individual or a duo (e.g., father-son or brother-brother). Each case result includes a description of the participant's phenotype (including clinical features, symptoms, and investigations), prior genetic testing performed and results, and LR-WGS variant filtration and LR-RNA-guided filtering in the present study, as applicable.

4.1. Case descriptions: Solved cases demonstrating diagnostic resolution through LR-WGS and LR-RNA-seq

The cases are considered solved based on the finding of a genomic variant with sufficient evidence to suggest or determine pathogenicity, as well as elevated clinical correlation with the participant's phenotype.

4.1.1. Complementary LR-WGS and LR-RNA-seq resolved a large-scale copy number variant in brother duo (NM03 and NM06)

Family History: NM03 (age 21) and NM06 (age 17) are full brothers. Their parents are non-consanguineous individuals of European (Dutch, Mennonite, Scottish, English, Irish and Spanish) descent. They have three full siblings, two brothers and one sister, with no neuromuscular features.

Clinical Description: Standardized phenotypic (HPO) terms representing individuals' neuromuscular features are listed in **Table 4.1**. Phenotypes for both NM03 and NM06 were characterized by episodic exercise- and temperature-related neuromuscular symptoms, including cramping, pain, stiffness, and reduced exercise tolerance, with onset at ages one and two, respectively. On physical examination by a clinical geneticist at ages five and 13, NM03 and NM06 both demonstrated slight-to-mild proximal muscle weakness, including the shoulder abductors and hip flexors, and NM06 demonstrated a Trendelenburg gait pattern. Both individuals also demonstrated some joint hypermobility, while NM06 also exhibited a mild valgus deformity at the ankle. Despite these similarities, NM03 exhibited a somewhat more pronounced neuromuscular presentation; from infancy he displayed both fine and gross motor delays, with early gait abnormalities including a waddling gait with lack of knee flexion. By most recent genetics evaluation at age 16, NM03's proximal weakness and episodic neuromuscular symptoms persisted, requiring use of a wheelchair for ambulation during periods of severe myalgias. By age 16, NM03 also demonstrated lumbar scoliosis. In contrast, NM06 did not display any motor delays, and by the most recent genetics evaluation at age 13, his episodic symptoms had significantly reduced in frequency and severity, although sporadic post-exertional myalgia and reduced sweating persisted.

Clinical investigations were supportive of myopathy in both individuals, including persistently elevated serum CK for both individuals and elevated serum myoglobin in NM03 (**Table 4.2**). Only NM03 had a muscle biopsy performed; this demonstrated myopathic features, including a slight increase in proportion of fibers with internal nuclei, rare atrophic and degenerating fibers, small foci of inflammatory cell aggregates, decreased myoadenylate deaminase activity and abnormal patchy reduction in alpha-dystroglycan. In sum, the brother duo exhibited a shared pattern of childhood-onset, episodic, temperature-sensitive myopathic symptoms alongside myopathic clinical investigations and proximal weakness on physical examination.

Table 4.1. Human Phenotype Ontology (HPO) terms selected for NM03 and NM06 based on their clinical presentation.¹ The presence of an ‘X’ indicates that the term is applicable to that individual.

HPO Term Description	HPO Code	NM03	NM06
Exercise intolerance	HP:0000118	X	X
Exercise-induced myalgia	HP:0003738	X	X
Proximal muscle weakness	HP:0003701	X	X
Cold-induced muscle cramps	HP:0003449	X	X
Exercise-induced leg cramps	HP:0008991	X	X
Easy fatiguability	HP:0003388	X	X
Gait disturbance	HP:0001288	X	X
Elevated circulating creatine kinase concentration	HP:0003236	X	X

Reduced muscle myoadenylate deaminase activity	HP:0034529	X	X
Abnormal muscle fiber morphology	HP:0004303	X	
Reduced muscle fiber alpha-dystroglycan	HP:0030099	X	
Muscle stiffness	HP:0003552	X	
Hip flexor weakness	HP:0012515	X	
Neck flexor weakness	HP:0003722		X
Pelvic girdle muscle weakness	HP:0003749		X
Motor delay	HP:0001270	X	
Areflexia of lower limbs	HP:0002522	X	
Pes planus	HP:0001763		X
Myoglobinuria	HP:0002913	X	

¹The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3. Terms ordered from first to last in the table according to the order used in the HPO hierarchy, beginning with terms applicable to both individuals.

Table 4.2. Summary of neuromuscular investigations in NM06 and NM03.

<u>NM03</u>			<u>NM06</u>	
Investigation	Finding	Age ¹	Finding	Age ¹
Serum CK	Persistently elevated ²	≥ 3	Persistently elevated ³	7-11
Serum myoglobin	Elevated ⁴	8	NA	
Muscle MRI	Normal (thighs and calves)	9 and 11	Normal (calves)	8
EMG	Not performed		Not performed	
Muscle biopsy	Myopathic features ⁵	4	Not performed	
NCS	Not performed		Not performed	

¹Age in years; ²283-5268 U/L (reference: <350 U/L); ³901-1968 U/L (reference: <350 U/L); ⁴88 ug/L (reference: < 50ug/L). ⁵Decreased myoadenylate deaminase activity, small foci of inflammatory cell aggregates, slight increase in portion of fibers with internal nuclei, rare atrophic and degenerating fibers, abnormal patchy reduction in alpha-dystroglycan. CK, creatine kinase; EMG, electromyography; MRI, magnetic resonance imaging; NA, not available/not found during chart review; NCS, nerve conduction studies.

Prior Genetic Testing (NM03 and NM06): NM03 underwent three forms of genetic testing, while NM06 underwent one. NM03 underwent karyotyping and WES, both of which were normal. NM03's muscle biopsy results demonstrating decreased myoadenylate deaminase activity prompted germline sequencing of *AMPDI*; this detected two compound heterozygous variants: c.34C>T (p.Q12X) and c.143C>T (p.48L), which were later confirmed in NM06 as well. Although the associated condition (AR myopathy due to myoadenylate deaminase deficiency, OMIM #615511) showed phenotypic correlation with NM03 and NM06, the two variants occur at increased frequency in reference populations (gnomAD maximum frequency 12.54% and 13.52%, respectively) and have multiple classifications in ClinVar as benign (Variation IDs: 18271, 50897). Thus, the *AMPDI* variants were of low or uncertain likelihood for causing NM03 and NM06's phenotypes, and an alternative genetic diagnosis was suspected.

LR-WGS Variant Filtration: NMD gene-based variant filtration yielded **four variants** for NM03 and NM06 each. Of these four variants, two were present in both NM03 and NM06. The first, a small variant, was not prioritized as a variant of interest due to a lack of phenotypic correlation. The second variant, a CNV, was a hemizygous 346 kb *DMD* deletion (GRCh38 chrX:31730001-32076000) and was prioritized as a variant of interest for both individuals based on phenotypic correlation with *DMD*-associated dystrophinopathies (OMIM #300376, 310200) (**Table 4.3**). This CNV was confirmed in LR-WGS reads using IGV, with no coverage observed in the deletion area with clear breakpoints in both NM03 and NM06 (**Figure 4.1**). Further, the CNV was visualized using LR-RNA-seq from NM03's muscle biopsy (**Figure 4.2**).

Table 4.3. Pathogenicity and phenotype information for hemizygous *DMD* copy number deletion variant in NM03 and NM06.

Gene Chr	Variant	Zygos- ity	Coding elements	ACMG pathogen- icity	ClinVar pathogen- icity	OMIM phenotype (Inheritance)
<i>DMD</i> ChrX	346kb deletion: 31,730,001- 32,076,000	Hemi- zygous	Exons 45-51 (/79)	Pathogenic ¹	NA	Duchenne/Becker muscular dystrophy (XLR)

¹Applied criterion: PVS1, PS4 and PP3. ACMG, American College of Medical Genetics; Chr, chromosome; NA, not available; XLR, X-linked recessive.



Figure 4.1. IGV snapshot of LR-WGS read alignments demonstrates 346kb *DMD* deletion in NM06 and NM03 (chrX:31,730,001-32,076,000). Deletion is indicated by the red bracket. Exons are denoted by vertical lines on the Human GRCh38 sequence; exons within or directly flanking the deletion are denoted with dark blue arrows or bracket, and intron 44 is denoted with a light blue bracket. DEL, deletion; chr, chromosome; GRCh38, Genome Reference Consortium Human Build 38; LR-WGS, long-read whole genome sequencing.

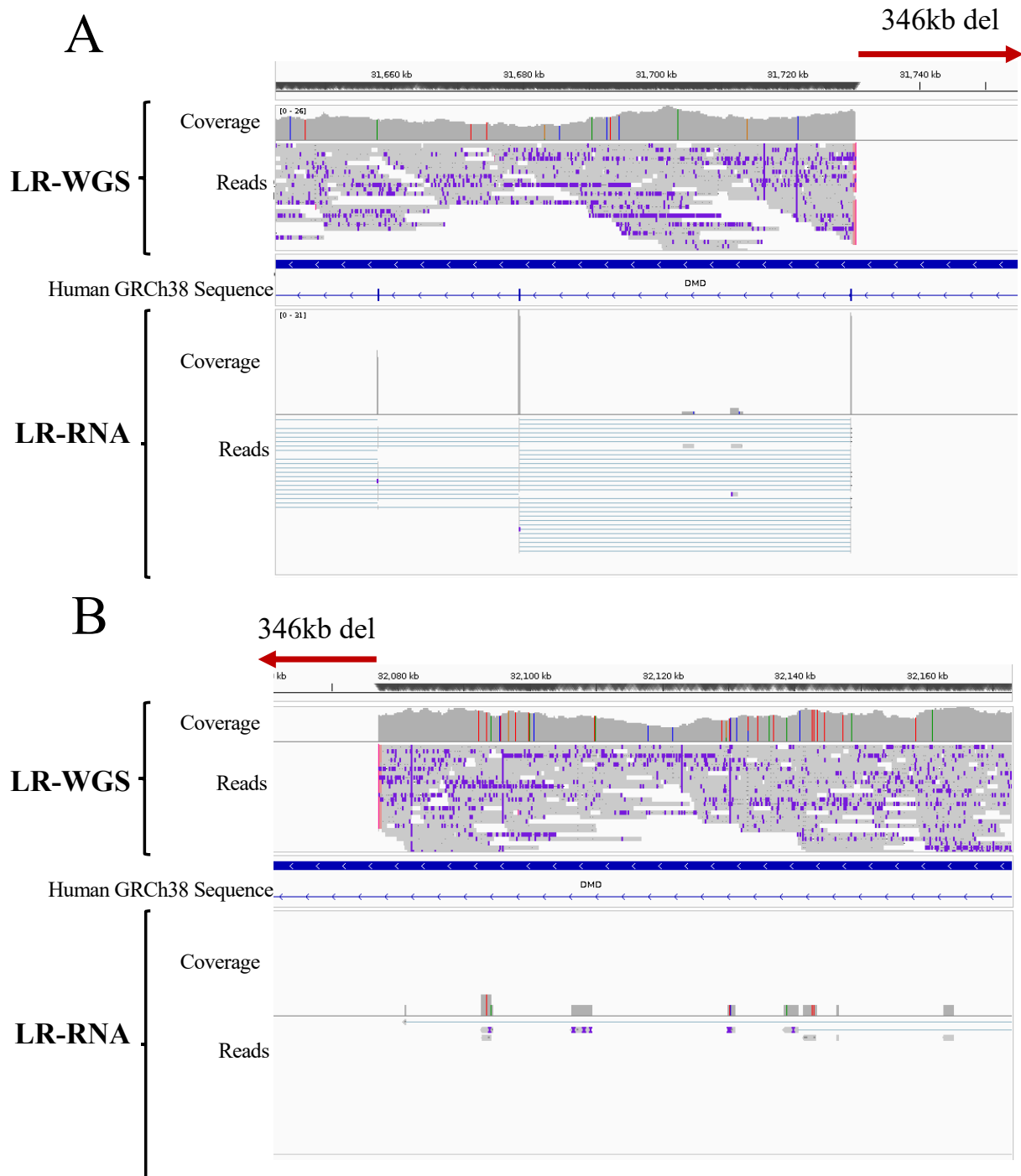


Figure 4.2. IGV snapshot of LR-WGS and LR-RNA-seq supports 346kb *DMD* deletion in NM03 (GRCh38 chrX:31,730,001-32,076,000). Deletion is indicated by the red arrow. A) Demonstrates start (5') of deletion at the locus chrX:31,730,001. B) Demonstrated end (3') of deletion at the locus chrX:32,076,000. DEL, deletion; chr, chromosome; GRCh38, Genome Reference Consortium Human Build 38; IGV,

Integrative Genomics Viewer; LR-RNA-seq, long-read RNA sequencing; LR-WGS, long-read whole genome sequencing.

DMD is located on chromosome Xp21.2-p21.1 and produces multiple tissue-specific isoforms of dystrophin expressed in skeletal muscle, cardiac muscle, the central nervous system and other human tissues (Boyce et al., 1991; GTEx Consortium, 2020; Hoffman et al., 1987; Scott et al., 1988). Dystrophin is a large protein which stabilizes the sarcolemma during muscle contraction by acting as a mechanical link between the intracellular actin cytoskeleton and the extracellular matrix (Duan et al., 2021). The *N*- and *C*-terminal binding domains are connected by a central rod domain encoded by repetitive elements in exons 8 to 64 (Verhaart Aartsma-Rus, 2019). Without functional dystrophin, the lack of cellular scaffold during repeated cycles of muscle contraction causes membrane damage, leading to inflammation and progressive muscle degeneration (Duan et al., 2021). As such, pathogenic variants in *DMD* that cause dystrophin deficiency are associated with a spectrum of progressive muscle disorders: Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) (Escobar-Huertas et al., 2024).

Both DMD and BMD are characterized by progressive skeletal muscle weakness, with initial predominance in proximal and lower-limb muscles (Flanigan, 2014). The study of genotype-phenotype correlations in dystrophinopathies has established the reading-frame rule, whereby variants that disrupt the reading frame (e.g., frameshift, nonsense) result in absent dystrophin expression and cause the more severe DMD (Hoffman et al., 1988). In DMD, symptom onset occurs before age five with gross motor delays and rapidly progressive proximal weakness (Darras et al., 2022). Loss of ambulation occurs by age 12 and mortality results in the third decade due to respiratory insufficiency and progressive cardiomyopathy (Darras et al., 2022). In contrast, in-frame variants, most commonly multi-exon deletions, permit production of a shortened but

partially functional protein and are associated with the milder BMD phenotype (Monaco et al., 1988). In BMD, symptom onset occurs at age 12 (range: 5-60 years old), including more slowly progressive proximal muscle weakness, exercise-induced cramping or fatigue and gross motor delays (Thada et al., 2024). Life expectancy in individuals with BMD is 40 to 50 years, most commonly due to dilated cardiomyopathy; respiratory insufficiency is much less frequent in BMD than in DMD (Schiava et al., 2024). Consistent with the reading-frame rule, the majority of pathogenic variants in individuals with BMD are in-frame deletions affecting the distal portion of the rod domain (exons 45 to 70), notably deletions of exons 45 to 47 and 45 to 48 (Comi et al., 1994; Nakamura et al., 2020; Ricci et al., 2024).

The detected *DMD* CNV is a hemizygous, in-frame deletion located on chromosome Xp21.1:31,730,001-32,076,000, spanning exons 45 to 51 of the *DMD* gene. According to ACMG pathogenicity criteria for sequence variants, this variant was classified as “***Pathogenic***”. **PVS1** was applied as a multi-exon deletion occurs in the central rod domain, a functionally important portion of dystrophin, deletion of which is consistent with the established molecular mechanism underlying BMD (Abou Tayoun et al., 2018; Delalande et al., 2018). Furthermore, *DMD* is highly constrained for loss-of-function variation (pLI = 1.0; observed/expected LoF ratio = 0.15 (Karczewski et al., 2020)) and loss of dystrophin function is a well-established disease mechanism in both DMD and BMD (Jin et al., 2024). Secondly, **PS4** was applied: the prevalence of the variant in affected individuals is significantly higher than in controls. Although the frequency of the variant in gnomAD is unknown due to the short-read methodology of this database, the variant is absent from common structural variants in the NCBI dbVar dataset (**Figure 4.3**) (Lappalainen et al., 2013). Further, substantial overlap was observed between the present deletion and previously reported pathogenic copy-number losses in both ClinVar and the Database of Genomic Variants

(**Figure 4.3**). In addition, an in-frame, hemizygous *DMD* deletion with 84% overlap to the present variant is reported in Decipher and classified as *Likely Pathogenic* (Patient: 271765). Lastly, **PP3 – Medium** was applied: co-segregation with disease in multiple affected relatives. The variant segregated with both affected male individuals in the duo, thereby following X-linked inheritance of *DMD* dystrophinopathies and suggesting inheritance from an asymptomatic carrier mother. Confirmation of the variant in the mother would provide additional evidence of segregation and further strengthen the pathogenic interpretation.

Additional contextual evidence supported the biological plausibility of this variant. The 346 kb deletion is located within the exon 45 to 55 region of *DMD*, one of two major mutational hotspots in the gene, which is responsible for more than half of all reported *DMD* deletions (Bladen et al., 2015; Geng et al., 2021). Further, the proximal breakpoint occurs within intron 44, the largest intron of *DMD* and a recognized recurrent site of genomic rearrangement (Geng et al., 2021). Using LRS methods, Geng et al. (2021) demonstrated that the genomic instability in this area may originate from the high density of repetitive elements (e.g., long interspersed nuclear elements, Alu sequences) and secondary DNA structures (e.g., hairpin loops). Localization of the breakpoint within this known instability hotspot provides a plausible mechanism for its origin earlier in the family lineage.

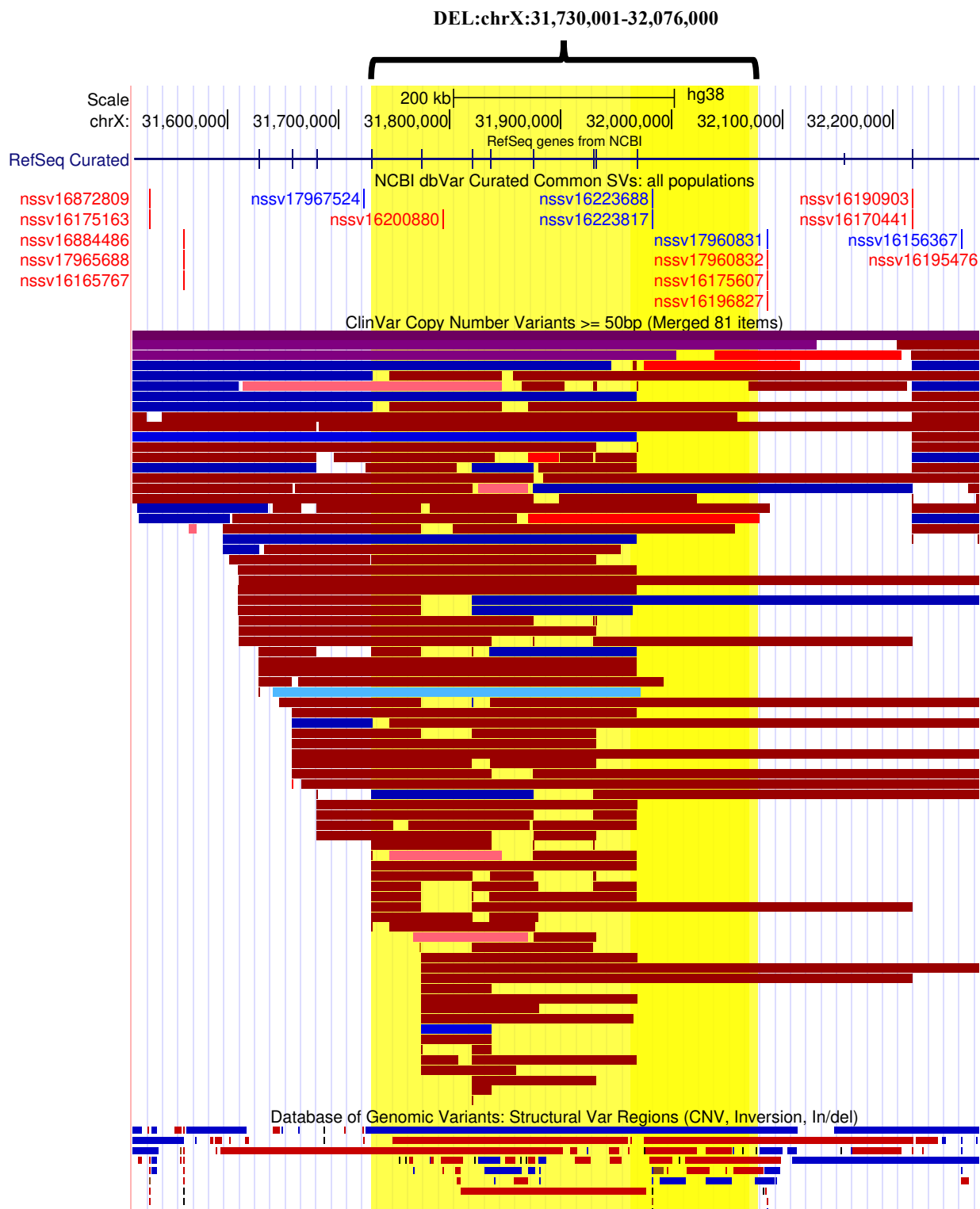


Figure 4.3. UCSC Genome Browser snapshot of *DMD* 346 kb deletion in NM03 and 06 demonstrates genomic overlap with ClinVar-reported deletions, and minimal overlap with common structural variants in healthy populations (NCBI

dbVar). The deletion is located at GRCh38 chrX:317300001-3207600. The range of the deletion is indicated in yellow. Located underneath “Clinvar Copy Number Variants >= 50bp...”, CN gains are in blue and CN losses of uncertain significance in red and CN losses of pathogenic classification in dark red. DEL, deletion; CN, copy number; CNVs, copy number variants; GRCh38, Genome Research Consortium human build 38; NCBI dbVar, National Center for Biotechnology Information Database of Genomic Structural Variation; UCSC, University of California Santa Cruz.

Clinical Correlation and Conclusions: After review by our multidisciplinary committee, the *DMD* CNV was considered a strong candidate for the NMD phenotypes of NM03 and 06. There was strong concordance of variant type (in-frame multi-exon deletion) with BMD, as well as strong phenotypic correlation with the participants’ features (pediatric-onset, proximal weakness with gross motor delay, exercise-induced pain or fatigue, elevated serum CK and myoglobin) (Bladen et al., 2015; Geng et al., 2023). Dystrophin immunostaining on NM03’s muscle biopsy was normal using three separate antibodies to stain the C-terminal, N-terminal and internal regions of the dystrophin protein. At the time of assessment this finding was incorrectly considered to have ruled out *DMD* variation as a cause for NM03’s phenotype. However, all three regions assessed were unaffected by this particular in-frame deletion and, as such, the normal immunostaining result did not exclude the presence of *DMD* deletion. The in-frame *DMD* deletion identified by this study would be expected to produce a shortened dystrophin protein that remains detectable by immunohistochemistry, despite reduced functionality. This also demonstrates the limitations of clinical testing in capturing the broad heterogeneity of genetic and molecular mechanisms underlying NMDs, and the utility of comprehensive omics technologies in resolving diagnoses. While *DMD* would have been assayed on NM03’s previous research-based SR-WES, it is plausible that the deletion was not identified due to bioinformatic tools misinterpreting the lack of coverage across exons to reduced or no coverage by SRS, rather than a true deletion. As the variant

was identified in a research setting, its presence was confirmed in both individuals by chromosomal microarray in a clinically accredited laboratory, while two unaffected siblings (one male, one female) did not demonstrate the CNV on chromosomal microarray.

Identification of the *DMD* variant holds clinical implications for both NM03 and NM06, as well as relatives. This genetic finding may facilitate early initiation of corticosteroid treatment, which has shown a meaningful benefit in delaying loss of independent ambulation, ability to navigate stairs and lifting hand to mouth to feed independently (Biggar et al., 2006; Matthews et al., 2016). With genetic diagnosis of BMD, management for common complications and early intervention is possible, notably comprehensive and ongoing cardiac examination for DCM surveillance (Finsterer & Stöllberger, 2008). Lastly, counselling regarding potential future disease progression (i.e. loss of ambulation and respiratory involvement) and supportive psychological care is indicated (Manzo et al., 2025; Mori-Yoshimura et al., 2019). Genetic counselling is indicated for the family, as this X-linked condition can have variable penetrance of muscle weakness, myalgia and DCM in female carriers (Cho & Choi, 2013). Thus, familial studies can be offered to screen female relatives, including the mother and sisters, who could then access cardiac screening. Finally, variant identification by this study supports effective reproductive counselling regarding family planning, both for NM03 and NM06 and family members (Kim et al., 2026).

Overall, the findings in NM03 and NM06 demonstrate the value of LR-WGS for detecting complex CNVs and resolving precise breakpoints within repetitive and intronic regions, which are challenging to interrogate using conventional SRS-based approaches. Future improvements in the interpretation of *DMD* CNVs would benefit from larger LRS reference datasets, robust case-control studies of structural variation, and more detailed reporting of intragenic *DMD* deletions and associated phenotypes in databases such as ClinVar.

4.1.2. LR-WGS identified a novel *ORAI* variant in a recently identified NMD gene (NM21)

Clinical Description: NM21 was a male who died at age 76. His phenotype was characterized by adult-onset progressive muscle weakness beginning at age 40, with lower limb and proximal predominance. By age 61, he reported easy fatigability, muscle cramping, episodic myalgia with dark brown urine, and neuropathic pain affecting the lower legs. Weakness progressed over subsequent years, resulting in frequent falls and use of a scooter and wheelchair for ambulation, and he also developed significant elbow flexor, left hand and axial weakness. Additional manifestations that NM21 developed included decreased lung capacity, tinnitus and left-sided hearing loss. Lastly, NM21 exhibited congenital, chronically constricted pupils (miosis) which did not respond to dilation. Clinical investigations performed between ages 61 and 72 years of age supported a myopathic process, including persistently elevated serum CK, myopathic findings on EMG, fibrofatty replacement on muscle MRI and muscle biopsy abnormalities consisting of fatty replacement of fibres, endomysial fibrosis, fibre size variability, increased internal nuclei and lobulated fibres (**Table 4.5**). Notably, the muscle biopsy also demonstrated tubular aggregates. These structures, which represent densely packed membrane tubules derived from the sarcoplasmic reticulum, are a pathological feature characteristic of tubular aggregate myopathy (TAM) (Serano et al., 2026). Standardized phenotypic (HPO) terms to represent NM21's neuromuscular features and chronic miosis are listed in **Table 4.4**.

Table 4.4. Human Phenotype Ontology (HPO) terms selected for NM21 based on their clinical presentation.¹

HPO Term Description	HPO Code
Muscle fiber tubular inclusions	HP:0100301

Increased variability in muscle fiber diameter	HP:0003557
Fatty replacement of skeletal muscle	HP:0012548
Distal muscle weakness	HP:0002460
Proximal muscle weakness	HP:0003701
Limb-girdle muscle weakness	HP:0003325
Axial muscle weakness	HP:0003327
EMG: myopathic abnormalities	HP:0003458
Exercise-induced leg cramps	HP:0008991
Difficulty climbing stairs	HP:0003551
Difficulty standing	HP:0003698
Frequent falls	HP:0002359
Increased muscle fatiguability	HP:0003750
Unsteady gait	HP:0002317
Elevated circulating creatine kinase concentration	HP:0003236
Elevated circulating myoglobin concentration	HP:0033438
Dark urine	HP:0040319
Exercise-induced myoglobinuria	HP:0009305
Recurrent myoglobinuria	HP:0003652
Miosis	HP:0000616
Myalgia	HP:0003326

¹The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3. Terms ordered from first to last in the table according to the order used in the HPO hierarchy.

Table 4.5. Summary of neuromuscular investigations in NM21.

Investigation	Finding	Age ¹ (Years)	Additional Information
Serum CK	Persistently elevated	68-72	638-985 U/L (reference: <175U/L)
Serum myoglobin	Elevated	68-72	213-278 ug/L (reference: <50 ug/L)
Muscle MRI	Selective fibrofatty replacement	61	
EMG	Myopathic	61 and 67	
Muscle biopsy²	Severe myopathic with tubular aggregates	63	Myopathic features: fatty replacement of fibres, endomysial fibrosis, fibre size variability, increased number of internal nuclei, lobulated fibres, tubular aggregates.
NCS	Normal	61 and 67	

¹Age in years. ²Sample site: vastus. CK, creatine kinase; EMG, electromyography; MRI, magnetic resonance imaging; NCS, nerve conduction studies.

Family History: NM21's parents were non-consanguineous individuals of European (English and Scottish) descent. He had one sister, who died at 46 years old of cervical cancer, and no biological children. NM21, his sister, and his mother all had scoliosis. NM21's nephew, approximately 66 years old, is affected with an inflammatory myopathy.

Prior Genetic Testing: NM21 underwent two forms of genetic testing, both of which were non-diagnostic. Single-gene sequencing of *CAPN3* to assess for calpainopathy-limb girdle muscular dystrophies was negative. Subsequently, a 43-gene neuromuscular sequencing panel with CNV detection was performed in 2019. This testing was non-diagnostic but did note four heterozygous VUS. The first variant was a heterozygous *DOK7* VUS with no second pathogenic allele identified, thereby conferring at most carrier status for AR congenital myasthenic syndrome (OMIM #54300). Secondly, a potential splice-altering VUS in *DES* lacked sufficient evidence for pathogenicity (bioinformatic splicing predictions unavailable, not reported in affected or healthy individuals) and was not felt to correlate with the participant's phenotype. Lastly, two VUS were

identified in *TTN*, pathogenic variants in which are associated with a broad spectrum of AD and AR myopathies (OMIM #188840). Firstly, a nonsense VUS was not considered likely to cause a neuromuscular phenotype because it is located in exon 48 of the Novex-3 transcript, which is not expressed in principal cardiac or skeletal muscle isoforms (Roberts et al., 2015). Secondly, a missense variant was classified as a VUS at the time of reporting due to its low allele frequency in gnomAD, absence from reported affected individuals in the literature, and conflicting *in silico* predictions. Given the limited evidence of pathogenicity, a lack of clear phenotypic correlation with the condition, and high background genetic variation in *TTN*, the participant's phenotype could not be attributed to this variant. Since the original report, the variant has been submitted to ClinVar as Uncertain Significance and Likely Benign by five and eight submitters, respectively.

LR-WGS Variant Filtration: NMD gene-based variant filtration yielded **nine variants** for further assessment, including five SNVs, three CNVs and one SV. Of these, six variants were not considered of interest due to a lack of phenotypic correlation, including two variants in genes associated with AR conditions for which no second pathogenic allele was identified. The previously reported *DES* variant was also retained by the filtering pipeline and again interpreted as not causal for the participant's phenotype. The eighth variant, a 160 kb CNV on chromosome 12, could not be verified in IGV and is therefore considered a sequencing artifact. The last variant, ***ORAI1* c.417G>A, p.(M139I)** (GRCh38 chr12:121641154:G:A, novel variant with no rsID available) was prioritized in our bioinformatic pipeline according to absence from reference populations ("Rare0") and high predicted deleteriousness ("Missense-High", Revel ≥ 0.45). The *ORAI1* variant was prioritized during manual review of candidate variants due to clinical correlation with AD tubular aggregate myopathy 2 (OMIM #615883) (**Table 4.6 and 4.7**) (**Figure 4.4**). Both *TTN* variants from past clinical testing were present in the LR-WGS data; however, they

were not retained in the filtering process due to high general population allele frequency (gnomAD maximum allele frequency $\geq 0.1\%$) and low bioinformatic deleteriousness predictions.

Lastly, NM21's blood sample failed TRGT genotyping at the *RFC1* locus, where biallelic repeat expansions are associated with cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS). Failure to genotype the *RFC1* repeat region has been reported in individuals with biallelic pathogenic expansions, as highly repetitive sequences can reduce sequencing coverage and result in allelic dropout (Höps et al., 2025). NM21's phenotype is not known to be consistent with the clinical spectrum of CANVAS, however, NM21 has never been formally assessed by a clinician for any CANVAS features, such as limb and gait ataxia, vestibular reflexes or sensory neuropathy. Further, NM21's predominant phenotype was characterized by longstanding myopathy and wheelchair dependence; given that CANVAS can be late-onset and slowly progressive, it is possible for CANVAS manifestations to have been overlooked or masked by the individual's primary NMD. Thus, an *RFC1* expansion cannot be excluded, particularly given the possibility of a dual diagnosis in which a milder or clinically distinct CANVAS phenotype may have been masked by the individual's primary neuromuscular disorder. This finding illustrates the range of results that may arise from LR-WGS and underscores the importance of detailed phenotyping for accurate variant interpretation.

Table 4.6. Allele frequency, pathogenicity evidence and phenotype information for heterozygous *ORAI1* variant in NM21. GRCh38 chr12:121641154:G:A, rsID not available.

Variant	Zygosity	GA4K/gnomAD ² (%)	<i>In Silico</i> predictions	ClinVar pathogenicity	ACMG pathogenicity	OMIM phenotype (Inheritance)
c.417G>A p.M139I	Het ¹	0/0	Pathogenic (5) Uncertain (1)	NA	Uncertain ³	Tubular aggregate myopathy 2 (AD)

¹Slivar analysis was negative for compound heterozygosity; no second pathogenic allele was identified in *ORAI1*; ²gnomAD v4.1.1., using maximum frequency in any population; ³Applicable criterion: PM2. ACMG, American College of Medical Genetics; AD, autosomal dominant; GA4K, genome answers for kids; GRCh38, Genome Research Consortium human build 38; Het, heterozygous; OMIM, Online Mendelian Inheritance in Man.

Table 4.7. *In silico* prediction tool values for variant *ORAI1* c.417G.A, p.(Met139Ile) (chr12:121641154:G:A) in NM21.

Tool	Value	Prediction	Tool interpretation
MutationTaster	1	Pathogenic	Range 0-1, higher values more likely to be deleterious
AlphaMissense	0.999	Pathogenic	Strong > 0.984 Moderate >0.869 Supporting >0.761
PolyPhen-2	0.192	Uncertain/low	Range 0-1, higher values more likely to be deleterious
<u>Aggregate impact score metrics</u>			
REVEL	0.53	Pathogenic (supporting)	Strong >0.932 Moderate >0.773 Supporting >0.29
CADD PHRED	24.40	Pathogenic	Higher scores more likely to be deleterious (>20 predicted among 1.0% most deleterious in human genome)

CADD, Combined Annotation-Dependent Depletion; REVEL, rare exome variant ensemble learner.

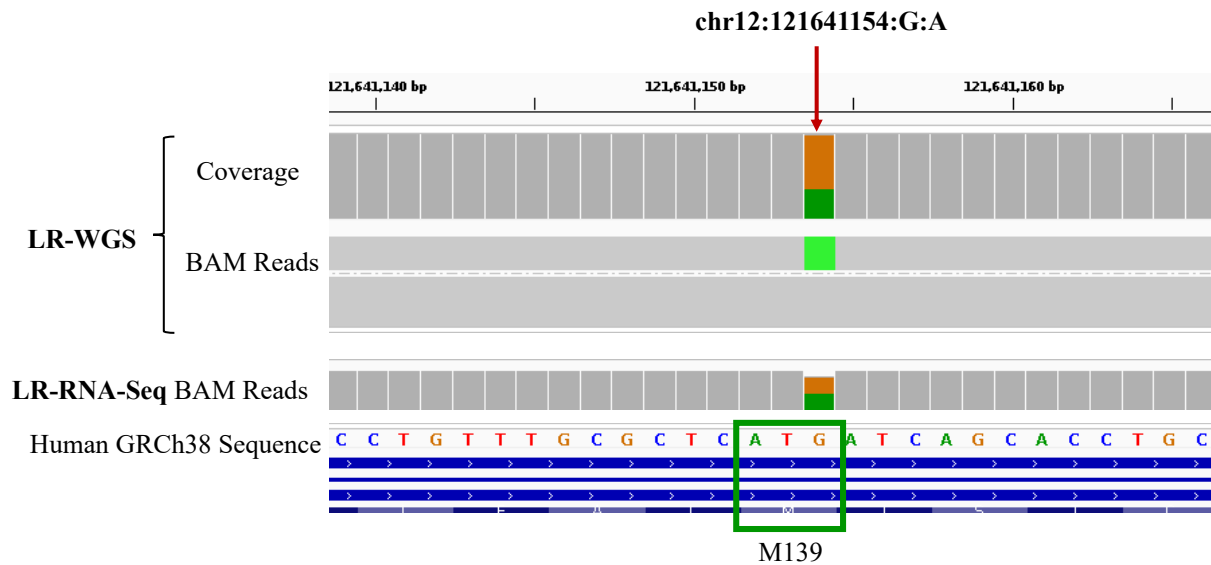


Figure 4.4. IGV snapshot of NM21's long-read genome and RNA sequencing read alignments for NM21 demonstrate heterozygous missense *ORAI1* variant (chr12:121641154:G:A). Variant nomenclature according to GRCh38, no variant rsID available. In LR-WGS and LR-RNA the wild-type allele *ORAI1* c.417G is indicated in orange as indicated by the red arrow, while the variant allele c.417A is indicated in green. The impacted residue (M139) is indicated by the green box. chr, chromosome; GRCh38, Genome Reference Consortium Human Build 38; IGV, Integrative Genomics Viewer; LR-RNA-seq, long-read RNA sequencing; LR-WGS, long-read whole genome sequencing.

ORAI1: *ORAI1*, located on chromosome 12q24, encodes the ORAI1 protein (Feske et al., 2006). ORAI1 is a component of the calcium-release activated calcium (CRAC) channel at the plasma membrane, and is activated by STIM1 to enable store-operated calcium ion entry (Yeung et al., 2020). Calcium ion signalling is essential for skeletal muscle contraction and for regulated cellular processes such as transcription, proliferation, differentiation, and exocytosis (Morin et al.,

2020). *ORAI1* is ubiquitously expressed in human tissues, including skeletal muscle, showing only minimal expression in the brain (GTEx Consortium, 2020; McCarl et al., 2009).

Gain-of-function missense variants in *ORAI1* are recognized to cause AD tubular aggregate myopathy 2 with congenital miosis (OMIM #615883) (Endo et al., 2015; Garibaldi et al., 2017; Nesin et al., 2014). In addition to characteristic tubular aggregates on muscle biopsy, elevated serum CK and congenital pupillary miosis, affected individuals have adult-onset, slowly progressive diffuse muscle weakness, with lower- and proximal-muscle predominance (Endo et al., 2015; Garibaldi et al., 2017; Nesin et al., 2014). *ORAI1* variants have been suggested to cause hyperactive CRAC signalling; sustained calcium influx through CRAC is hypothesized to gradually disrupt protein folding within the sarcoplasmic reticulum, thus initiating the formation of tubular aggregates (Nesin et al., 2014). Indeed, *in vitro* studies of *ORAI1* variants identified in TAM-affected individuals demonstrated constitutive activation of CRAC and increased Ca^{2+} entry, thereby supporting a gain-of-function disease mechanism (Endo et al., 2015; Garibaldi et al., 2017). Further, many individuals with heterozygous missense variants in the aforementioned *STIM1* also present with TAM and demonstrate excessive extracellular Ca^{2+} entry (Morin et al., 2014, 2020). Given the close interaction of *ORAI1* and *STIM1*, and their shared role in CRAC signalling, this finding supports dysregulated calcium homeostasis as a shared pathogenic mechanism underlying the muscular phenotype (Serano et al., 2026). Given the relatively recent association of *ORAI1* with NMDs, this gene was not included in earlier neuromuscular gene panels, including the one NM21 received in 2019.

ORAI1 c.417G>A, p.(Met139Ile): This variant was considered a high-interest variant according to ACMG criteria because it is absent from the gnomAD population database (**PM2**) and as multiple *in-silico* prediction tools support a deleterious effect (PP3). MutationTaster (1.00),

AlphaMissense (0.99), REVEL (0.53) and CADD PHRED (24.40) all predicted a deleterious effect, while PolyPhen-2 produced an uncertain or lower score (0.19) (**Table 4.7**). Polyphen-2 is an older tool which predicts the functional impact of missense variants using evolutionary conservation, protein sequence and structural features (Adzhubei et al., 2013). In contrast, REVEL integrates a broader range of genomic and functional annotations from 13 missense algorithms, including PolyPhen-2, SIFT and Mutation Taster, and was trained to distinguish rare pathogenic from neutral missense variants (Ioannidis et al., 2016). CADD uses a model trained on a much larger and less biased set of conserved and simulated *de novo* variants, and employs 60 annotations spanning conservation, epigenetic and functional features; the PHRED score of 24.40 places this variant among the top 1% of small variants predicted to be deleterious genome wide (Rentzsch et al., 2019). Thus, *in silico* evidence from more recently developed, comprehensive tools supported a deleterious effect of this variant. However, PM2 and PP3 were insufficient to establish pathogenicity to establish pathogenicity in the absence of additional supporting evidence such as functional or segregation data. As such, this variant is classified as a VUS.

From a structural perspective, the p.(M139I) substitution in this genetic variant replaces one hydrophobic, uncharged amino acid with another of similar volume (Pommié et al., 2004). However, the replaced isoleucine is more hydrophobic than methionine (hydropathy indices 4.5 vs. 1.8), and is also slightly larger than methionine (166.7 vs 162.9Å³) (Kyte & Doolittle, 1982; Pommié et al., 2004). *ORAI1* c. 417G>A, p.(Met139Ile) impacts the second-to-last amino acid in the second *ORAI1* transmembrane, which extends from residues 120-140 (Cancer Genome Atlas Research Network, 2014; Frischauf et al., 2017a). As such, the identified variant is consistent with known disease mechanisms for TAM2, as all TAM-associated *ORAI1* variants have been reported in the transmembrane domains of the protein (Morin et al., 2020). Of note, a missense variant at

the amino acid directly adjacent to Met139 (L138F) has been reported in ClinVar as Likely Pathogenic and has been shown to cause adolescent-onset slowly progressive myopathy through constitutive Orai1 activation (Endo et al., 2015). Further, *in vitro* studies demonstrated that residue 138 interacts with neighbouring residues, including T92 and H134, to stabilize and mediate rotation of the pore helix during important channel functions (Yeung et al., 2023). This same group also demonstrated that missense variant L138F disrupts normal biochemical function and regulation of the channel through steric clash with neighbouring amino acids, thus demonstrating the importance of the volume of these nearby residues (Yeung et al., 2023). While the role or variation of M139 was not described in this study, the residue's proximity to this key interactive site in ORAI1 highlights the need to consider the effects of amino acid substitution. Indeed, a different *ORAI1* missense variant at this same residue (M139V) was found in a stomach carcinoma sample; overexpression of *ORAI1* M139V in HEK cells led to significantly increased Ca^{2+} concentrations (Frischauf et al., 2017). Altogether, *ORAI1* c.417G>A, p.M139I was consistent with known disease mechanism for TAM2 and is located in a functionally important transmembrane domain, which has shown sensitivity to amino acid substitution. Overall, the variant's rarity in healthy reference populations, supportive *in silico* predictions and functional evidence from the same and neighbouring residues within a key transmembrane domain suggested that *ORAI1* c.417G>A may have a pathogenic effect despite VUS classification by ACMG guidelines.

Clinical Correlation and Conclusion: After review by our multidisciplinary committee, this *ORAI1* variant was considered a strong candidate underlying NM21's phenotype. There was strong clinical correlation with the participant's phenotype; notably, the tubular aggregates on muscle biopsy and the participant's chronic miosis are both rare features in NMDs. Further, there was

concordance between both variant type (missense) and residue locus (transmembrane domain, proximity to L139) and the established disease mechanism. Lastly, most bioinformatic tools consistently predicted a deleterious effect. Given the variant's uncertain pathogenicity status, further evidence is needed to support reclassification towards Likely Pathogenic or Pathogenic. Such information could include functional studies in model organisms or the identification of additional affected individuals through resources such as GeneMatcher. Segregation analysis is not possible in this case due to the paucity of living relatives. Lastly, as the variant was identified in a research setting, validation in a clinically accredited laboratory is required. The identification of this variant with relatively recent gene-disease associations, as well as recently emerged functional information of nearby amino acid substitution, highlights the utility of performing comprehensive genetic analysis, such as LR-WGS, as new evidence continues to emerge.

4.1.3. LR-WGS facilitated reclassification of FLNC missense variant through expanded case-level evidence from additional affected individuals (NM09)

Clinical Description: NM09 is a 60-year-old female with progressive proximal lower limb weakness, notably of the hip girdle muscles, with onset at age 42. Disease manifestations also included compromised balance and buckling of the left knee, numbness and tingling in her thighs and numbness of her fingers and toes. Physical examination by a clinical geneticist at age 49 demonstrated absent knee and ankle reflexes and reduced sensation to touch in her lower back, upper thighs and anterior legs. She was unable to toe-walk or heel-walk and demonstrated overriding of the great toes bilaterally. She displayed a modified positive Gowers sign, with widened stance and knee instability indicating hip stabilizer weakness. Over subsequent years, NM09 described increasing aching pain in the lower back and feet. Additional medical history for NM09 included adolescent thoracic scoliosis, and ligamentous knee injury at 42 years old requiring meniscal debridement. Lastly, an echocardiogram at age 52 demonstrated borderline left ventricular hypertrophy and mild valvular abnormalities, including mild aortic stenosis, mitral and tricuspid regurgitation, and trace pulmonary insufficiency. Standardized phenotypic (HPO) terms to represent NM09's neuromuscular and cardiac features are listed in **Table 4.8**. Clinical testing performed between 48-50 years of age demonstrated **myopathic features**, including mildly elevated serum CK, myopathic EMG and fatty replacement of multiple muscle groups on MRI (**Table 4.9**). Further, muscle biopsy of the quadriceps femoris revealed chronic myopathic changes, including lobulated fibres and nemaline rods, increased fibre-size variation, grouped atrophy, and mild endomysial fibrosis.

Table 4.8. Human Phenotype Ontology (HPO) terms selected for NM09 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

¹The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3. Terms ordered from first to last in the table according to the order used in the HPO hierarchy.

Table 4.9. Summary of neuromuscular investigations in NM09.

Investigation	Finding	Age ¹	Additional Information
Serum CK	Mildly elevated	49	121 IU/mL (reference: 28-110 IU/mL)
Muscle MRI	Significant fibrofatty replacement	50	Severe bilateral atrophy in thigh and calf muscles, significant fatty replacement in gastrocnemius and soleus, moderate replacement of peroneal muscles, core of fibrofatty atrophy in gluteus medius and patchy fatty replacement of biceps femoris and adductor magnus.
EMG	Myopathic	49	
Muscle biopsy²	Myopathic with fatty replacement	48	Myopathic features: lobulated fibres and nemaline rods, increased fibre size variation, grouped atrophy and mild endomysial fibrosis. Fatty replacement with areas of streaking hyper-erythema.
NCS	Normal	61 and 67	

¹Age in years. ²Sample site: quadriceps femoris. CK, creatine kinase; EMG, electromyography; MRI, magnetic resonance imaging; NCS, nerve conduction studies.

Family History: NM09's parents are non-consanguineous individuals European (Irish and Czech) descent. NM09 has a maternal first cousin once removed who uses a wheelchair for ambulation, for unknown reasons. Of NM09's large sibship, she has one full sister with a mildly elevated CK level (134U/mL; reference: 28-110IU/mL), but no neuromuscular symptoms. NM09's other siblings, three full sisters and three full brothers, are all reported to be healthy. NM09 has no children.

Prior Genetic Testing: NM09 underwent two previous forms of genetic testing. Firstly, a three-gene nemaline myopathy sequencing panel (*ACTA1*, *NEB* and *TPM2*) was negative for pathogenic variants. Second, a 43-gene myopathy panel was non-diagnostic, revealing two heterozygous VUS in *NEB* and *FLNC*. The *NEB* finding consisted of a copy number gain within the triplicate region. While most humans have six total copies of this genetic material, NM09 was found to have eight (Donner et al., 2004). Pathogenicity depends on allelic distribution; if the two additional copies

are on a single allele (i.e. 3,5) rather than one additional copy on each allele (i.e. 4,4), this variant could be pathogenic. Biallelic pathogenic variants, including triplicate-region copy number gain, in *NEB* cause AR nemaline myopathy 2 (NEM2) (OMIM #256030) (Kiiski et al., 2016). As no second *NEB* variant was detected in NM09, the triplicate-region copy number gain was thought to cause carrier status at most for NM09, depending on the organization of the eight copies. Of note, the highly homologous sequence of the *NEB* TRI cannot be captured effectively with SRS. However, LRS, in conjunction with new bioinformatic tools, have facilitated improved detection and phasing of copy number information for homologous genes, such as the Paraphrase tool for *NEB* and *SMN1/2* (Chen et al., 2023; PacBio, 2022/2026; Sagath et al., 2025).

The second VUS found through the above testing was a heterozygous, rare missense *FLNC* variant located in exon two of 48 coding exons: **c.577G>A, p.(A193T)** (GRCh38 chr7:128835550:G:A, rs387906587). At the time of reporting (2016), this variant was classified as a VUS based on **PM2** (absence from control populations). In 2011, Duff et. al. reported that this variant segregated with disease in an Italian family but did not publish a pathogenicity classification for the variant. The group reported four individuals with a 30-year-old onset of hand weakness and slow running, with proximal and distal leg weakness beginning after age 40, and proximal upper limb weakness beginning after age 50 (Duff et al., 2011). Serum CK was normal to two-fold increased, and two of the four individuals developed cardiomyopathy. Finally, functional studies of the p.A193T filamin demonstrated slightly decreased thermal stability, increased actin-binding activity and the formation of intracellular protein aggregates (Duff et al., 2011). At the time of reporting, pathogenic variants in *FLNC* were reported to cause various AD conditions, including myofibrillar myopathy, distal myopathy 4 and hypertrophic cardiomyopathy 26 (Duff et al., 2011; Shatunov et al., 2009; Valdés-Mas et al., 2014; Vorgerd et al., 2005), while

association with several other AD cardiomyopathies would become available afterwards (Begay et al., 2016; Brodehl et al., 2016; Brun et al., 2020) (OMIM #102565). Familial testing of NM09's asymptomatic full siblings was negative for this variant, and NM09's muscle biopsy findings were consistent with a myofibrillar myopathy. In sum, although available information only allowed VUS classification according to ACMG guidelines at the time, this *FLNC* variant was considered highly likely to be underlying NM09's phenotype, due to high phenotypic correlation and available functional variant information.

LR-WGS Variant Filtration: NMD gene-based variant filtration yielded **four variants** for further assessment, including three small variants and one SV. The SV and two of the small variants were not considered relevant due to a lack of phenotypic concordance. The last variant was the same *FLNC* variant identified in prior genetic testing (**Table 4.10**) (**Figure 4.5**), which was prioritized in our bioinformatic filtering pipeline according to absence from reference populations ("Rare0") and high predicted deleteriousness ("Missense-High", REVEL ≥ 0.45). The *FLNC* was prioritized as a variant of interest during manual review of candidate variants based on clinical correlation with myofibrillar myopathy 5 (OMIM #609524) and recently reported variant pathogenicity information.

Table 4.10. Allele frequency, pathogenicity evidence and phenotype information for heterozygous *FLNC* missense variant in NM09. GRCh38 chr7:128835550:G:A, rs387906587.

Variant	Zygosity	GA4K/gnomAD ² (%)	<i>In silico</i> predictions	ClinVar pathogenicity	ACMG pathogenicity	OMIM phenotype (Inheritance)
c.577G>A p.A193T	Het ¹	0/0	Pathogenic (6)	Pathogenic	Pathogenic ³	Distal myopathy 4 (AD) Myofibrillar myopathy 5 (AD)

Hypertrophic CM 26 (AD)
Restrictive CM 5 (AD)
Dilated CM 1PP (AD)
ARVD (AD)

¹Slivar analysis for compound heterozygosity negative; no second pathogenic allele in *FLNC*; ²gnomAD v4.1.1., using maximum frequency in any population; ³Applicable criterion: PM2, PM6, PP3, PP5, PS3, PS4. ACMG, American College of Medical Genetics; AD, autosomal dominant; ARVD, arrhythmogenic right ventricular dysplasia; CM, cardiomyopathy; GA4K, genome answers for kids; Het, heterozygous; OMIM, Online Mendelian Inheritance in Man.

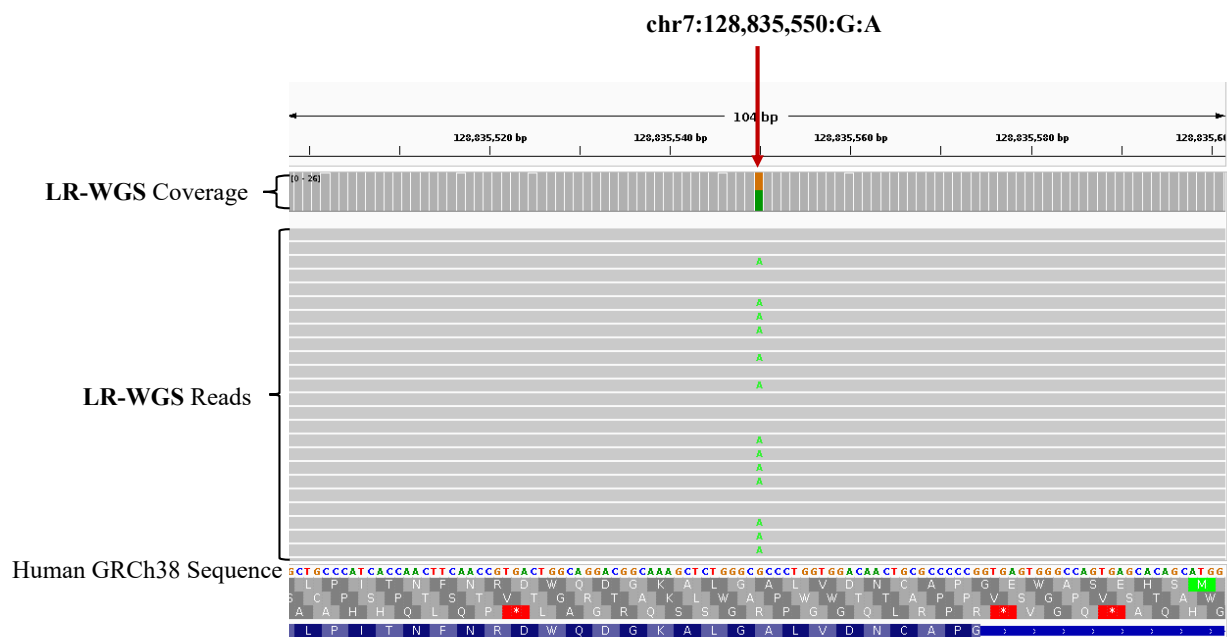


Figure 4.5. IGV snapshot of LR-WGS read alignments demonstrate a heterozygous missense *FLNC* variant in NM09 (chr7:128835550:G:A). Variant nomenclature according to GRCh38; variant rsID: rs387906587. In LR-WGS the wild-type allele chr7:128835550:G is indicated in orange as indicated by the red arrow, while the variant allele chr7:128835550:A is indicated in green. chr, chromosome; GRCh38, Genome Reference Consortium Human Build 38; LR-RNA-seq, long-read RNA sequencing; LR-WGS, long-read whole genome sequencing.

FLNC Variant Reclassification: With integration of the following evidence and ACMG criteria,

the variant was upgraded from a predicted VUS to Pathogenic. Criterion **PM2** (absence from

controls) and **PS3** (well-established *in vitro* functional studies supportive of a damaging effect) still applied to the classification, as with the previous classification. Further, since its last report in 2016, the variant has been reported as pathogenic five times (2018-2025) by reputable laboratories, allowing the application of **PP5 criteria**. The variant has been reported multiple times in unrelated individuals, allowing application of criteria **PS4** (prevalence in affected individuals is significantly increased compared with controls). Bioinformatic *in silico* tools allowed application of **PP3** through a highly deleterious predicted variant effect (**Table 4.11**) and intolerance of the *FLNC* gene to loss-of-function variants (pLI = 1; o/e LoF = 0.25; Decipher Haploinsufficiency = 21.56). Lastly, we applied criterion **PM6** (assumed *de novo* without confirmation of paternity/maternity). Although NM09's parents were not available for familial variant testing, the absence of the variant in all seven of NM09's siblings provided supporting evidence for an assumed *de novo* occurrence.

Table 4.11. *In silico* prediction tool values for variant *FLNC* c.577G>A (p. A193T) in NM09.
GRCh38 chr7:128835550G:A, rs387906587.

Tool	Value	Prediction	Tool interpretation
AlphaMissense	0.80	Pathogenic (supporting)	Strong > 0.984 Moderate >0.869 Supporting >0.761
MutationTaster	1	Pathogenic	Range 0-1, higher values more likely to be deleterious
MutationAssessor	3.67	Deleterious (high)	High deleterious >3.5 Medium deleterious >1.935 Low deleterious >0.8
SIFT	0.00	Deleterious	Supporting pathogenic evidence <0.001
PolyPhen-2 (HumVar)	0.99	Deleterious	Range 0-1, higher values more likely to be deleterious

<u>Aggregate impact score metrics</u>			
CADD PHRED	26.00	Pathogenic	Higher scores more likely to be deleterious (>20 predicted among 1.0% most deleterious in human genome)
REVEL	0.91	Pathogenic (moderate)	Strong >0.932 Moderate >0.773 Supporting >0.29

CADD, Combined Annotation-Dependent Depletion; SIFT, Sorting intolerant from tolerant; REVEL, rare exome variant ensemble learner.

Clinical Correlation and Conclusions: LR-WGS enabled the identification and reinterpretation of a known *FLNC* missense variant, the pathogenicity of which was last assessed in 2016. Since that time, new variant evidence, stronger reference datasets and more bioinformatic tools to predict pathogenicity of small variants have emerged, allowing reclassification from VUS to Pathogenic (II) based on ACMG criterion **PM2 and PS3**, plus more recent application of criterion **PM6, PP3, PP5 and PS4**. NM09's progressive proximal lower limb muscle weakness, myopathic features, fatty atrophy and reduced sensation and reflexes in the lower limbs, with onset at age 42, aligned with reported disease presentation of similar symptoms between the third to fifth decade of life for individuals with *FLNC* variants (Previtali et al., 2019; van den Bogaart et al., 2017). Further, NM09's late-onset cardiac abnormalities aligned with hypertrophic CM reported in several individuals with *FLNC* variants and co-occurring myopathies (Duff et al., 2011). Although this sequencing was research-grade, the variant has already been validated in prior clinical genetic testing and thus genetic counselling is indicated to communicate an official diagnosis of myofibrillar myopathy 5. Lastly, as multiple individuals with the same genetic variant have progressed to respiratory muscle involvement, health counselling and surveillance for this feature is indicated for this participant (van den Bogaart et al., 2017).

In summary, this case illustrates the utility of iterative variant reanalysis in genetic testing to apply new variant- and disease-level information to previously unsolved cases. The case also demonstrates the value of comprehensive genetic testing to identify actionable health information beyond the given indication, as the identification of the *FLNC* variant in NM09 allowed screening and detection of hypertrophic CM. Lastly, the case demonstrates the importance of functional studies in elucidating variant pathogenicity and enabling genetic diagnosis. While a strong variant-disease association has been made for *FLNC* variants, the underlying disease pathology remains to be determined. In the single functional study, abnormal protein accumulations or actin-binding dysregulation were postulated to be involved in the disease mechanism (Duff et al., 2011); however, muscle biopsies from affected individuals, including NM09, are not reported to show protein accumulations. Further functional studies, whether *in vitro/in vivo* or using transcriptional/proteomic data from participant muscle biopsies, are needed to understand the disease mechanism underlying this variant and potentially identify any therapeutic targets.

4.1.4. LR-RNA demonstrated splice-altering effect of a previously identified TTN predicted splicing variant (NM12)

Clinical Description: NM12 is a 13-year-old female with a phenotype characterized by congenital-onset neuromuscular features and facial dysmorphism; these features are summarized through standardized phenotypic (HPO) terms in **Table 4.12**. NM12 was born at 36 weeks' gestation (i.e. late pre-term infant) following a slow delivery with some fetal decelerations. She had a birth weight of 2956g (50-90th percentile), a length of 53cm (50th percentile), and a head circumference of 34cm (>98th percentile). At birth, she demonstrated severe peripheral hypotonia and upper limb contractures. Physical examination by a clinical geneticist at three days old revealed facial dysmorphism with a high forehead, frontal bossing, high-arched eyebrows, high-arched palate, microretrognathia, bilateral infraorbital creases, thin upper lip vermillion and thin-simple auricles. Further, NM12 displayed bilateral ptosis, as well as dolichocephaly with a prominent occiput and metopic ridge. She required a three-week hospitalization due to feeding difficulties, fatigue and hyperbilirubinemia.

By age two, NM12 continued to exhibit mild arthrogryposis, feeding challenges and minimal finger joint contractures, as well as gross motor delay. On the most recent genetics follow-up at age seven, NM12 demonstrated diffuse, non-progressive muscle weakness, particularly in the knee and hip flexors and shoulder abductors, as well as poor muscle bulk, decreased deep tendon reflexes, a waddling gait and a positive Gowers sign. She was hyperflexible, with swan-necking of the distal joints, and had long fingers and thoracic scoliosis. She continues to struggle with gross motor skills, including standing from a seated position, engaging in physical activity with peers, and balancing on one foot. She demonstrates normal cognition and completes a regular school program.

Clinical testing performed between infancy and 3 years of age demonstrated **myopathic features (Table 4.13)**. Firstly, muscle MRI in infancy and at age three revealed symmetric fibrofatty replacement and paucity of muscle tissue in several areas (thighs, quadriceps and soleus), with compensatory hypertrophy of the adductor muscles. Muscle biopsy of the left quadriceps in infancy also demonstrated myopathic features (myofiber size variation and oxidative enzyme deficit), as well as multiple foci of contractile element disorganization, therefore suggestive of multi-minicore myopathy.

Table 4.12. Human Phenotype Ontology (HPO) terms selected for NM12 based on their clinical presentation.¹

HPO Term Description	HPO Code
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:0002194
Abnormal finger morphology	HP:0001167

Waddling gait	HP:0002515
Reduced tendon reflexes	HP:0001167
Dolicocephaly	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 vermillion	HP:0000219
High palate	HP:0000218
Simple ear	HP:0020206

¹The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3. Terms ordered from first to last in the table according to the order used in the HPO hierarchy.

Table 4.13. Summary of neuromuscular investigations in NM12.

Investigation	Finding	Age ¹	Additional Information
Serum CK	Normal	0-2	
Muscle MRI	Fibrofatty replacement and atrophy	0 and 3	
EMG	Normal	1	
Muscle biopsy²	Myopathic with multiple minicores	0	Myopathic features: myofiber size variation, oxidative enzyme deficit Multiple foci of contractile element disorganization suggestive of multiminicore myopathy
NCS	Normal	1	

¹Age in years. ²Sample site: left quadriceps femoris. CK, creatine kinase; EMG, electromyography; MRI, magnetic resonance imaging; NCS, nerve conduction studies.

Family History: NM12 is of Middle Eastern (Syrian and Palestinian) descent; her parents are first cousins (i.e. consanguineous). She has one full brother (age 16). No family members are known to be affected by neuromuscular disorders.

Prior Genetic Testing: NM12 underwent seven forms of prior genetic testing (**Table 4.14**), all of which were non-diagnostic, with one VUS of interest noted. Single-gene analyses for *SMN1*, *DMPK*, *CHST3*, and *FBN2*, as well as karyotype analysis, were all normal. Chromosomal microarray was normal, although multiple regions of homozygosity were noted totalling 153 Mb. Lastly, singleton, SRS-based exome sequencing was performed in another research study; this identified multiple VUS that were not considered of interest due to a lack of phenotypic correlation with the participant. A homozygous *TTN* missense variant was not considered clinically relevant, as it was located in an exon expressed only in cardiac transcripts, not in the N2AS soleus/skeletal long isoform transcript in skeletal muscle; thus, the variant was less likely to cause a congenital neuromuscular phenotype. However, one homozygous *TTN* VUS (c.34291+3A>G, GRCh38 c.34291+3A) was considered of interest due to predicted splicing impact through computational analysis. Further, the *TTN* variant was in close proximity (3 bp) to the splice junction of exon 138, which is incorporated into the N2A skeletal long isoform and expressed in skeletal muscle tissue. The research team commented that additional studies were necessary to determine the true functional impact of this splicing variant, notably through RNA-seq.

Table 4.14. Summary of genetic testing performed for NM12. Tests are ordered chronologically.

Test	Result	Additional findings
Karyotype	Normal, XX	
Chromosomal microarray	Non-diagnostic	Multiple ROH (>5Mb each, totalling 153 Mb)
<i>SMN1</i> gene dosage	Normal	

<i>DMPK</i> CTG repeat analysis	Normal	
<i>CHST3</i> sequencing and deletion/duplication analysis	Normal	
<i>FBN2</i> sequencing and deletion/duplication analysis	Normal	
WES (SRS, singleton)	Non-diagnostic	• Homozygous <i>TTN</i> predicted splicing variant • Homozygous <i>TTN</i> missense variant • Three heterozygous VUS in genes was associated with AD, AD/AR and AR genetic conditions (AD)*

*No phenotypic correlation between associated disease and NM12. AD, autosomal dominant; AR, autosomal recessive; ROH, runs of homozygosity; SRS, short-read sequencing; VUS, variant of uncertain significance.

LR-WGS Variant Filtration: NMD gene-based variant filtration yielded **12 variants** for further assessment, including nine small variants, one CNV and two SVs. Of these, five of the small variants had previously been detected in clinical exome sequencing, including the *TTN* missense and splicing variants. Of the remaining six variants, four were not prioritized due to a lack of phenotypic correlation, and two due to multiple reports of benign pathogenicity in ClinVar. Ultimately, the *TTN* homozygous splicing variant (**c.34291+3A>G, chr2:178677618:T:C**) was prioritized as a candidate variant of interest based on strong phenotypic correlation, fulfillment of the AR inheritance pattern and expression of the relevant exon in skeletal muscle tissue (**Tables 4.15 and 4.16**). Pathogenic *TTN* variants have been reported in several NMD phenotypes; however, the AR congenital and myofibrillar myopathies may match the phenotype of this participant (OMIM #611705, 603689) (Amberger et al., 2015). The *TTN* variant was prioritized in our bioinformatic filtering pipeline due to absence from reference populations (“Rare0”) and predicted splicing impact (“Splicing-High, dbSNV ADA score > 0.6).

Table 4.15. Allele frequency, pathogenicity evidence and phenotype information for homozygous *TTN* predicted splicing variant in NM12. GRCh38 chr2:178677618T>C.

Variant	Zygosity	GA4K/gnomAD ¹ (%)	<i>In silico</i> predictions	ClinVar pathogenicity	ACMG pathogenicity	OMIM phenotype (Inheritance)
c.34291+3A>G	Homozygous	0/0	Splice-altering (4)	NA	Uncertain ²	1. LGMD R10 titin-related (AR) 2. Myofibrillar myopathy 9 with early respiratory failure (AD) 3. Tardive tibial muscular dystrophy (AD) 4. Congenital myopathy 5 with CM (AR) 5. CM dilated 1G and hypertrophic 9 (AD)

¹gnomAD v4.1.1., using maximum frequency in any population; ²Applicable criterion: PM2, PP3. ACMG, American College of Medical Genetics; AD, autosomal dominant; GA4K, genome answers for kids; GRCh38, Genome Research Consortium human build 38; Het, heterozygous; OMIM, Online Mendelian Inheritance in Man.

Table 4.16. *In silico* prediction tool values for homozygous *TTN* predicted splicing variant in NM12. *TTN* c.34291+3A>G, GRCh38 chr2:178677618T>C.

Tool	Value	Prediction	Tool interpretation
dbscSNV RF	0.914	Likely splice-altering	Scores range 0 to 1 and are interpreted as probability of variant being splice altering
dbscSNV ADA	0.9987	Likely splice-altering	Scores range 0 to 1 and are interpreted as probability of variant being splice altering

SpliceAI	DS Acceptor gain: 0 DS Acceptor loss: 0 DS Donor gain: 0.47 DS Donor loss: 0.48 DP Acceptor gain: 3 DP Acceptor loss: 47 DP Donor gain: 22 DP Donor loss: 3	Predicted splicing high - No predicted acceptor gain/loss - Moderate predicted donor gain 22bp from variant - Moderate predicted donor loss 3bp from variant	Scores range 0 to 1; values > 0.2 suggest likely splice-altering effects
<u>Aggregate impact score metrics</u>			
CADD PHRED	23.5	Pathogenic	Higher scores more likely to be deleterious (>20 predicted among 1.0% most deleterious in human genome)

ADA, AdaBoost; AG, acceptor gain; CADD, Combined Annotation-Dependent Depletion; dbSNV, database of single nucleotide variants within splicing consensus regions; DP, Delta Position; DS, Delta Score; REVEL, rare exome variant ensemble learner; RF, Random Forest.

TTN: c.34291+3A>G (chr2:178677618T>C): This intronic variant was visually confirmed using IGV (**Figure 4.6**). The variant is located 3 bp upstream of a canonical splice site within intron 147. According to ACMG classification, the variant was classified as a “**VUS**” based on absence from control datasets (gnomAD) (**PM2**) and multiple lines of computational evidence supporting a deleterious effect (**PP3**). *In silico* tools predicted a potential impact of the variant on normal splicing (dbscSNV RF ADA = 0.9987; dbscSNV RF = 0.914, SpliceAI maximum delta score = 0.48), with SpliceAI predicting loss of a canonical donor splice site 3 bp away from the variant (DS DL = 0.48) and potential gain of a donor splice site 22 bp away from the variant (DS DG = 0.47). CADD PHRED aggregate impact score further supported deleteriousness through a score of 23.5.

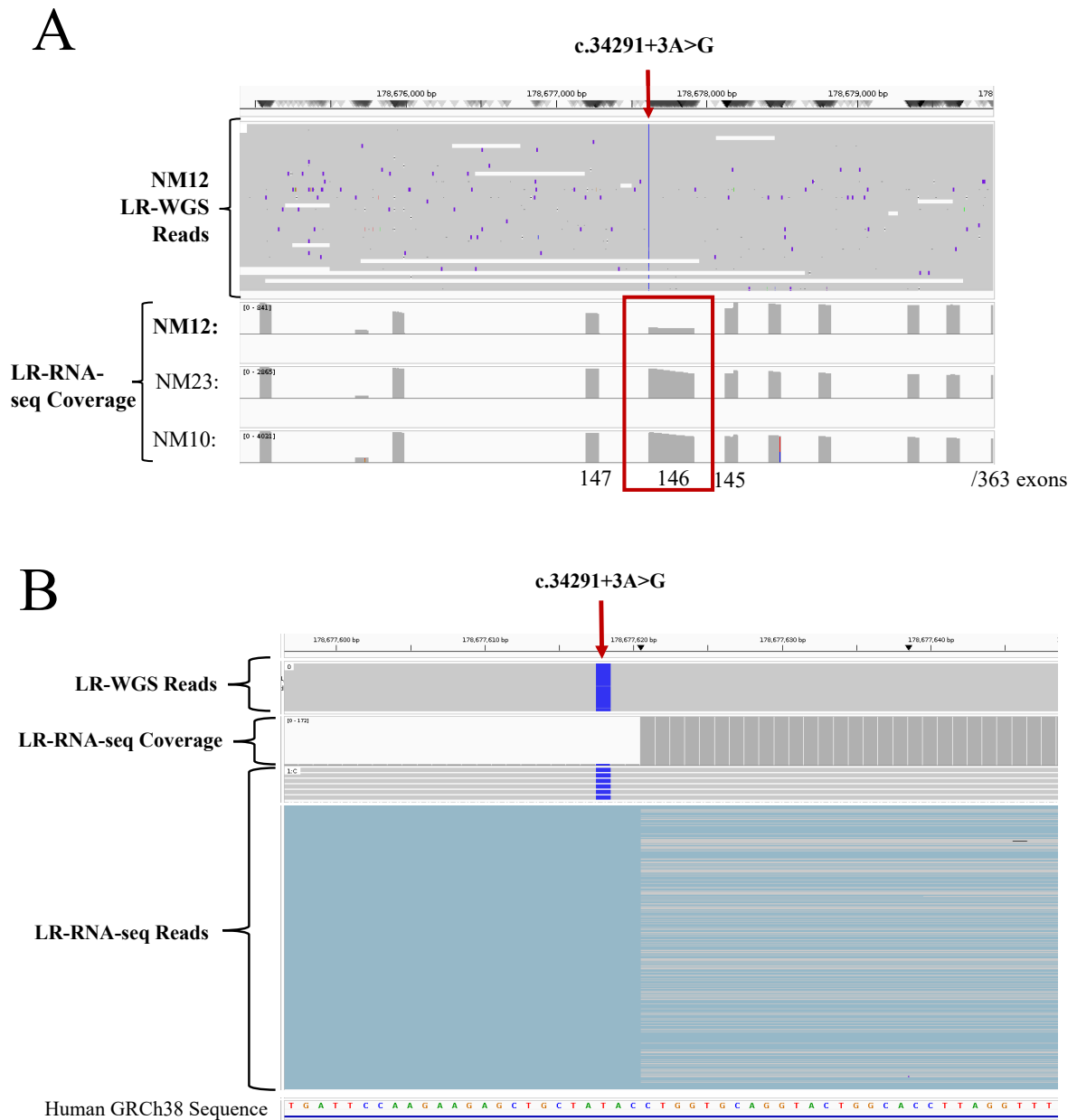


Figure 4.6. IGV snapshots of LR-WGS and LR-RNA-seq alignments in *TTN* demonstrate demonstrate homozygous splicing variant (chr2:178677618:T:C) with alterations to expression of exon 146 compared to other individuals in NMD cohort. Variant nomenclature according to GRCh38. The variant (c.34291+3A>G) is 3 base pairs downstream from the exon 146 donor splice site; exon 146 is indicated in the red box in image A. A) Image of LR-WGS reads in NM12, as well as LR-RNA-seq coverage for NM12 in the first row (identified in bold), while the next two rows display LR-RNA-seq coverage in two other individuals in the NMD cohort who do not have

identified *TTN* pathogenic variants. B) Image of LR-WGS reads, as well as LR-RNA_seq reads for NM12 which demonstrate five reads in support of the missense variant. The remaining LR-RNA-seq reads in at this locus are shown in the bottom portion of LR-RNA-seq Reads, with a blue background and reads shown in grey. chr, chromosome; GRCh38, Genome Reference Consortium Human Build 38; IGV, Integrative Genomic Viewer; LR-RNA-seq, long-read RNA sequencing; LR-WGS, long-read whole genome sequencing.

LR-RNA-seq provided additional evidence supporting a functional impact of this *TTN* variant. Firstly, *TTN* was identified by IsoRanker analysis as a gene expression outlier in NM12 compared to the cohort, demonstrating significant downregulation (**Figure 4.7 B**). *TTN* expression in NM12 was measured at 1.28 TPM compared with a cohort mean of 9.68 TPM, representing a 7.6X decrease in expression. This corresponded to a Z-score of -2.08 and a test statistic of 5.89 by IsoRanker analysis. Reduced expression was also evident on visual assessment of LR-RNA-seq reads in IGV, with *TTN* supported by an average of 841 reads in NM12 compared with 2,865-4,031 reads in other cohort members. Additionally, *TTN* was identified in IsoRanker analyses as a splicing outlier, with upregulation of known and novel splice sites. IsoRanker prioritized 16 isoforms for *TTN* in NM12, four of which were upregulated (Z-score approximately 4.01) and 12 which were downregulated ($[-2.40 - 1.01]$) in comparison to other individuals in this cohort (**Figure 4.7 A**). A qualitative review of LR-RNA-seq data in IGV showed markedly reduced coverage of exon 146, the exon immediately preceding this intronic variant, compared with adjacent exons and with corresponding coverage in other individuals from the NMD cohort (**Figure 4.6**).

Clinical correlation and conclusion: After review by our multidisciplinary committee, the homozygous *TTN* splicing variant is considered a strong candidate for explaining NM12's phenotype. While the variant was previously classified as a VUS based on computational splicing

predictions, LR-RNA-seq from disease-relevant tissue supported the interpretation of a true splice-altering impact in exon 147, as well as overall gene down-regulation. Formal evaluation of these transcriptomic findings by a certified clinical molecular geneticist will be required to confirm their contribution towards variant classification. This case demonstrates the utility of the IsoRanker-guided LR-RNA-seq approach; even in the absence of prior variant identification through clinical exome sequencing or LR-WGS, the observed aberrant *TTN* splicing would have prioritized the gene for further genomic investigation. Moreover, this case demonstrates the contribution of functional, disease-relevant information to determine variant impact.

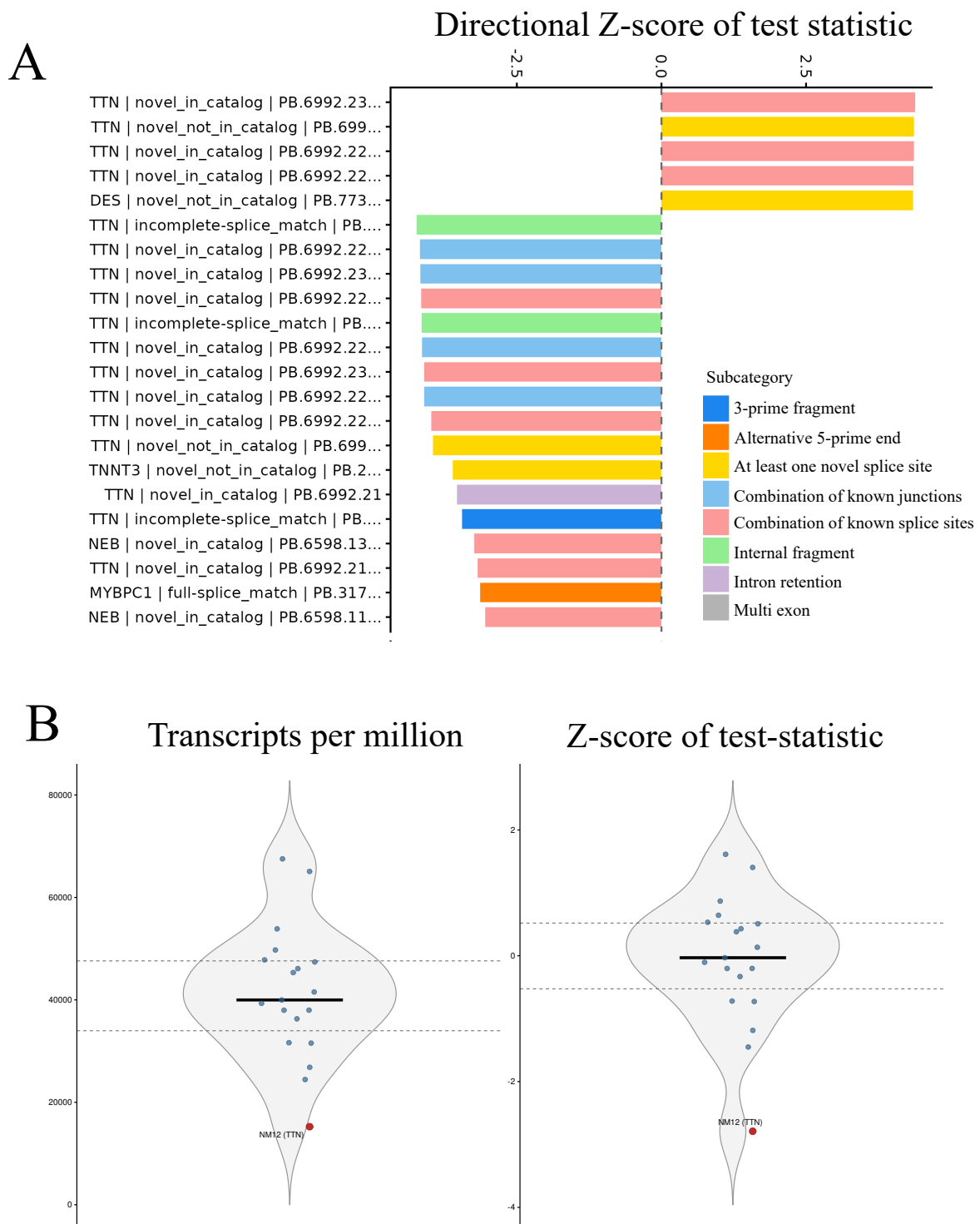


Figure 4.7. LR-RNA-seq with IsoRanker identifies aberrant *TTN* splicing and decreased *TTN* gene expression in NM12 compared with other individuals in NMD cohort ($n = 19$). A) Isoform expression outliers in NM12 using a directional Z score $\geq [2.5]$. Bar colors indicate isoform subcategory according to PacBio Iso-Seq annotation

(Guizard et al., 2023). B) Gene-level expression outliers in *TTN*; *TTN* showed reduced gene expression in NM12 (1.28 TPM compared to the rest of the cohort (mean 9.68 TPM), with a directional Z-score of -2.08, test statistic of 5.89 and z-score of test statistic of 3.93. LR-RNA-seq; long-read RNA sequencing; NMD, neuromuscular disease; PacBio Iso-Seq, Pacific Biosciences Isoform Sequencing; TPM, transcripts per million.

4.1.5. Complementary LR-WGS and LR-RNA-seq approach allowed identification of biallelic NUP188 variants with accurate direct phasing (NM05)

Clinical Description: NM05 was a male who died at 4 years and 2 months old. In the neonatal period NM05 demonstrated poor respiratory effort and the requirement of tracheotomy and ventilation due to muscle weakness and central obstructive apnea. NM05 was assessed by clinical genetics at that time, who noted multiple congenital anomalies (cleft palate, rocker bottom feet, anterior placed anus, abnormal thumbs, long fingers and toes and undescended testes), microcephaly, facial dysmorphisms (broad/flat nose, microcephaly and small anterior fontanelle), cardiac defect (large patent ductus arteriosus, small atrial septal defect, mild tricuspid regurgitation), neuromuscular weakness, infantile spasms, mixed severe bilateral hearing loss and bilateral cataracts. Later in development, NM05 demonstrated severe global developmental delay, muscular hypotonia, respiratory failure, failure to thrive, encephalopathy and seizures. Standardized phenotypic (HPO) terms to represent NM05's neuromuscular, respiratory and congenital features are listed in **Table 4.17**. Regarding clinical investigations, muscle biopsy at three months of age showed reduced alpha-dystroglycan immunoreactivity and was otherwise normal. Records of EMG, NCS and MRI were not available for review, as they were ordered by and performed at a centre outside of the province.

Table 4.17. Human Phenotype Ontology (HPO) terms selected for NM05 based on their clinical presentation.¹

HPO Term Description	HPO Code
Small anterior fontanelle	HP:0000237
Primary microcephaly	HP:0011451

Respiratory insufficiency due to muscle weakness	HP:0002747
Generalized muscle weakness	HP:0003324
Generalized neonatal hypotonia	HP:0008935
Abnormal thumb morphology	HP:0001172
Long fingers	HP:0100807
Long toe	HP:001051
Rocker bottom foot	HP:0001838
Neonatal electro-clinical tonic seizure	HP:0032821
Global developmental delay	HP:0001263
Congenital encephalopathy	HP:0007239
Neurogenic bladder	HP:0000011
Bilateral cryptorchidism	HP:0008689
Patent ductus arteriosus after premature birth	HP:0011649
Atrial septal defect	HP:0001631
Congenital sensorineural hearing impairment	HP:0008527
Cleft palate	HP:0000175
Juvenile cataract	HP:0001118
Mixed hearing impairment	HP:0000410
Short umbilical cord	HP:0001196

¹The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3. Terms ordered from first to last in the table according to the order used in the HPO hierarchy.

Family History: NM05's parents were non-consanguineous individuals of European (Irish, French, English, Eastern European) and Métis ancestry. NM05 had no siblings, and no other family

members were affected with neuromuscular symptoms. NM05's paternal twin uncles and paternal grandfather had early-onset cataracts, with diagnosis in the third and fifth decades, respectively.

Prior Genetic Testing: NM05 underwent seven forms of prior genetic testing. Of these, five genetic tests returned normal results: karyotype, chromosomal microarray, Fanconi anemia testing, cerebro-oculofascioskeletal syndrome testing, and sequencing of nuclear ($n=19$) and mitochondrial ($n=37$) genes. A neuromuscular gene sequencing panel ($n=129$) revealed no diagnostic pathogenic changes, with two heterozygous VUS identified in *COL12A1* and *KLHL40*; both were considered non-diagnostic due to a lack of phenotypic correlation and the absence of a second variant in *COL12A1*. Two VUS of unknown phasing were also observed in *NEB*; pathogenic variants in *NEB* cause nemaline myopathy, typically following AR inheritance, though AD inheritance can also occur.

Lastly, a research-based trio exome using SRS was also non-diagnostic, although several heterozygous VUS were revealed. Specifically, three heterozygous variants were in genes associated with AR conditions, with no second variant found. A paternally inherited pathogenic variant in *VWF* was identified, thereby providing a secondary finding of Von Willebrand disease with reduced penetrance. Finally, a rare missense VUS was identified in *UNC45B*: **c.712T>A, p.(S238T)** (GRCh38 chr17:35155368:T:A, rs766448432) (**Table 4.18**). *UNC45B* has been proposed as a candidate gene for AD cataract 45 (OMIM #616279), based on linkage studies in one family and functional studies supporting a role for *UNC45B* in lens maturation in zebrafish (Hansen et al., 2014). At the time of the initial exome analysis, the variant was considered a potential explanation for NM05's personal and paternal family history of cataracts, although there were significant differences in age of onset between the family reported by Hansen et al. (2014) (early childhood), NM05 himself (congenital) and NM05's paternal relatives (20-50 years-old).

Further, variant reanalysis during the present study classified the variant as a VUS based on the sole applicable criterion **BP4**: Computational prediction tools supported a benign effect on the gene (**Table 4.19**). Overall limited evidence supported both gene-disease association and pathogenicity of this variant, and the genetic finding did not explain NM05's broader syndromic and neuromuscular phenotype. Thus, the variant was not considered a sufficient explanation for his presentation by the present study, and another genetic explanation was sought.

Table 4.18. Allele frequency, pathogenicity evidence and phenotype information for heterozygous *UNC45B* variant in NM05. GRCh38 chr17:35155368:T:A, rs766448432.

Variant	Zygo- sity	GA4K/ gnomAD ² (%)	<i>In silico</i> predictions	ClinVar pathogen- icity	ACMG pathogen- icity	OMIM phenotype (Inheritance)
c.712T>A p.S238T	Het ¹	0/0.003	Benign (7) Uncertain (1) Deleterious (1)	Uncertain	Uncertain ³	?Cataract 43 (AD)

¹Slivar analysis for compound heterozygosity negative; no second pathogenic allele in *UNC45B*.
²gnomAD v4.1.1., using maximum frequency in any population. ³Applicable criterion: BP4. ACMG, American College of Medical Genetics; AD, autosomal dominant; chr, chromosome; GA4K, genome answers for kids; GRCh38, Genome Reference Consortium Human Build 38; Het, heterozygous; OMIM, Online Mendelian Inheritance in Man.

Table 4.19. *In silico* prediction tool values for variant *UNC45B* c.712T>A, p.(S238T) in NM05. GRCh38 chr17:35155368:T:A, rs766448432.

Tool	Value	Prediction	Tool interpretation
AlphaMissense	0.136	Benign (Moderate)	Strong benign < 0.073 Moderate benign <0.147 Supporting benign <0.331
MutationAssessor	1.56	Low likelihood of functional impact	Probability of functional impact: High > 3.5 Medium > 1.935

			Low > 0.8 Neutral < 0.8
MutationTaster	0.994	Deleterious	Range 0-1, higher values more likely to be deleterious
PolyPhen-2 (HumVar)	0.437	Benign	Range 0-1, higher values more likely to be deleterious
SIFT	0.313	Benign (Supporting)	Pathogenic <0.001 Moderate benign >0.327 Supporting benign >0.08
<u>Aggregate impact score metrics</u>			
CADD PHRED	19.32	Uncertain	Higher scores more likely to be deleterious (>20 predicted among 1.0% most deleterious in human genome)
MPC	0.094	Benign	Deleterious ≥ 2.0
REVEL	0.05	Benign (Moderate)	Strong benign <0.016 Moderate benign < 0.183 Supporting benign <0.29
CADD, Combined Annotation-Dependent Depletion; MPC, Missense badness, PolyPhen-2, and Constraint; SIFT, Sorting intolerant from tolerant; REVEL, rare exome variant ensemble learner.			

LR-WGS Variant Filtration: NMD gene-based variant filtration yielded **two variants** for further assessment, including one SNV and one SV. These variants were not considered relevant to diagnosis due to a lack of association with the individual's phenotype. All VUS from prior genetic testing, as well as variants associated with carrier status, were also identified in the LR-WGS data but were filtered out before manual interpretation due to maximum allele frequency >0.1% in gnomAD (*KLHL40*, *NEB* c.5565C>G, *MFF*, *VWF*), low pathogenicity prediction (*CCDC174*, *COL12A1*, *NEB* c.23894C>A, *UNC45B*) or lack of gene association with an NMD (*BCHE*, *TYR*, *CYP26C1*).

LR-RNA-Guided Filtering: Analysis of LR-RNA-seq data with IsoRanker identified gene expression outliers in NM05 for one NMD-associated gene and 10 genes associated with other OMIM phenotypes. Four additional genes exhibited isoform expression outliers with NMD or OMIM phenotypes. Among these genes, *NUP188* showed decreased expression in NM05 [1.28 transcripts per million (TPM)] compared to the rest of the cohort (mean 8.90 TPM) (**Figure 4.8**). This represented a 6.9X reduction in *NUP188* gene expression, with an IsoRanker Z-score of -2.08 and test statistic of 3.93 (**Figure 4.8**). *NUP188* was prioritized as a gene of interest based on clinical correlation with AR Sandestig-Stefanova syndrome (SANDSTEF) (OMIM #618804).

NUP188: Located at 9q34.11, this gene encodes the 188kDa nuclear pore protein Nup188 (Miller et al., 2000). As a scaffold to the Nup93 subcomplex of the nuclear pore complex (NPC), Nup188 supports regulation of macromolecule traffic between the nucleus and cytoplasm (Beck & Hurt, 2017; Miller et al., 2000). Nup188 may also have roles in chromatin organization, transcriptional regulation and chromosome segregation (Itoh et al., 2013; Labade et al., 2016). *NUP188* is ubiquitously expressed in humans, including in skeletal muscle and brain tissues (GTEx Consortium, 2020; Lonsdale et al., 2013).

Biallelic loss-of-function variants in *NUP188* were first associated with AR SANDSTEF (OMIM #618804) in 2019, based on two unrelated individuals with distinct homozygous nonsense variants (Sandestig et al., 2020). Muir et al. (2020) published a further six individuals from four unrelated families. Functional studies further supported gene and variant-level disease association; *Nup188*-knockout flies exhibited aberrant dendrite tiling, as well as motor deficits and seizures (Muir et al., 2020). Further, patient-derived fibroblasts showed from two individuals with SANDSTEF showed reduced protein transport into the nucleus compared to controls (Muir et al., 2020). Clinically, SANDSTEF is characterized by microcephaly, congenital anomalies, hypotonia,

respiratory failure, developmental impairment and abnormal brain imaging (Muir et al., 2020; Sandestig et al., 2020). These features demonstrate strong clinical correlation with NM05's phenotype, although, as brain imaging was not performed for NM05, those findings could not be correlated.

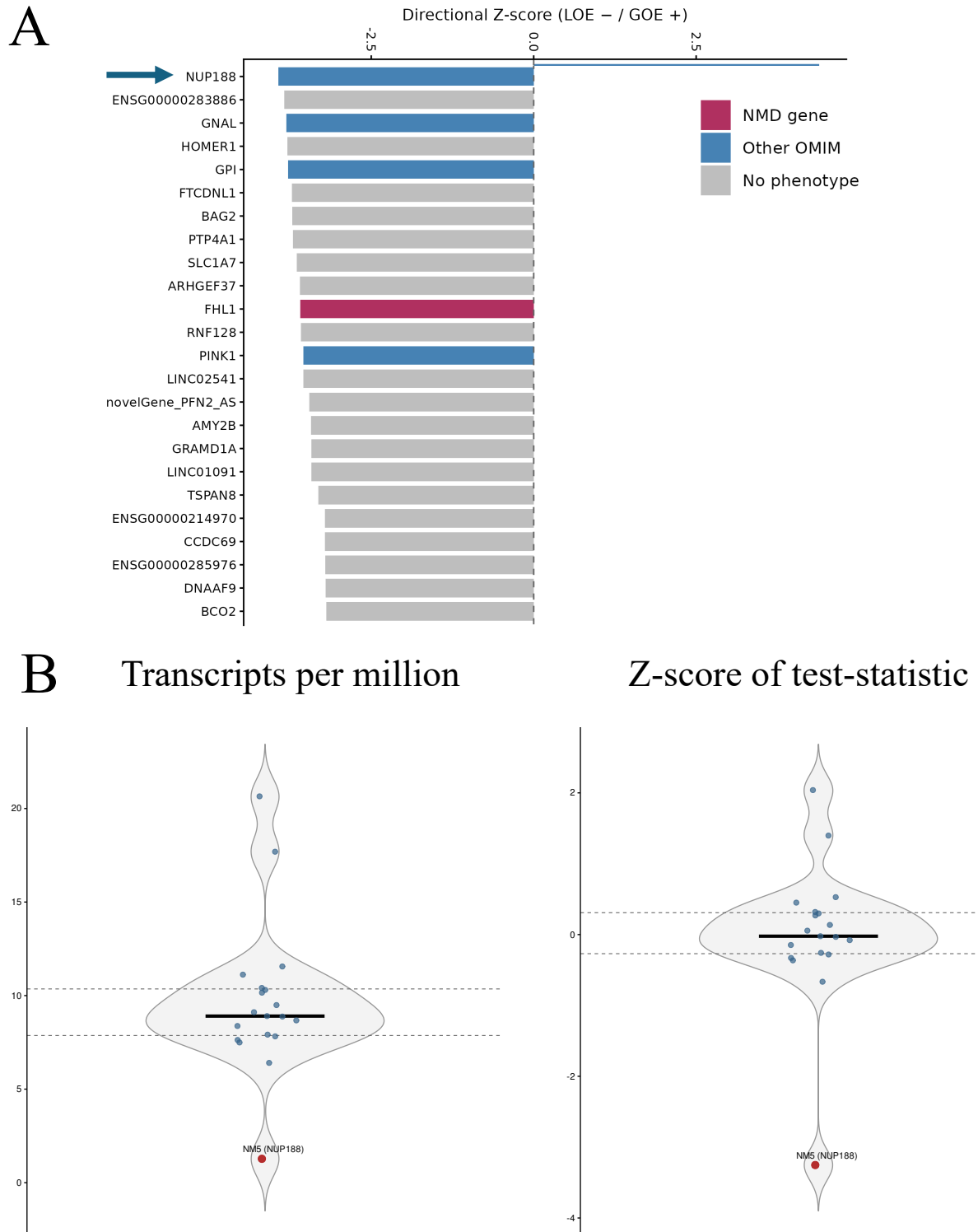


Figure 4.8. LR-RNA-seq with IsoRanker identifies reduced *NUP188* gene expression in NM05 compared with other individuals in NMD cohort ($n = 19$). A) Gene-level expression outliers in *NUP188*. Genes associated with NMDs are shown in purple, genes associated with other clinical phenotypes in OMIM in blue, and genes

without known clinical phenotype associations in grey. Negative directional Z-scores indicate decreased expression, while positive directional Z-scores indicate increased expression relative to cohort. B) *NUP188* gene level expression analyses for NM05; *NUP188* showed reduced expression in NM05 (1.28 transcripts per million [TPM]) compared to the rest of the cohort (mean 8.90 TPM), with a directional Z-score of -2.08 and test statistic of 3.93. LR-RNA-seq; long-read RNA sequencing; NMD, neuromuscular disease; OMIM, Online Mendelian Inheritance in Man; TPM, transcripts per million.

Filtering of LR-WGS for *NUP188* identified two compound heterozygous frameshift *NUP188* variants in *trans*. These variants were not identified in initial LR-WGS variant filtration, as *NUP188* is not specifically associated with NMDs but rather a multisystemic presentation including both muscle weakness and hypotonia, as well as other congenital features. Visualization in IGV confirmed the presence and *trans* configuration of the two variants (Figure 4.9).

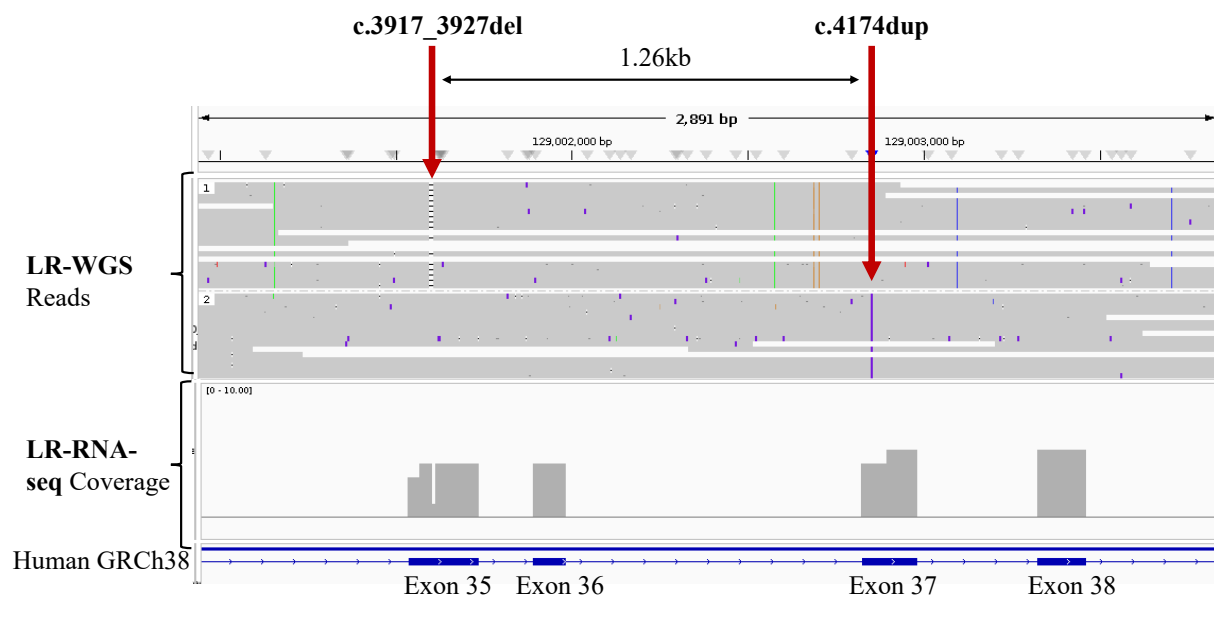


Figure 4.9. IGV snapshot of *NUP188* compound heterozygous variants in NM05 demonstrate ability of LR-WGS to accurately resolve variant phasing at lengths >1kb. Red arrows indicate compound heterozygous, likely pathogenic variants c.3917_3927del, p.(V1306Afs*25) and c.4174dup, p.(S1392F fs*27) on opposite alleles at a distance of 1.26kb, as demonstrated by the black double arrow. del, deletion;

GRCh38, Genome Reference Consortium Human Build 38; IGV, Integrative Genomics Viewer; LR-RNA-seq, long-read RNA sequencing; LR-WGS, long-read whole genome sequencing.

Likely Pathogenic *NUP188* Variants: Relevant information for both variants is listed in **Table 4.20** and includes location, allele frequency, pathogenicity evidence and gene-phenotype associations. The first *NUP188* is a frameshift deletion c.3917_3927del, p.(V1306Afs*25) in exon 35 (GRCh38 chr9:129001596:TGCAGGTAACCC:T). This variant was classified as Likely Pathogenic based on the following ACMG criterion: Null variant in a gene where loss of function is a known mechanism of disease (**PVS1**) and low allele frequency in healthy reference populations, with a maximum gnomAD frequency of 0.0003% in European individuals (**PM2**). The second variant is an insertion frameshift c.4174dup, p.(Ser1392Phefs*27) in exon 37 (GRCh38 chr9:129002852:C:CT). This variant was classified as Likely Pathogenic based on the same ACMG criteria (**PVS1** and **PM2**) and its low allele frequency in healthy reference populations, with a maximum gnomAD frequency of 0.0006% in European individuals. Neither variant has been reported in ClinVar, and computational pathogenicity prediction tools were not available for either variant. In reported cases of SANDSTEF, the variant mechanism is loss of function with homozygous or compound heterozygous truncating or frameshift variants identified (Muir et al., 2020). The downstream mechanism is not fully elucidated; however, loss-of-function *NUP188* variants are hypothesized to cause phenotypic changes through reduced nuclear transport, as well as disruption of left-right patterning and other cellular functions (Fakhro et al., 2011; Muir et al., 2020). There are no homozygous truncating variants in *NUP188* in gnomAD, and the gene does not show constraint metrics consistent with dosage sensitivity (pLI = 0; o/e LoF 0.51).

Table 4.20. Allele frequency, pathogenicity evidence and phenotype information for two compound-heterozygous *NUP188* variants in NM05.

Variant	Exon	GA4K (%)	gnomAD ¹ (%)	ClinVar pathogenicity	ACMG pathogenicity	OMIM phenotype (Inheritance)
c.3917_3927 deletion², p.(V1306A fs*25)	35/44	0	0.0003	NA	Likely pathogenic ³	Sandestig-Stefanova syndrome (AR)
c.4174 duplication⁴, p.(S1392F fs*27)	37/44	0	0.0006	NA	Likely pathogenic ³	Sandestig-Stefanova syndrome (AR)

In silico predictions were not available for these variants. These two compound heterozygous variants were on opposite alleles (in trans), 1.26kb apart. ¹gnomAD v4.1.1., using maximum frequency in any population. ²GRCh38 chr9:129001596:TGCAGGTAACCC:T. ³ACMG criterion applied: PVS1 and PM2. ⁴GRCh38 chr9:129002852:C:CT. AR, autosomal recessive; ACMG, American College of Medical Genetics; fs, frameshift; GA4K, genome answers for Kids; NA, not available.

Clinical Correlation and Conclusions: After review by our multidisciplinary committee, the compound heterozygous *NUP188* variants were considered high-interest variants likely to be causal for NM05's phenotype. In addition to elevated phenotypic correlation with NM05, the variants were consistent with both the AR inheritance pattern and the reported variant types for SANDSTEF. Importantly, the two variants were separated by more than 1 kb and therefore could not have been directly phased using conventional SRS approaches. Thus, confirming biallelic configuration and, ultimately, genetic diagnosis would instead have required additional studies, such as parental testing. In contrast, LR-WGS enabled direct phasing of the variants, illustrating the additional diagnostic value provided by this single long-read assay. Moreover, LR-RNA-seq facilitated variant identification and interpretation by providing evidence of reduced gene

expression consistent with a potential loss-of-function disease mechanism. Together, these findings support the utility of a multi-omic, LRS-based approach for genetic diagnosis. Clinical validation of the identified variants remains necessary before a formal diagnosis can be established.

4.2. Case Descriptions – Genomic findings of uncertain clinical significance

4.2.1. LR-WGS resolved a rare PRX structural variant of unknown significance (NM25)

Clinical Description: NM25 was a female who died at 14 years and 8 months of age. NM25's phenotype is summarized as complex multisystemic, with static muscle weakness and ptosis, respiratory insufficiency and multiple congenital anomalies. Standardized phenotypic terms to represent NM25's clinical features (HPO terms) are listed in **Table 4.21**. Given the complexity of NM25's phenotype, the following section will be divided into neonatal medical history, developmental history and neuromuscular history.

Table 4.21. Human Phenotype Ontology (HPO) terms selected for NM25 based on their clinical presentation.¹

HPO Term Description	HPO Code
Progressive congenital scoliosis	HP:0008458
Weakness of facial musculature	HP:0030319
Inability to puff cheeks	HP:6001102
Generalized amyotrophy	HP:0003700
Decreased muscle mass	HP:0003199
Generalized muscle weakness	HP:0003323
Generalized neonatal hypotonia	HP:0008935
EMG myopathic abnormalities	HP:0003458
Difficulty climbing stairs	HP:0003551
Difficulty descending stairs	HP:0033235

Difficulty running	HP:0009046
Small joint hypermobility	HP:0430046
Hypoplastic/small distal phalanx of the fifth finger	HP:0004227
Proximal placement of thumb	HP:0009623
Flexion contracture of finger	HP:0012785
Clubbing of fingers	HP:0100759
Pes planus	HP:0001763
Attention-deficit hyperactivity disorder	HP:00007018
Compulsive behaviors	HP:000722
Autism	HP:0000717
Cognitive impairment	HP:0100543
Impaired heel-walking ability	HP:6000490
Impaired toe-walking ability	HP:0034052
Clumsiness	HP:0002312
Slurred speech	HP:0001350
Patent ductus arteriosus	HP:001643
Abnormal inferior vena cava morphology	HP:0025576
Ventricular septal defect	HP:0001629
Atrial septal defect	HP:0001631
Myopathic facies	HP:0002058
Horizontal nystagmus	HP:0000666
Bilateral ptosis	HP:0001488
Midline liver	HP:0034188

Omphalocele	HP:0001539
Recurrent aspiration pneumonia	HP:0002100
Neonatal respiratory distress	HP:0002643
Ventilatory dependence with inability to wean	HP:0005946
Hypernasal speech	HP:0001611

¹The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3. Terms ordered from first to last in the table according to the order used in the HPO hierarchy.

Neonatal Medical History: NM25 was born at term with weight and length in the 25th percentile for infants (3062g, 50cm), while head circumference was in the 25-50th percentile (34.5cm). NM25 demonstrated multiple congenital anomalies, including a large omphalocele (diameter: 6.5cm), midline liver, congenital heart defects (atrial septal defect, inferior retro-aortic defect, ventricular septal defect, patent ductus arteriosus and duplicated inferior vena cava), mildly hypoplastic fifth fingernail and proximally implanted thumb. She also demonstrated congenital hypotonia, scoliosis and horizontal nystagmus, and was fed via a gastronomy tube. NM25 demonstrated respiratory insufficiency; she required immediate respiratory support after birth via intubation and ventilation. Tracheotomy was inserted at three months of age, removed at 14 months of age, and re-inserted at 14 years of age; in the years without tracheotomy, NM25 required continuous oxygen therapy via nasal prongs. NM25 was hospitalized for the first 16 months of life, during which time she experienced multiple desaturations, cardiac arrests and infantile seizures.

Developmental History: NM25 demonstrated global developmental delay (GDD) and cognitive impairment. She walked at 19 months and was never able to run, jump or climb stairs independently. She had a single word at two years, 20 words by three years, and was speaking simple sentences at 5.5 years. She was able to follow a two-step command after five

years. She attended elementary school with an individualized education plan and support educational assistants, and performed grade levels below her age. NM25 received diagnoses of autism spectrum disorder, attention deficit/hyperactivity disorder and obsessive-compulsive disorder.

Neuromuscular and Respiratory History: In early childhood, NM25 exhibited an abnormal, slightly wide-based gait and was described by family as clumsy and prone to frequent tripping. She exhibited static diffuse hypotonia and muscle weakness, as well as fluctuating slurred speech with a hyper-nasal quality, and dysconjugate gaze.

At age seven, NM25's respiratory and neuromuscular symptoms progressed; she was significantly more fatigued with activity and preferred use of her wheelchair to walking. She frequently underwent respiratory desaturation (blood oxygen saturation <90%), required high oxygen flow rates, and exhibited more prominent snoring, diaphoresis and finger clubbing. At the time, it was unclear whether her symptoms were of respiratory, cardiovascular, neuromuscular or other origin, as relevant investigations were normal (echocardiogram, diaphragmatic exclusion on fluoroscopy, restrictive lung process on sleep study), and as cardiac surgery to repair congenital defects (age nine) did not improve her phenotype. At age seven, she also demonstrated behavioral changes (e.g., more disruptive and easily frustrated, shorter attention span). By age eight, NM25 demonstrated a very myopathic facial appearance with bilateral ptosis and difficulty puffing out her cheeks. She exhibited hyperflexibility in her small joints, notably at the ankles, had a planovalgus deformity on weight-bearing and demonstrated isolated finger flexion contractures. Musculature was diffusely atrophic and was globally weak, with distal weakness more significant than proximal. She could not tiptoe or heel walk. Over the remainder of her life, NM25 choked

more consistently on food and water, demonstrated more facial droopiness and underwent progression of scoliosis.

Family History: NM25's parents were unrelated individuals of First Nations (Cree) descent. NM25's mother had dysferlin-related LGMD R2 with an identified homozygous *DYSF* pathogenic variant; NM25's maternal aunt, grandmother and great-grandmother have reportedly similar symptoms to her mother. NM25 has three full sisters, all unaffected by neuromuscular symptoms. Her two younger sisters had delayed speech and possible cognitive delay or impairment.

Prior Genetic Testing: NM25 underwent seven forms of prior genetic testing (**Table 4.22**). *SMN1/2* gene dosage analysis and *DMPK* CTG repeat analysis were normal, and chromosomal microarray (CMA) was non-diagnostic but did identify five runs of homozygosity totalling 51.89 Mb. An SRS-based 46-gene structural myopathy panel identified inheritance of one pathogenic *DYSF* variant (associated with LGMD R2, OMIM #253601) inherited from her mother. This gene panel also identified a heterozygous *POMGNT1* VUS (associated with several AR muscular dystrophy-dystroglycanopathies, OMIM #606822) for which no second pathogenic allele was found on *POMGNT1* deletion-duplication analysis.

Notably, singleton SR-WES was non-diagnostic but offered several VUS. Firstly, the participant was homozygous for a *CACNA2D2* missense variant: c.3218C>T, p.(P1073L) (GRCh38 chr3:50364961:G:A, rs753600233). Biallelic pathogenic variants in *CACNA2D2* are associated with AR cerebellar atrophy with seizures and variable developmental delay (OMIM #61850). The only evidence to support pathogenicity for the variant was rarity in reference populations (current maximum gnomAD frequency: 0.005%), and the variant is now classified in ClinVar as a VUS (Variation ID: 656108, Accession: VCV000656108.7). This variant was located within one of the regions of homozygosity of NM25 noted upon earlier chromosomal microarray,

thus explaining its homozygous state. However, the phenotype for this condition did not align closely with NM25's, and the individual's lack of brain MRI findings at 14 days and subsequently at 2.5 years of age dissuaded from attributing the variant to the patient's phenotype. Other variants detected in this clinical exome included two rare *TTN* missense variants of unknown phasing; these were of low clinical suspicion, as most disease-causing variants in *TTN* are frameshift, and as VUS are common in *TTN*.

Table 4.22. Summary of genetic testing performed for NM25. Tests are ordered chronologically.

Test	Result	Additional findings
Karyotype	Normal, XX	
CMA	Non-diagnostic	Five ROH (>5Mb each, totalling 51.89Mb)
<i>SMN1</i> and <i>SMN2</i> gene dosage	Normal	
<i>DMPK</i> CTG repeat analysis	Normal	
46-gene structural myopathy panel (SRS)	Non-diagnostic	• Heterozygous for maternal <i>DYSF</i> variant: asymptomatic carrier status LGMBD R2 • Heterozygous <i>POMGNT1</i> VUS
<i>POMGNT1</i> deletion-duplication analysis	Normal	No second allele variant; ruled out <i>POMGNT1</i> AR muscular dystrophy-dystroglycanopathies
Exome sequencing (SRS, singleton)	Non-diagnostic	• Homozygous <i>CACNA2D2</i> missense VUS* • Two <i>TTN</i> missense VUS, unknown phasing • Heterozygous VUS in two genes associated with AD conditions*

*No phenotypic correlation between associated disease and NM25. AD, autosomal dominant; AR, autosomal recessive; CMA, chromosomal microarray; LGMD R2, limb-girdle muscular dystrophy R2; SRS, short-read sequencing; VUS, variant of uncertain significance.

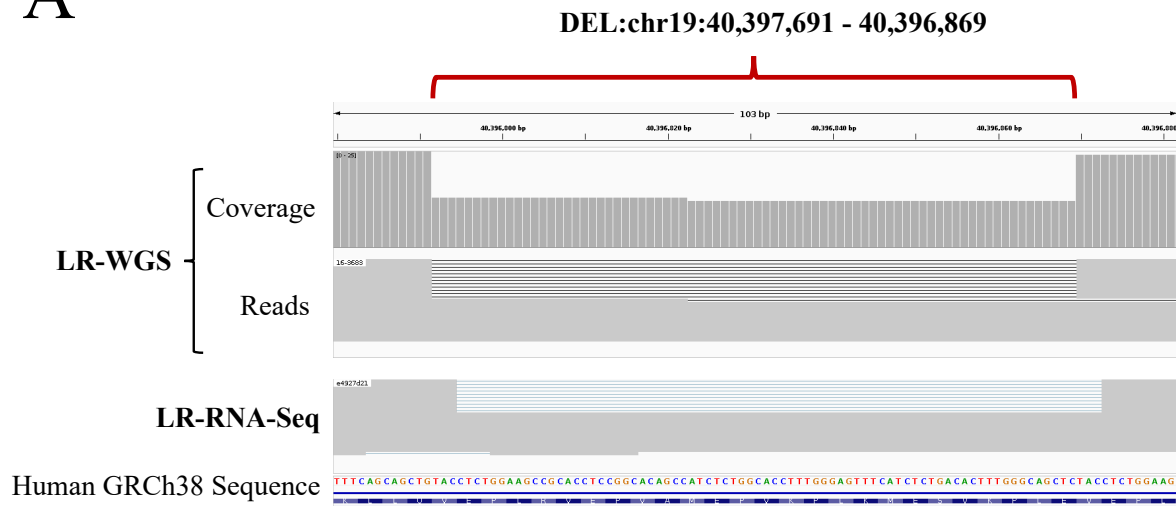
LR-WGS Variant Filtration: NMD gene-based variant filtration yielded **eight variants** for further assessment, including five small variants and three SVs. Of these, three small variants had been identified on previous clinical testing (*POMGNT1*, *DYSF* and *TTN*). No second biallelic *TTN* variant which met our filtering criteria was identified, and that variant had since been reported as Likely Benign in ClinVar. Of the remaining five variants, three were considered not relevant due to a lack of phenotypic correlation and one was not relevant due to gene association with an AR phenotype and the absence of a second variant. The last variant identified was an in-frame, 78 bp deletion (SV) identified in the seventh exon in *PRX* (DEL:Chr19: 40396791-40396869) (**Table 4.23**) (**Figure 4.10**). This SV was prioritized as a variant of interest based on the available literature on the variant and its potential clinical correlation with AD/AR Dejerine-Sottas syndrome (DSS).

Table 4.23. Pathogenicity and phenotype information for heterozygous *PRX* structural variant (loss) in NM25.

Gene Chr	Variant	Zyg- osity	Coding elements	ACMG pathogen- icity	ClinVar pathogen- icity	OMIM phenotype (Inheritance)
<i>PRX</i> chr19	78bp del: 40396791- 40396869	Het	Exon 7/7 Del: E495-V520	VUS (+0.3)	VUS (4)	1. CMT4F (AR) 2. Dejerine-Sottas syndrome (AD/AR)

ACMG, American College of Medical Genetics; AD, autosomal dominant; AR, autosomal recessive; Chr, chromosome; CMT4F, Charcot-Marie-Tooth Type 4F; del, deletion; Het, heterozygous; VUS, variant of uncertain significance.

A



B

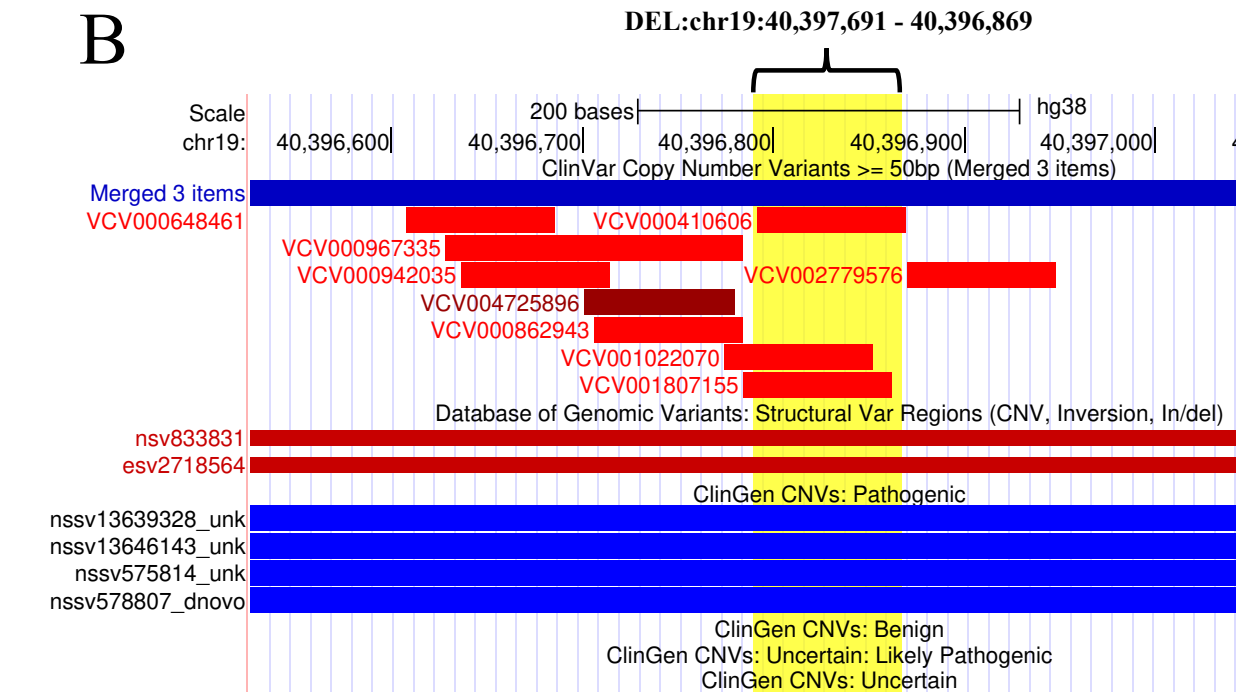


Figure 4.10. Complementary LR-WGS and LR-RNA-seq identify 78bp *PRX* deletion (chr19:40,397,691-40,396,869) in NM25 which shows significant overlap with other deletions of unknown significance reported in ClinVar. The deletion is located at GRCh38 chr19:40397691-40396869. A) IGV snapshot of LR-WGS and LR-RNA-seq alignments for deletion, which is indicated by red bracket. B) UCSC Genome Browser snapshot of deletion overlapping genomic regions with three deletions of uncertain significance in ClinVar; The range of NM25's deletion is indicated in yellow. Located underneath "ClinVar Copy Number Variants ≥ 50 bp...", CN gains are in blue,

CN losses of uncertain significance in red and CN losses of pathogenic classification in dark red. DEL, deletion; chr, chromosome; CNVs, copy number variants; CN, copy number; GRCh38, Genome Research Consortium human build 38; IGV, Integrative Genomics Viewer.

PRX: *PRX* maps to chromosome 19q13.13-q13.2 and encodes periaxin (Gillespie et al., 1994). *PRX* is expressed in myelinating Schwann cells in the peripheral nervous system (Gillespie et al., 1994; Scherer et al., 1995). Periaxin is a structural protein which binds to the dystroglycan complex, thereby maintaining myelin sheath stability, regulating Schwann cell shape and interacting in cell signalling (Court et al., 2009; Scherer et al., 1995; Sherman et al., 2001). Indeed, *PRX*-null mouse models demonstrate severe demyelinating neuropathy, as well as mislocalization and loss of the dystroglycan complex (Gillespie et al., 2000; Sherman et al., 2001). Periaxin consists of a highly basic domain and a highly acidic domain, which are separated by a long repeat domain with penta- and higher-order repeats from residues 430-730, located in the seventh and last exon (Boerkoel et al., 2001). The vast majority of pathogenic *PRX* variants are biallelic or compound heterozygous, located in the seventh exon, and involve frameshift or premature truncations, including in the repeat domain (Boerkoel et al., 2001; Peddareddygar et al., 2013). The postulated disease mechanism for such variants is inhibition of periaxin binding to the cytoskeleton, resulting in disruption of stable myelin production and normal cell signalling (Boerkoel et al., 2001; Datta et al., 2019).

DSS and Charcot-Marie-Tooth disease type 1 (CMT1) form a spectrum of neuropathy phenotypes ranging from congenital hypomyelinating neuropathy to adult-onset neuropathy. In addition to *PRX*, variants in five other genes have been associated with these conditions (Lupski & Garcia, 2001). *PRX*-associated DSS is characterized by infant-onset, slowly progressive CMT-like neuropathy, with presenting symptoms being delayed motor development, generalized

hypotonia and muscle weakness, respiratory insufficiency and distal sensory loss (Baets et al., 2011). Some affected individuals have also demonstrated early foot deformities (i.e., pes cavus), scoliosis and sensory ataxia (Baets et al., 2011).

PRX Structural Variant: This 78 bp, in-frame deletion removes 26 amino acids from the seventh exon of *PRX* (Glu495-520). The deletion is located within the *PRX* repeat domain, which spans from codon 400-700, of 1461 total codons in the protein (Boerkoel et al., 2001; Peddareddygar et al., 2013). Based on ACMG pathogenicity guidelines, the variant is classified as a VUS (score: +0.3) based on the inclusion of scoring criteria **1A** (contains protein-coding material), **3A** (contains 0-24 genes) and **2E** (in-frame deletion within the same gene = +0.3) (Riggs et al., 2020). The variant has been previously reported in affected individuals; there are four entries in ClinVar, all classified as VUS (Variation ID: 410606, Accession: VCV000410606.10). Peddareddygar et al. (2013) described an affected individual heterozygous for the variant; the individual demonstrated adolescent-onset gait difficulties and was diagnosed with CMT disease at age 35. This reported individual's unaffected family members (i.e. parents, children) did not undergo familial variant testing to investigate if the variant segregated with disease. The authors hypothesized that although the deletion would not disrupt the mRNA reading frame, it may affect the core repeats, disrupt higher-order repeats in the region, and ultimately inhibit periaxin function (Peddareddygar et al., 2013). However, the authors also suggested that because most forms of *PRX*-related disease are AR, the reported individual may have had a second, undetected variant in the other *PRX* allele. Lastly, the *PRX* variant was identified in multiple French-Canadian individuals with CMT, although the variant zygosity was not reported (Volodarsky et al., 2021). Visualization of the *PRX* variant was performed using UCSC Genome Browser (**Figure 4.10 B**). No information was

available about the variant's frequency in healthy populations (e.g., NCBI dbVar Curated Common SVs); however, the variant was shown to overlap with other deletion SVs of uncertain significance.

Clinical Correlation and Conclusions: After review by our multidisciplinary committee, the *PRX* variant remained of uncertain significance, but was considered unlikely to explain NM25's phenotype. This assessment was based on the uncertain pathogenicity of the variant itself, limited evidence that the variant could cause disease in a heterozygous state and the lack of strong phenotypic concordance with NM25. If the variant were indeed pathogenic in a heterozygous state, the variant could account for a small subset of NM25's phenotype, including infant-onset muscle weakness, generalized hypotonia and respiratory insufficiency. Alternatively, the variant may be non-pathogenic or NM25 may represent an incidental carrier for AR DSS/CMT. Additional evidence to understand the variant's pathogenicity could include family studies, identification of other affected individuals heterozygous for this variant or functional studies in model organisms. More broadly, this case demonstrates the ability of LR-WGS to accurately resolve SVs, as well as the challenge of interpreting these variants in the absence of technology-matched tools. As LRS becomes more widely adopted, improved SV interpretation tools and larger population-scale LRS datasets will be essential for accurate variant classification. Furthermore, expanding genomic reference datasets to better represent individuals of non-European ancestries would support more equitable and accurate variant interpretation. This is particularly relevant for NM25, as she was of Cree ancestry and as Indigenous populations remain underrepresented in existing genomic databases Canadian initiatives such as Silent Genomes aim to address these disparities through the development of Indigenous-led genomic resources and data governance frameworks (Caron et al., 2020). Future work could include requesting access to Silent Genomes data to assess the

frequency of this variant in Indigenous populations. Lastly, as the variant was identified in a research setting, validation in a clinically accredited laboratory is required.

4.2.2 LR-WGS characterized ATXN8OS STR expansion of uncertain clinical significance in father-son duo (NM02 and NM01)

Clinical Investigations: NM02 (age 50) and NM01 (age 26) are a father-son duo. Standardized phenotypic terms representing individuals' clinical features (HPO terms) are listed in **Table 4.24**. Both individuals exhibited mild, congenital-onset, static, proximal limb weakness and exertional fatigue, and abnormal gait patterns. NM02 and NM01 described difficulty keeping up with peers during physical activities and had difficulty climbing stairs or jumping. NM01 also demonstrated post-exertional fatigue and muscle pains in childhood. While NM02 wasn't assessed in clinical genetics until 35 years old, genetics assessment of NM01 at age 13 noted weakness of neck flexion and shoulder hip girdles, globally hypoactive reflexes, brachydactyly, hypertrophic scarring, laxity of knees and elbows. Both individuals had keratosis pilaris. Over time, NM01's phenotype gradually improved; follow-up with clinical geneticist at age 26 yielded a completely normal exam with no evidence of muscle weakness, gait difference or other features of NMD. Conversely, NM02 experienced onset of progressive proximal and distal weakness with lower predominance at age 20, along with proximal lower limb atrophy, persistent muscle aches (exacerbated with rest) and worsening gait pattern. At age 31, muscle weakness and atrophy began progressing more rapidly, with difficulty arising from chair or squat, and inability to stand on toes. By age 34, NM02 also developed frequent muscle cramps worse at night-time, as well as prominent quadricep fasciculations and occasional brachioradialis fasciculations. Lastly, NM02 also reported chronic mild-moderate dull back pain with onset at age 25, and chronic migraine/tension headaches which were unilateral moderate-severe with photophonophobia. In summary, both individuals were affected with mild congenital static proximal limb weakness and

exertional fatigue, while NM02 (father) was also affected with a 20-year-old onset of progressive and distal muscle weakness and lower limb atrophy.

Table 4.24. Human Phenotype Ontology (HPO) terms selected for NM01 and NM02 based on their clinical presentation¹. The presence of an ‘X’ indicates that the term is applicable to that individual.

HPO Term Description	HPO Code	NM01	NM02
Low back pain	HP:0003419	X	X
Proximal muscle weakness in lower limbs	HP:0008994	X	X
Muscle spasm	HP:0003304	X	X
Difficulty climbing stairs	HP:0003551	X	X
Difficulty running	HP:0009046	X	X
Broad-based gait	HP:0002136	X	X
Keratosis pilaris	HP:0032152	X	X
Hyperlordosis	HP:0003307	X	
Limb-girdle muscle weakness	HP:0003325	X	
Fatty replacement of skeletal muscle	HP:0012548		X
Distal muscle weakness	HP:0002460		X
Limb-girdle muscle weakness	HP:0003325	X	
EMG: axonal abnormality	HP:0003482		X
EMG: neuropathic changes	HP:0003445		X
EMG: chronic denervation signs	HP:0003444		X

EMG: myopathic abnormalities	HP:0003458	X
Exercise-induced muscle cramps	HP:0003394	X
Easy fatigability	HP:0003388	X
Frequent falls	HP:0002359	X
Quadriceps muscle atrophy	HP:0009050	X
Limited knee flexion/extension	HP:0005085	X
Joint hypermobility	HP:0001382	X
Reduced tendon reflexes	HP:0001315	X
Impaired toe-walking ability	HP:0043052	X
Limb fasciculations	HP:0007289	X
Elevated circulating creatine kinase concentration	HP:0003236	X
Myalgia	HP:0003326	X
Exercise-induced myalgia	HP:0003738	X
Exercise intolerance	HP:0003546	X

¹The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3. Terms ordered from first to last in the table according to the order used in the HPO hierarchy, beginning with terms applicable to both individuals.

NM02 and NM01 also demonstrated differences in clinical testing (**Table 4.25**). NM02 demonstrated persistently elevated serum CK, and EMG and NCS from ages 33 to 35 revealed both chronic neurogenic changes, as well as myopathic abnormalities. Conversely, NM01's serum CK, EMG and NCS were all normal between ages nine and 14. Both NM02 and NM01 demonstrated myopathic features on muscle MRI of the lower limbs at ages 37 and 13,

respectively, however, NM02 demonstrated more prominent atrophy and fatty replacement, as well as increased short tau inversion recovery (STIR) sequence consistent with denervation or inflammation. Finally, only NM01 received muscle biopsy, which was relatively unremarkable at age 14 with some slight prevalence of type 2 fibers. Collectively, NM02's phenotype at the time of clinical investigations was more prominently affected, with a potentially mixed myopathic-neuropathic phenotype.

Table 4.25. Summary of neuromuscular investigations in NM02 and NM01.

<u>NM02</u>			<u>NM01</u>	
Investigation	Finding	Age ¹	Finding	Age ¹
Serum CK	Persistently elevated ²	35	Normal	12-14
Muscle MRI	Myopathic (hips and tibia fibula), increased STIR signal	37	Mild myopathic ³	13
EMG	Myopathic	33-35	Normal	9 and 14
Muscle biopsy	Not performed	-	Unremarkable, slight prevalence type 2 fiber	14
NCS	Chronic neurogenic changes ⁴	33-35	Normal	9 and 14

¹Age in years; ²600-900 U/L (reference: 52-175 U/L); ³Findings: symmetric atrophy and fatty replacement most prominent in the gluteus maximus and tensor fascia lata muscles, and more subtle in the sartorius, gracilis, biceps femoris, semimembranosus and adductor magnus; ⁴Findings: denervation, decreased amplitude. CK, creatine kinase; EMG, electromyography; MRI, magnetic resonance imaging; NA, not available/not found during chart review; NCS, nerve conduction studies.

Family History: NM02 and NM01 are of Mennonite ancestry from Paraguay; there may be distant consanguinity in NM02's parents. NM02 has five full siblings, while NM01 has one full brother and sister. Two of NM01's maternal uncles may have a neuropathy with underlying myopathy. The family history is otherwise negative for any other neuromuscular disorder.

Prior Genetic Testing: NM02 underwent four previous assays, including a six-gene panel, 46-gene panel, single-gene testing via multiplex ligation-dependent probe amplification (MLPA) for *SMN1* and *SMN2*, and PCR analysis to determine CAG repeat count in *AR*, which is associated

with SBMA X-linked 1. NM01 also underwent four previous assays, including a chromosomal microarray and three targeted genetic tests. These last consisted of *DMPK* CTG repeat analysis for myotonic dystrophy type 1, *SMN1* gene dosage studies (i.e. MLPA), and *FKRP* sequencing for dystroglycan-related LGMD R9.

LR-WGS and LR-RNA-seq Analyses: LR-WGS Variant Filtration did not yield any variants of interest for NM01 or NM02; the filtering strategy yielded **four candidate variants for NM01** and **five for NM02**, all excluded due to lack of phenotypic correlation. LR-RNA-guided filtering did not identify any variants or genes of interest for NM01 for whom a muscle biopsy was available. However, TRGT STR analysis identified elevated CTG repeat count in *ATXN8OS* in both NM02 and NM01. NM02 demonstrated one allele with 74 CTG repeats (12 reads, 99.6% purity), while NM01 demonstrated one allele with 72 CTG repeats (21 reads, 98.8% purity) (**Figure 4.11**). CTG repeat expansions (>54) have been associated with spinocerebellar ataxia 8 (SCA8), an autosomal dominant condition with reduced penetrance (Cleary et al., 2021).

ATXN8OS: *ATXN8OS* (ataxin-8 opposite strand) is located on chromosome 13q21.33 and encodes a long non-coding RNA transcript (Koob et al., 1999; Moseley et al., 2006). The 5' end of *ATXN8OS* overlaps with *KLHL1* (Nemes et al., 2000); the encoded actin-binding protein (KLHL1) is implicated in neuronal structure and function through modulation of calcium channel activity (Aromolaran et al., 2009; He et al., 2006). As such, it is hypothesized that *ATXN8OS* may regulate *KLHL1* expression through an antisense mechanism (Nemes et al., 2000). Both *ATXN8OS* and *KLHL1* are primarily expressed in human brain tissues, including the cerebellum, which is most affected by SCA8 (Lonsdale et al., 2013; Nemes et al., 2000).

CTG repeat expansions in *ATXN8OS* have been associated with SCA8 through multiple affected families with linkage disequilibrium, as well as unrelated individuals (Koob et al., 1999;

Silveira et al., 2000). Clinically, SCA8 is characterized by slowly progressive gait and limb ataxia, scanning dysarthria and eye-movement abnormalities with onset between 18-72 years of age (Ikeda et al., 2000; Koob et al., 1999). Additional manifestations may include upper motor neuron findings, extrapyramidal signs, brainstem signs, sensory neuropathy and cognitive features (Cleary et al., 2001/2021).

ATXN8OS Repeat Expansions: Interpretation of *ATXN8OS* remains challenging, as the true repeat size at which individuals are at risk of developing SCA8 is not well understood (Gupta & Jankovic, 2009). The current *ATXN8OS* interpretation states that unaffected alleles have 50 or fewer CTG repeats (Cleary et al., 2021). Pathogenic alleles are characterized by 54 to 250 CTG repeats, although reduced penetrance occurs for CTG repeats of all sizes (Cleary et al., 2021). In particular, expanded alleles between 50-100 CTG repeats lack clear evidence of either pathogenicity or benignity. For example, while the original individual affected with SCA8 demonstrated 80 CTG repeats (Silveira et al., 2000), mean repeat length in unaffected individuals from a large SCA8 kindred was 92.4 (vs. 116.8 in affected individuals) (Koob et al., 1999). However, studies of control chromosomes from the general population have demonstrated relative rarity of repeat expansions; among studies of 1200 and 1818 control chromosomes, fewer than 1% of alleles had >37 or 40 repeats, respectively (Koob et al., 1999; Silveira et al., 2000). Some more recent studies have supported that expanded repeat size can be seen in unaffected individuals, and outlined that additional genetic modifiers are likely to contribute to disease risk. These include haplotype background and CGG interruptions within the CAG repeat tract (Ikeda et al., 2004; Perez et al., 2021). In sum, the 74- and 72-repeat alleles identified in NM02 and NM01, without CGG interruptions, are best interpreted as variants of uncertain clinical significance with a possible lifetime risk of developing SCA8.

Clinical Correlation and Conclusion: NM02 and NM01's LR-WGS did not yield a clear candidate variant to explain their neuromuscular presentations. LR-WGS did identify an *ATXN8OS* repeat expansion in both individuals, of uncertain clinical significance. SCA8 does not overlap clinically with the congenital proximal weakness observed in either participant, nor with the adult-onset progressive proximal and distal weakness in NM02. However, if a true pathogenic finding were present, SCA8 could have contributed to NM02's neuropathic findings on EMG. Moreover, NM02 and NM01 may be pre-symptomatic and at risk of developing SCA8 later in life. As such, genetic counselling may be useful to navigate uncertainty of this result and its potential implications. Further information to elucidate the relevance of these findings in future would include more population-level genotype information (e.g., large, population-level studies to understand the frequency of this repeat size in unaffected individuals), as well as research to elucidate risk factors contributing to SCA8 development in individuals with increased CTG repeat length.

In summary, the *ATXN8OS* findings in NM01 and NM02 demonstrate the utility of LR-WGS for detecting and characterizing repeat expansions that may not be readily captured by conventional SRS-based genomic approaches. These cases also underscore the need for improved evidence on repeat-length thresholds, penetrance, and genotype–phenotype relationships to enable more accurate interpretation and counselling regarding *ATXN8OS*-associated disease risk.

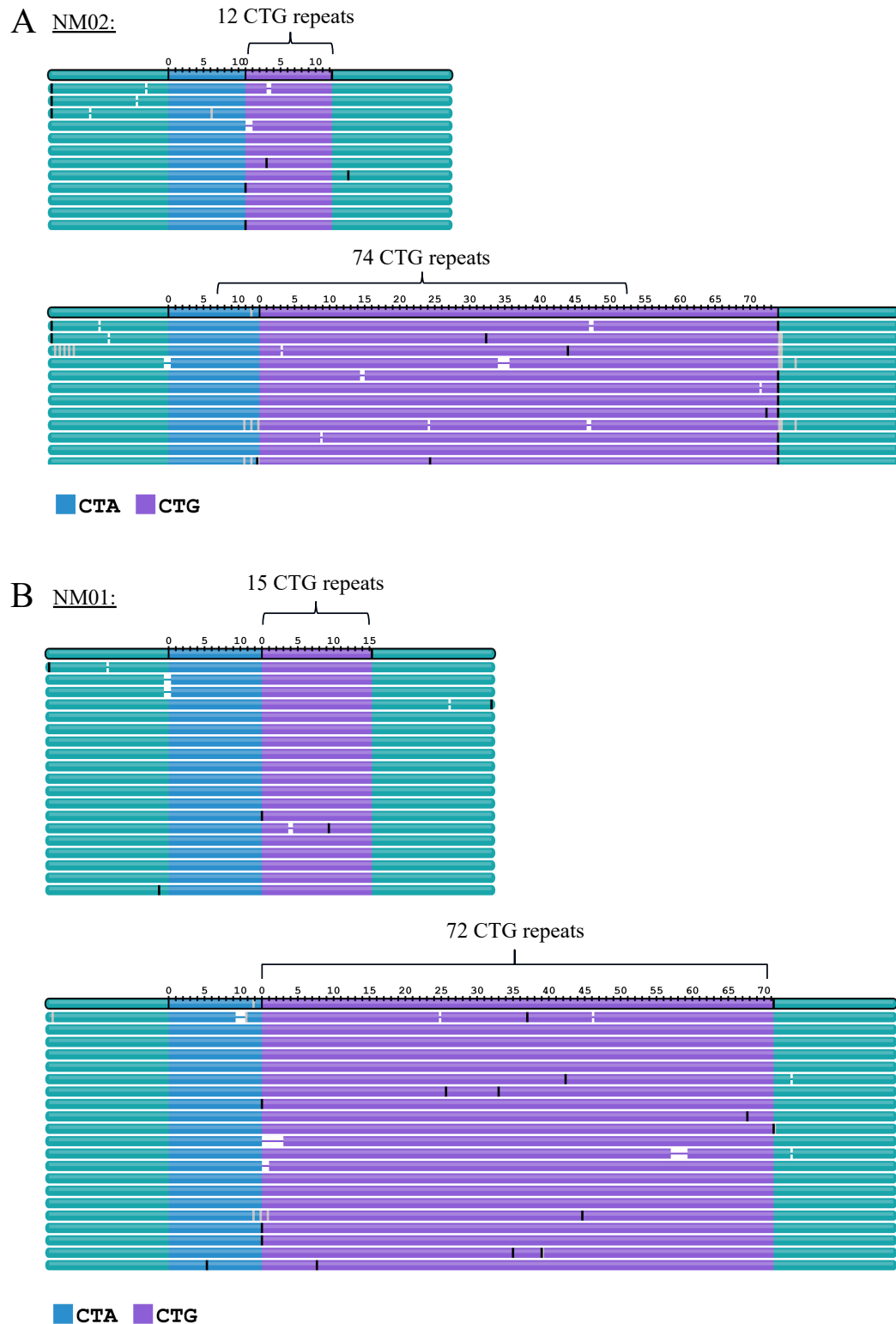


Figure 4.11. Father (NM02) and son (NM01) duo with NMDs demonstrate heterozygosity for elevated CTG repeat number in *ATXN80S*, associated with spinocerebellar ataxia 8. Results derived from TRGT analysis using LR-WGS from

whole blood, and visualized on TRVZ. Both *ATXN8OS* alleles are shown for each individual; the non-expanded alleles are demonstrated above the expanded allele. CTG repeat length is indicated with a black bracket and repeat number annotation above each allele. A) CTG repeat profile for NM02 at *ATXN8OS* locus. B) CTG repeat profile for NM01 at *ATXN8OS* locus. NMD, neuromuscular disorder; TRGT, Tandem Repeat Genotyping Tool; TRVZ, Tandem Repeat Visualizer.

4.3. Case Descriptions – Functional insights from disease-relevant tissue transcriptomic profiles

4.3.1. *LR-RNA-seq from disease-relevant tissue identified mitochondrial transcriptomic dysregulation supporting a working diagnosis of mitochondrial myopathy (NM17)*

Clinical Description: NM17 is a 55 year old female with slowly progressive ophthalmoplegia (weakness of muscles controlling eye movement) and ptosis (muscle weakness of the upper eyelid), alongside mild proximal muscle weakness and exercise intolerance/fatigability, starting at age 37. Physical examination at age 42 also demonstrated mild right scapular winging. At age 46, she began developing significant joint stiffness and pain, worse in the morning and exacerbated with rest. With normal inflammatory markers and non-response to medications for inflammatory arthritis medications, the joint features were believed to be related to a mitochondrial myopathy. NM17 developed sensorineural hearing loss (SNHL) at age 43, first right-sided, then bilateral. In later years her phenotype also included lower-limb ataxia and progressive balance disturbances, early-morning headaches and poor sleep quality, and leg muscle cramping at rest. **Table 4.26** provides a summary of standardized phenotypic (HPO) terms to represent NM17's neuromuscular features, including ocular, skeletal muscle and ataxic symptoms, and hearing loss. Clinical testing performed between 42 and 54 years of age demonstrated features consistent with a mitochondrial myopathy, including persistently elevated serum CK, myopathic EMG abnormalities and muscle biopsy findings of abnormal mitochondrial distribution, ragged red fibers and cytochrome c oxidase deficiency affecting 5-10% of fibers (**Table 4.27**).

Table 4.26. Human Phenotype Ontology (HPO) terms selected for NM17 based on their clinical presentation.¹

HPO Term Description	HPO Code
Arthralgia/arthritis	HP:0005059
Amyotrophy involving the shoulder musculature	HP:0001465
Scapular winging	HP:0003691
Shoulder girdle muscle weakness	HP:0003547
Ragged-red muscle fibers	HP:003200
Weakness of facial musculature	HP:0030319
Myopathy	HP:0003198
Proximal amyotrophy	HP:0007126
Pelvic girdle amyotrophy	HP:0008946
Shoulder girdle muscle atrophy	HP:0003724
Proximal muscle weakness	HP:0003701
EMG myopathic abnormalities	HP:0003458
Difficulty climbing stairs	HP:0004302
Multiple mitochondrial DNA deletions	HP:0003689
Mitochondrial myopathy	HP:0003737
Increased muscle fatiguability	HP:0003750
Joint stiffness	HP:0001387
Elevated circulating creatine kinase concentration	HP:0003236
External ophthalmoplegia	HP:0000544
Diplopia	HP:0000651

Bilateral ptosis	HP:0001488
Bilateral sensorineural hearing impairment	HP:0008619

¹The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3. Terms ordered from first to last in the table according to the order used in the HPO hierarchy.

Table 4.27. Summary of neuromuscular investigations in NM17.

Investigation	Finding	Age ¹	Additional Information
CK	Persistently elevated	51-54	Levels between 341-800+U/L (reference <190U/L)
Muscle MRI	Not performed	-	-
EMG	Myopathic	42	Mild increase in broad, low amplitude polyphasic units with early recruitment at left bicep, left deltoid and vastus medius. Milder effect in tibialis anterior bilaterally.
Muscle biopsy²	Mitochondrial myopathy	43	Features consistent with mitochondrial myopathy: abnormal mitochondrial distribution, ragged red fibers and 5-10% of fibers deficient in cytochrome c oxidase.
NCS	Normal	42	-

¹Age in years. ²Sample site: left quadriceps. CK, creatine kinase; EMG, electromyography; MRI, magnetic resonance imaging; NCS, nerve conduction studies.

Family History: NM17's parents were non-consanguineous individuals of First Nations (Mi'kmaq) and European (French and Scottish) descent. She had eight half-siblings and one identical twin sister. NM17 had two daughters, one of whom may be demonstrating the initial manifestation of their neuromuscular symptoms. No other family members are known to have features of NMDs.

Prior Genetic Testing: NM17 underwent six forms of previous genetic testing, one of those being confirmation of zygosity with her identical twin. Targeted genetic testing for oculopharyngeal muscular dystrophy via *PABPN1* repeat analysis was negative. Given NM17's muscle biopsy was suggestive of mitochondrial myopathy, comprehensive mitochondrial analysis

was performed through four separate assays. In the muscle sample, no pathogenic mitochondrial DNA (mtDNA) variants were identified; however, multiple heteroplasmic independent mitochondrial deletions were present, suggesting a disorder of mitochondrial maintenance. Additional testing in muscle for common mtDNA variants and across 19 nuclear genes was non-diagnostic. Lastly, sequencing and deletion/duplication testing of 16 genes associated with mtDNA depletion conditions in both whole blood and muscle tissue was also negative. Given that NM17 has an unaffected identical twin sister, the suspected inheritance pattern for her NMD is either heteroplasmy for a rare mtDNA variant or mosaicism for a dominant nuclear gene. However, DNA from the twin sister was not available for any analyses in this study.

LR-WGS Variant Filtration: Variant filtration yielded seven variants for further assessment, including four SNVs and three SVs. Of these, six were not considered for further analysis due to a lack of association with the patient's phenotype; further, three variants were in genes associated with autosomal recessive conditions, for which no second variant was identified. The final variant is a 113 bp deletion in *DNAJB6*; this SV is a deep intronic variant between exons nine and ten. Pathogenic variants in *DNAJB6* cause LGMD D1; this condition has some phenotypic overlap with our participant (proximal muscle weakness and atrophy, elevated serum CK); however, the patient's ophthalmoplegia and ptosis, nor SNHL, are clinical features in this condition. Further, the variant is not consistent with the variant type reported in families affected by LGMD D1, namely missense variants with a dominant toxic effect that reduces the protective anti-aggregation effect of *DNAJB6* (Couthouis et al., 2014; Sarparanta et al., 2012). Altogether, the variant was not considered likely to be causal for NM17's NMD.

LR-RNA-seq Guided Filtering: LR-RNA-seq guided filtering did not yield any genes or variants of interest for NM17. Of the 25 top gene expression and isoform outliers identified by

IsoRanker, 15 genes had established gene-disease associations. Among these, six were associated with NMDs and nine with other OMIM phenotypes. However, 14 genes were not considered of interest due to lack of phenotypic correlation with NM17. The remaining gene (*ANXA11*) was highlighted by IsoRanker as having a novel splice site. The associated condition (AD inclusion body myopathy, OMIM #619733) showed some clinical correlation with the patient's phenotype; however, LR-WGS analysis of *ANXA11* using relaxed filters did not reveal any candidate SNVs, CNVs or SVs.

However, IsoRanker analysis identified upregulation of various mitochondrial genes in NM17 relative to the cohort, as well as increased alternative splicing of mitochondrial gene isoforms (**Figure 4.12**). Individually, these mitochondrial genes are not associated with a specific OMIM phenotype. However, the broader transcriptomic profile for NM17 suggests widespread dysregulation of mitochondrial gene expression and splicing in skeletal muscle tissue, not observed in IsoRanker results for other individuals in the cohort. These results provide functional molecular support for the suspected or working diagnosis of mitochondrial myopathy for NM17.

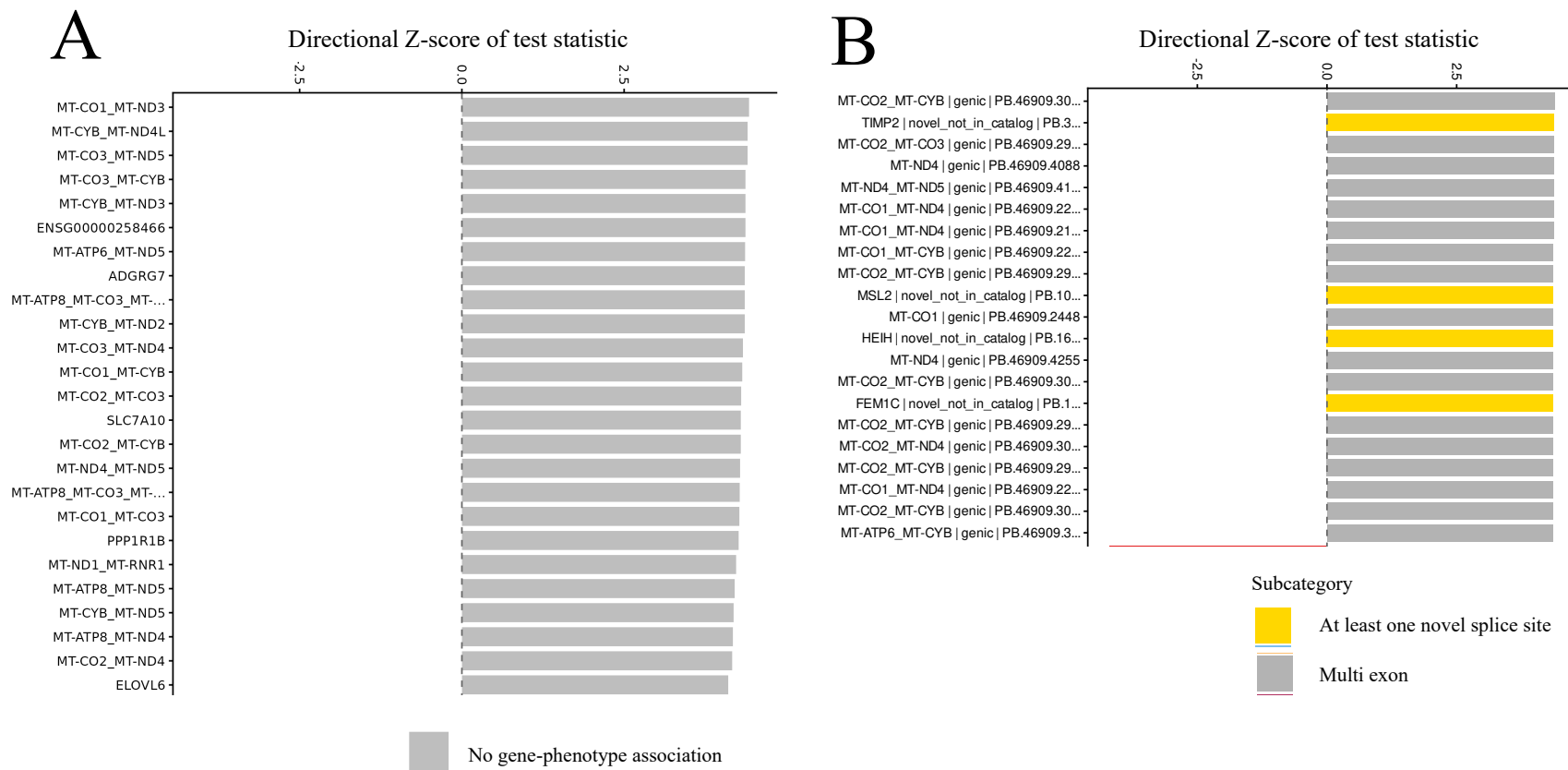


Figure 4.12. IsoRanker-guided LR-RNA-seq identifies mitochondrial gene upregulation and increased alternative splicing in NM17 relative to NMD cohort ($n = 22$). A) Gene-level expression outliers in NM17. Genes associated with NMDs are shown in purple, genes associated with other clinical phenotypes in OMIM are shown in blue, and genes without known clinical phenotype associations are shown in grey. Negative directional Z-scores indicate decreased expression relative to cohort, while positive directional Z-scores indicate increased expression relative to cohort. B) Isoform expression outliers in NM17 using a directional Z score $\geq [2.5]$. Bar colors indicate isoform subcategory according to PacBio Iso-Seq annotation (Guizard et al., 2023). LR-RNA-seq; long-read RNA sequencing; NMD, neuromuscular disease; PacBio Iso-Seq, Pacific Biosciences Isoform Sequencing.

Clinical Correlation and Conclusions: As reviewed with our multidisciplinary committee, no pathogenic variants were identified through our LR-WGS and LR-RNA-seq analyses in NM17. LR-RNA-seq provided functional information that the underlying disease mechanism in NM17 involves broad mitochondrial gene expression and splicing dysfunction. These findings demonstrate the value of integrated multi-omic analyses in uncovering disease-relevant molecular changes and in supporting clinical interpretation, even when a causative genomic variant remains unresolved. For this participant, further resolution of the molecular basis of disease may be elucidated through additional analyses. Firstly, comparative analysis of LR-WGS of NM04's identical twin sister, who is apparently unaffected with neuromuscular symptoms, would provide a powerful opportunity to identify potential germline mosaic variants. In addition, given the potential for a mosaic variant to affect skeletal muscle function, genetic analysis of skeletal muscle tissue in NM04 could improve sensitivity in detecting variants that may be absent or underrepresented in blood-derived DNA.

4.3.2. LR-RNA-seq identifies inflammatory and immune pathway activation (NM04)

Clinical Description: NM04 is a 65-year-old female. At age 40, she developed progressive muscle pain and cramping, mild asymmetric weakness of shoulder abductors and elbow flexors (with left predominance), and exercise intolerance. By age 56, NM04 reported dyspnea with minimal activity, as well as visible muscle twitching in her left thigh. NM04 did not report myotonia, but physical examination at one point demonstrated significantly delayed relaxation of grip. Standardized phenotypic (HPO) terms to represent NM04's neuromuscular features are listed in **Table 4.28**. Clinical testing performed between 53 and 56 years of age demonstrated myopathic features (**Table 4.29**). This included persistently elevated serum CK and several areas of fatty replacement in the thigh and calf musculature, with no increase of signal intensity to suggest an inflammatory component. Fatty replacement was severe at the medial gastrocnemius, moderate at the distal vastus lateralis and mild at the peroneal muscles, and was suggestive of a structural myopathy. At age 55, muscle biopsy of the right thigh had features consistent with necrotizing myopathy, including inflammatory infiltrate (mainly endomysial), expression of major histocompatibility complex class I, deposition of complement C5b-9, endomysial fibrosis, fibre-size variability, increased number of internal nuclei and scattered nuclear clumps.

Table 4.28. Human Phenotype Ontology (HPO) terms selected for NM04 based on their clinical presentation.¹

HPO Term Description	HPO Code
Shoulder girdle muscle weakness	HP:0003547
Increased variability in muscle fiber diameter	HP:0003557
Increased endomysial connective tissue	HP:0100297

Muscle fiber necrosis	HP:0003713
Inflammatory myopathy	HP:0009071
Difficulty climbing stairs	HP:0003551
Frequent falls	HP:0002359
Easy fatigability	HP:0003388
Fatty replacement of skeletal muscle	HP:0012548
EMG: myopathic abnormalities	HP:0003458
Handgrip myotonia	HP:0012899
Muscle stiffness	HP:0003552
Hyperuricemia	HP:0002149
Elevated circulating creatine kinase concentration	HP:0003236
Hyperglycemia	HP:0003074
Myalgia	HP:0003326
Trochanteric pain	HP:6001158
Exercise intolerance	HP:0003546

¹The listed HPO terms are the most specific terms matching the participant phenotype selected from the HPO hierarchy described in Section 2.3. Terms ordered from first to last in the table according to the order used in the HPO hierarchy.

Table 4.29. Summary of neuromuscular investigations in NM04.

Investigation	Finding	Age¹	Additional Information
Serum CK	Persistently elevated	53-55	495-1057 U/L (reference: 28-110 U/L)
Muscle MRI	Fatty replacement (thigh and calf muscles)	56	Fatty replacement severe at the medial gastrocnemius, moderate at the distal vastus lateralis and mild at the peroneal muscles
EMG	Normal	53	
Muscle biopsy²	Necrotizing myopathic	55	Necrotizing myopathic features: inflammatory infiltrate (mainly endomysial), expression of

		MHC class I and deposition of complement C5b-9, endomysial fibrosis, fibre size variability, increased number of internal nuclei and scattered nuclear clumps
NCS	Normal	54
¹ Age in years. ² Sample site: right thigh. CK, creatine kinase; EMG, electromyography; MRI, magnetic resonance imaging; NCS, nerve conduction studies.		

Family History: NM04's parents are second cousins of European (English and German) descent. She has three full brothers, as well as three daughters and one son, all apparently unaffected. No family members are known to be affected with neuromuscular disorders.

Prior Genetic Testing: NM04 underwent four forms of genetic testing. Targeted repeat analysis in *DMPK* and *CNBP* for myotonic dystrophy types I and II, respectively, was negative. A 43-gene panel for structural myopathy was undertaken and was non-diagnostic, but revealed two VUS; subsequently, her mother and father underwent targeted variant testing to determine inheritance for these two variants. The first VUS was a maternally-inherited *FKTN* missense variant; pathogenic variants in *FKTN* are associated with AR muscular dystrophy-dystroglycanopathies type A4, B4 and C4, as well as dilated cardiomyopathy 1X. As no second allele was identified in the gene panel, the *FKTN* variant was not considered causal for NM04's phenotype. Secondly, a missense variant was identified in *SMCHD1*. Variants in this gene are known to cause fascioscapulohumeral muscular dystrophy (FSHD) type II when combined with a *permissive haplotype of the FSHD1 locus at the terminus of chromosome 4q*. As NM04's father was found to carry the same *SMCHD1* variant and as he does not have any neuromuscular symptoms, this variant could not be deemed causative for NM04's neuromuscular phenotype.

Lastly, since enrolment in the present study, NM04 underwent genetic testing for FSHD type I via pulse field gel electrophoresis/Southern blot and hybridization of the 4q35 region. This testing, performed on a blood sample, identified a mosaic 31 kb deletion encompassing the entire

D4Z4 repeat area on the permissive 4qa haplotype. FSHD Type 1 is typically caused by heterozygous pathogenic contractions of the D4Z4 repeat locus from ≥ 11 repeat units in unaffected individuals to ≤ 7 repeat units in fully penetrant alleles (van Overveld et al., 2003). Repeat contraction disrupts methylation patterns and leads to expression of the adjacent gene *DUX4*, typically silenced in mature skeletal muscle, thus resulting in aberrant cell signalling and disease pathogenesis (Lim et al., 2020). As such, it is plausible that deletion of all D4Z4 repeats due to the 31 kb deletion observed in NM04 could also cause the condition. Some uncertainty regarding this diagnosis persists due to the mosaic nature of the deletion; the percentage of mosaicism and which tissues are affected is not known, for example, whether the deletion is present in skeletal muscle. Further, several unaffected individuals have been reported in a family with a chromosomal translocation resulting in monosomy of the D4Z4 region (Tupler et al., 1996); this evidence suggests that at least one D4Z4 repeat may be necessary to cause *DUX4* expression and thus questions the effect of a complete deletion of the repeat locus (Sidlauskaitė et al., 2020). However, given the strong phenotypic correlation between the patient's symptomatology, as well as clinical testing results (inflammatory and necrotic muscle biopsy, EMG and elevated CK), this genetic finding is strongly believed to be causal for the patient's neuromuscular condition.

As our study was performed after this FSHD1 finding, LR-WGS did not yield a new diagnosis for the participant; however, LR-RNA-seq provided evidence of inflammatory processes occurring in the muscle biopsy, thereby providing functional evidence to support the mosaic genetic diagnosis of FSHD Type 1 (**Section 3.5.2**).

LR-WGS Variant Filtration: Given the mosaic nature of NM04's FSHD type 1 diagnosis, we still performed NMD-centred LR-WGS filtering. This revealed four SNVs and no CNVs/SVs. None of the four variants were in genes with strong clinical correlation to NM04; further, two

variants were in genes associated with autosomal recessive conditions, for which a second allele was not detected.

Of note, our bioinformatic pipeline did not detect the 31 kb deletion at 4q35 identified in prior genetic testing. The 4q35 region is highly challenging to resolve accurately even with LR technologies, owing to the long and highly repetitive sequence architecture, extreme GC content and homology with chromosome 10q26 (Huang et al., 2024; Lyle et al., 1995). Moreover, the 4q35 locus is located within a subtelomeric region of chromosome 4 that is poorly represented in current genome assemblies, thus complicating read alignment and variant detection (Nurk et al., 2022). The upcoming Telomere-to-Telomere (T2T) Consortium reference genome may support improved sequencing resolution of this locus (Nurk et al., 2022). Further, additional computational tools to resolve large unit variable number tandem repeats in LR-WGS data may support analysis of this region, such as kivvi (Chen et al., 2025). Additionally, LR analytical frameworks integrating multiple sequencing modalities, including LR-WGS and optical genome mapping, demonstrated strength in resolving complex subtelomeric structural variation and rearrangements at 4q35 (Tardy et al., 2025).

LR-RNA-Guided Filtering for NM04 did not yield any genes or variants of interest. Analysis of LR-RNA data with IsoRanker identified gene expression and/or isoform outliers in 12 genes, two of which were associated with NMDs and ten with other OMIM. All 12 genes flagged by IsoRanker were not considered of interest, due to a lack of phenotypic correlation between the associated condition and NM04's clinical presentation.

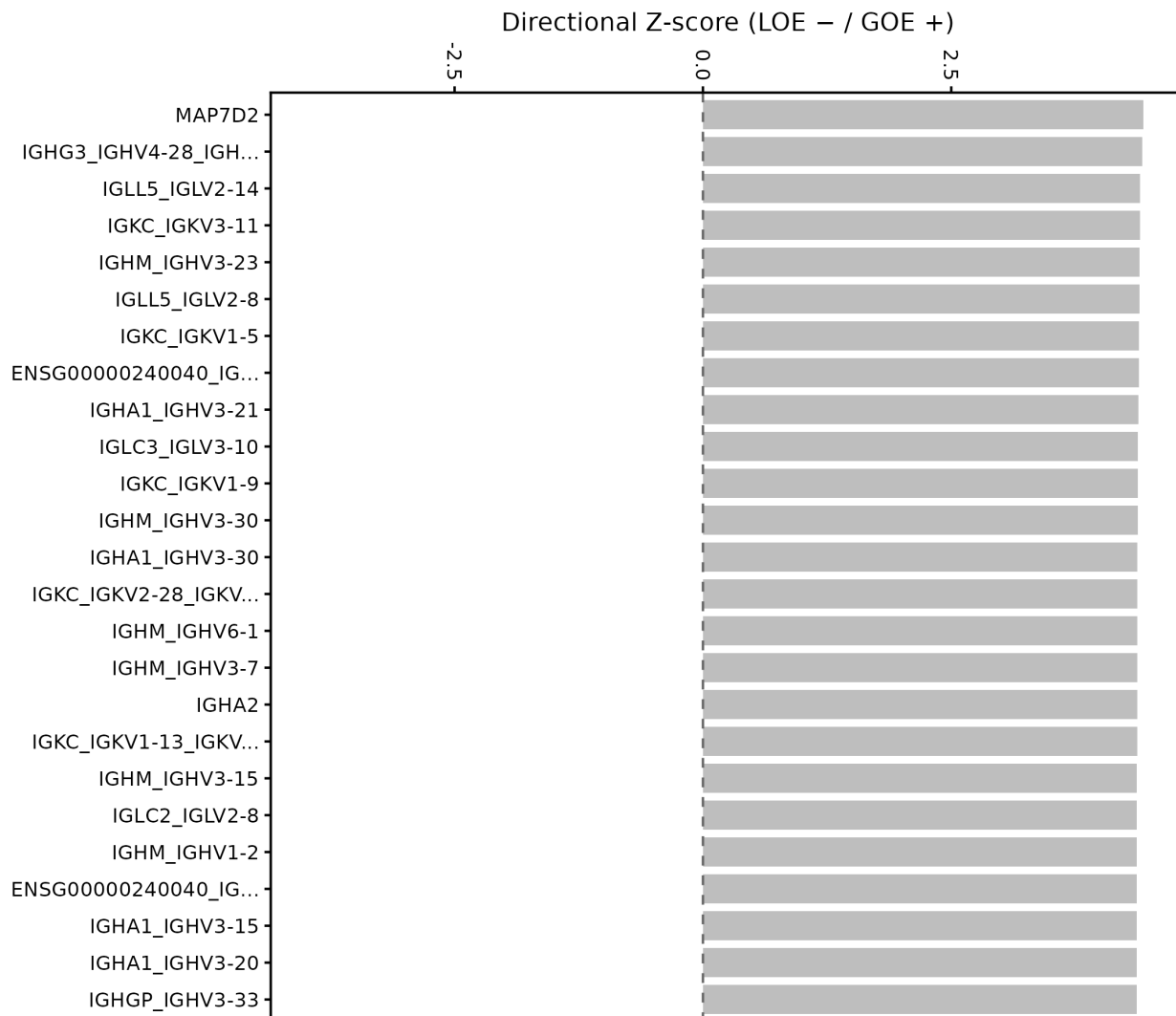


Figure 4.13. Long-read transcriptomic analysis using IsoRanker identifies increased immune-related gene expression in NM04 relative to the neuromuscular disease (NMD) cohort ($n = 18$). Gene-level expression outliers in NM04 relative to cohort. Gene names are displayed on the left side. Negative directional Z-scores indicate decreased expression, while positive directional Z-scores indicate increased expression relative to cohort.

Our LR-RNA-seq analysis demonstrated increased expression of several inflammatory and immune-associated genes, including *IGHG3*, *IGHA1*, *IGHA2*, *IGHM*, *IGKC*, *IGLC3* and *IGLL5*. (**Figure 4.13**). This finding is suggestive of activation of inflammatory pathways within skeletal muscle tissue. Several of these genes encode components of antibodies, namely *IGHG3*, *IGHA1*, *IGHA2*, and *IGHM*, which encode immunoglobulin heavy chains, and *IGLC3* and *IGLL5*, which encode immunoglobulin light chains (Jana et al., 2026). These genes are typically expressed in B cells and plasma cells; therefore, increased expression of immunoglobulin genes in bulk RNA sequencing of skeletal muscle suggests local infiltration or accumulation of immune cell types in the muscle tissue (Jana et al., 2026). Local infiltration of immune cell types is supported by the endomysial inflammatory infiltrate denoted on the muscle biopsy. Further, NM04's muscle biopsy showed C5b-9 membrane attack complex deposition, which has been associated with promotion of immune cell infiltration to muscle during FSHD1 pathogenesis (Hubregtse et al., 2024; Wong et al., 2022); indeed, using analysis of RNA with qPCR from 12 FSHD subjects and 13 unaffected first-degree relatives, the complement system was one of the most overrepresented pathways among genes differentially expressed in FSHD biceps compared with control muscles. Further, the pattern of multiple immunoglobulin types, notably of IGM, IGG and IGA, may reflect various components of the humoral immune system, including B cell activation, differentiation into antibody-secreting plasma cells and class-switch recombination and altogether an active adaptive immune response (Janeway et al., 2001). In sum, the transcriptomic expression profile identified using LR-RNA-seq may highlight chronic immune cell recruitment and activation rather than direct immunoglobulin expression by muscle fibers.

Moreover, this finding is consistent with the established pathophysiology of FSHD1, in which chronic inflammation is recognized as a key contributor to disease progression, fatty

replacement of muscle and muscle degeneration (Arahata et al., 1995; Dahlqvist et al., 2019). In FSHD1, contraction of the D4Z4 repeat array results in epigenetic relaxation and incomplete somatic repression of the *Dux4*, which is normally expressed only during embryonic development and in the testis, but silenced in mature skeletal muscle (De Iaco et al., 2017; Hendrickson et al., 2017; Lemmers et al., 2010, 2012). Aberrant re-expression of *DUX4* in skeletal muscle leads to activation of pathogenic transcriptional programs, including immune and inflammatory signaling pathways (Frisullo et al., 2011; Geng et al., 2012). As such, the inflammatory transcriptional signature identified in NM04 may reflect downstream effects of *DUX4*-mediated muscle toxicity and chronic immune activation occurring within disease-relevant tissue.

Clinical Correlation and Conclusions: LR-RNA-seq provides functional evidence supporting activation of disease-associated molecular pathways in NM04. This finding strengthens the mosaic diagnosis of FSHD1, based on a 31 kb deletion at 4q35. No additional variants of interest were identified in NM04, and LR-WGS did not capture the previously reported 31 kb 4q35 mosaic deletion related to the FSHD1 diagnosis, highlighting a limitation in resolving this complex sub telomeric region using current reference genomes (GRCh38).

Chapter 5: Discussion

NMDs are genetically heterogeneous inherited conditions for which half of affected individuals remain without genetic diagnosis through current clinical standards of genetic testing, including SRS-based single-gene testing, gene panels and WES (Beecroft et al., 2020; Chae et al., 2015; Gibson et al., 2018; Gonzalez-Quereda et al., 2020; Park et al., 2019; Westra et al., 2019). This persistent diagnostic gap highlights the need to generate further evidence regarding the clinical utility and interpretive value of alternative diagnostic approaches, particularly LRS and multi-omic approaches. This study evaluated the utility of integrated LR-WGS and LR-RNA-seq, using the latest PacBio Revio platform, for diagnostic applications in 22 individuals with genetically undiagnosed NMDs. Using enhanced variant-calling and annotation tools, together with a custom NMD-focused bioinformatic filtering pipeline, this multi-omic approach yielded high-interest variants for provisional genetic diagnoses in six of 22 participants (**27%**). However, one variant was shared by two affected siblings in the study; thus, independent high-interest variants were identified in five of 20 unrelated individuals (**25%**). This strategy also identified uncertain but potentially disease-relevant candidate variants in three participants (**14%**); as two of these variants were shared by the father-son duo, independent uncertain variants were thus identified in 2 of 20 unrelated families (**10%**). Lastly, transcriptomic evidence consistent with disease-relevant gene expression patterns in two participants (**9%**). Importantly, LR-RNA-seq directly enabled high-interest candidate variant identification and interpretation in two of 22 participants (**9%**), thereby accounting for 33% of all solved diagnostic cases. Of the six solved cases, five occurred in pediatric or congenital cases, while one occurred in an adult-onset case. The main study outcomes are summarized in **Figure 5.1**.

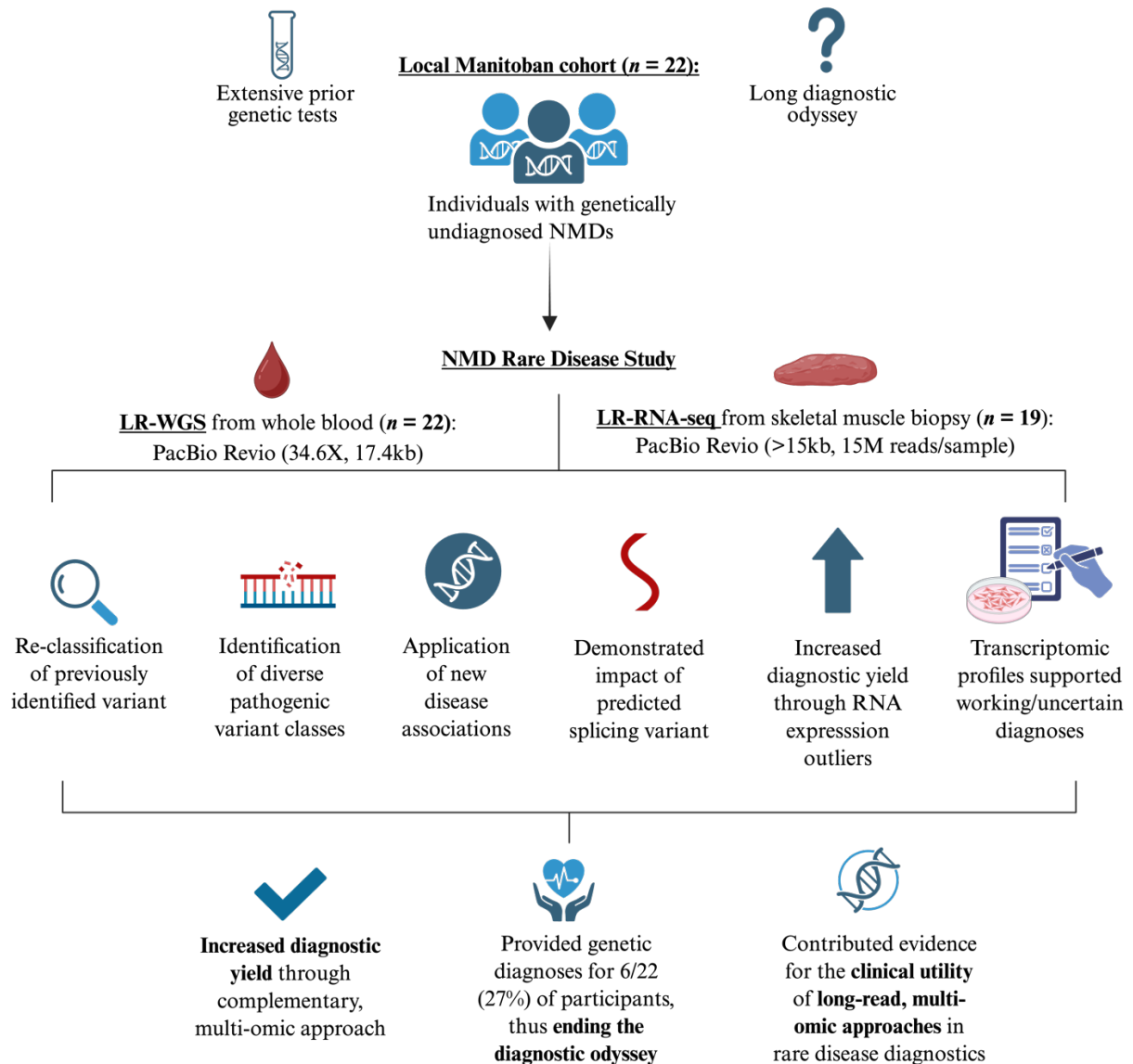


Figure 5.1. Integrated LR-WGS and LR-RNA-seq approach in rare disease cohort demonstrated clinical utility of this approach through improved diagnostic yield and identification and interpretation of multiple variant types, leading to a diagnostic rate of 6/22 (27%) of participants. A cohort of 22 individuals with NMDs were recruited for LR-WGS from 2017-2018; 19 of these individuals also had skeletal muscle biopsy available for LR-RNA-seq study. LR-WGS enabled identification of pathogenic variants from diverse variant classes, including small variants, SVs, CNVs and repeat expansions, as well as re-classification of a variant identified in previous SRS-based testing to pathogenic status. Further, LR-WGS as a comprehensive approach allowed for the application of new disease associations to result in genetic diagnosis for two participants. The complementary multi-omic approach enabled

increased identification of genes of interest through RNA expression outlier analysis (using IsoRanker), as well as improved interpretation of the impact of genomic variants. Lastly, LR-RNA-seq demonstrated transcriptomic expression profiles in two individuals which supported working or uncertain genomic diagnoses. Altogether, the comprehensive, LRS-based methodology of this study enabled genetic diagnosis for 6/22 (27%) of participants, or 5 individuals from 20 families (25%), thereby ending the diagnostic odyssey and providing evidence for these methods in broader clinical use.

These results of the present study were achieved in a cohort enriched for diagnostically challenging cases, as all participants had undergone extensive prior genetic investigations without a diagnosis (3.95 average prior genetic tests per participant, range: 1-10). This highlights the finding that important diagnostic information may be missed by conventional testing strategies, notably for the six individuals who did receive a genetic diagnosis through this study's approach. Furthermore, the previously unsolved nature of this cohort supports the strength of an integrated LR-WGS and LR-RNA-seq approach to capture variants in challenging diagnostic cases. The results therefore support the study's hypothesis that integrated LR genomic and transcriptomic analyses can improve genetic diagnosis rates by enabling comprehensive detection of variants across classes and genomic regions. Collectively, these pilot findings inform future implementation of LR and multi-omic approaches to increase the genetic diagnostic rate for both NMDs and other rare diseases.

Finally, effective data interpretation of comprehensive multi-omic data was guided by detailed phenotyping. Bioinformatic filtering of LR-WGS data used standardized phenotypic (HPO) terms to prioritize variants in genes with phenotype-gene associations relevant to the participant's clinical features (Robinson et al., 2008). This phenotype-driven approach is well-established to support automated variant filtration pipelines in other Mendelian disease cohorts, such as the 100,000 Genomes Project ([Cipriani et al., 2020](#); [Jacobsen et al., 2022](#); [Yuan et al.,](#)

2022). Detailed phenotyping was also vital after bioinformatic filtering to support manual variant interpretation and phenotype correlation. To facilitate efficient and consistent phenotype characterization, clinical data were systematically curated using a structured capture sheet. This approach enabled retrospective extraction of clinical features relevant to neuromuscular and inherited conditions. This was particularly valuable given the extensive clinical records accumulated during many participants' prolonged diagnostic odysseys, thereby providing a systematic framework to identify and prioritize phenotype information most relevant to genomic interpretation. More broadly, structured phenotyping frameworks support scalable integration into variant interpretation pipelines and improve diagnostic performance in rare disease genomics.

5.1. Long-read genomic sequencing enables comprehensive variant detection and interpretation

The use of long reads in LR-WGS, with a coverage of 30X and read lengths of 15-20 kb, demonstrated diagnostic utility by identifying diverse classes of genomic variation in a single assay. On average, LR-WGS identified approximately 5.01 million variants per participant prior to filtering, including 4.96 million small variants, 68 CNVs and 51,316 SVs. The value of this comprehensive approach is reflected in the diversity of high-interest and uncertain candidate variants identified in the present study, including four SNVs (*FLNC*, *NUP188*, *ORAI1*, *TTN*), one SV (*PRX*), two CNVs (*DMD*) and two repeat expansions (*ATXN8OS*). LR-WGS also successfully detected all small variants identified through participants' prior genetic testing, supporting its ability to perform comparably to SRS approaches in small variant identification. The only previously identified variant not detected by LR-WGS was a 31 kb deletion at chromosome 4q35 (Section 4.3.2). This likely reflects the challenges of accurately mapping reads within the highly repetitive subtelomeric D4Z4 locus; previous genetic methods succeeded in identifying this variant

as they directly assessed fragment size and did not rely on sequence alignment. Emerging resources and technologies may improve LRS performance in this and other challenging genomic regions, including the telomere-to-telomere (T2T) reference genome (Lu et al., 2024). Further, ongoing extension of read length shows improved read alignment and variant detection across such repetitive genomic regions, for example ultra-long read sequencing which can generate reads exceeding 100 kb in length (Lu et al., 2024). In sum, the comprehensive findings of variant identification by LR-WGS support the ability of this assay to assess more genetic conditions at once (Höps et al., 2025; Sen et al., 2025); this is particularly important in the context of NMDs, where substantial clinical and genetic heterogeneity requires a broad diagnostic approach (Barp et al., 2021; Benarroch et al., 2024).

Additionally, simultaneous detection of multiple variant classes facilitates assessment of compound heterozygosity in recessive disorders, thereby reducing the need for sequential follow-up testing (AlAbdi et al., 2023). For example, in previous SRS testing, four participants in this cohort had a heterozygous small variant identified in a gene associated with a recessive disorder. Due to the limited ability of SRS to detect SVs and CNVs, follow-up deletion-duplication testing had been required to reliably assess for a second pathogenic allele. In contrast, LR-WGS interrogates all variant classes in a single assay, enabling direct, confident exclusion of a second pathogenic allele and streamlining evaluation of heterozygous variants in genes associated with recessive conditions (AlAbdi et al., 2023; D. E. Miller et al., 2021). Additionally, the extended length of LR-WGS enables direct phasing of variants for the detection recessive disorders, either through direct phasing or indirect phasing across haplotype blocks (Zhao et al., 2025). This information is vital for NMDs, as NMD genes are large and complex, with frequent genetic variation contributing to compound heterozygous disease (Savarese et al., 2018). Although

segregation analysis in first-degree relatives has been a standard approach to phasing, this process is time-consuming, resource-intensive, and not feasible for individuals without multiple available family members for testing (Kernohan & Boycott, 2024). This strength of LRS was evidenced in our cohort, as LR-WGS in one participant directly established that two *NUP188* variants were in *trans*; this critical information confirmed concordance with the AR inheritance of the associated condition and supported a provisional genetic diagnosis. Collectively, these findings suggest that LR-WGS can optimize diagnostic workflows and improve variant interpretation, while providing comprehensive evaluation of the diverse genetic mechanisms underlying NMDs (AlAbdi et al., 2023; Bogaert et al., 2026). These findings also support growing evidence that LR-WGS may be more appropriately positioned as an early-tier or first-tier genomic, rather than a technology reserved for unresolved cases (Kamolvisit et al., 2025; Mortazavi et al., 2026). Caveats and challenges to implementation of a such approach are later discussed in **Sections 5.5 and 5.6**.

Several additional technical advantages of LR-WGS improved variant characterization and interpretation in this study. LR-WGS enables high-resolution delineation of variant structure and breakpoints, including within intronic or highly repetitive regions that may be poorly mapped using short reads (Logsdon et al., 2020; S. Zeng et al., 2019). In this cohort, LR-WGS successfully defined the intronic breakpoints of a large *DMD* CNV, thereby providing confidence in variant structure and mechanism, and supporting diagnosis of BMD due to the in-frame nature of this variant. Second, LR-WGS enhances the characterization of repeat expansion disorders by enabling direct assessment of both repeat length and internal sequence composition, each of which is critical for interpreting pathogenicity (Ebbert et al., 2018; Han et al., 2026). This was reflected in our cohort, where the detection of sequence interruptions within the expanded *FGF14* repeats allowed accurate interpretation as non-pathogenic in two individuals (**Section 3.2.3**).

5.2. WGS diagnostic power compared to WES

In the broader literature, WGS has demonstrated an increased diagnostic rate compared to WES, owing to the capture of both coding and non-coding variants, increased coverage over first exons, and improved detection of SVs and CNVs. In other cohorts, SR-WGS demonstrated an additional 7-8.2% diagnostic yield over SR-WES, and a substantial proportion of variants discovered on WGS required this approach for variant detection (Alfares, Aloraini, subaie, et al., 2018; Smedley et al., 2021; Wojcik et al., 2024). In this cohort, no additional pathogenic variants were uniquely identified by LR-WGS in non-coding regions. An intronic *TTN* splicing variant was characterized in this study, however it had been previously detected by SR-WES due to exon-padding strategies (**Section 4.1.4**). The absence of additional non-coding findings in this cohort may reflect the underrealized diagnostic potential of WGS, which is impacted by challenges in interpreting the functional consequences of non-coding variants (Hayeems et al., 2026; Lewis et al., 2025). Current approaches to interpreting non-coding variant pathogenicity are limited to evolutionary constraint and allele frequency, without widely used functional predictions of non-coding variant impact (Lewis et al., 2025). However, emerging computational tools may improve interpretation and filtering of non-coding variants. For example, PromoterAI predicts the regulatory impact of promoter variants on gene expression, while AlphaGenome provides genome-wide regulatory effect scores based on sequence-to-function predictions of chromatin access, transcription factor binding and gene expression effects (García-González & Gogolewski, 2026; Jaganathan et al., 2025). In addition, multi-omic approaches that directly assess downstream molecular consequences, such as gene expression and splicing outliers, further enable functional interpretation of non-coding variant effect (Yépez et al., 2026). In this study, this was demonstrated through LR-RNA-seq from disease-relevant skeletal muscle, which supported interpretation of a

TTN splicing variant through corresponding gene down-regulation and alternative splicing (**Section 4.1.4**). Collectively, these approaches highlight that the primary limitation in WGS-based diagnostics is not variant detection, but functional interpretation of non-coding effects. Lastly, these observations align with a recent influential study by Hayeems et al. (2026); the study did not find a significant diagnostic advantage of SR-WGS over SR-WES, but did note that the strength of SR-WGS was underrealized in part due to exclusion of non-coding variants which were not previously reported as disease-associated. Further, the study also identified that lack of advanced bioinformatic tools for variant annotation and filtering further restricted the identification of potentially diagnostic variants (Chen et al., 2016; Hayeems et al., 2026).

5.3. Comprehensive and reproducible methods support genetic re-analysis

The comprehensive nature of LR-WGS makes it an effective tool for iterative genetic re-analysis and variant interpretation as new evidence and tools emerge (Bamshad et al., 2019; Lionel et al., 2018). This includes more variant information (e.g., functional studies, new reference datasets, more affected individuals) or as new tools for interpretation can be applied (e.g., bioinformatic deleteriousness predictions) (Moon & Seo, 2025). In the present study, LR-WGS enabled reinterpretation of variants of interest identified in prior clinical testing, including reclassification of an *FLNC* missense variant from VUS to pathogenic following accumulation of supporting ClinVar evidence (**Section 4.1.3**). Re-interpretation of other VUS from past clinical testing to benign status was also significant, notably the *TTN* missense variants in two individuals reclassified as benign due to ACMG criterion BP4 (computational tools predicting a benign effect) and BP6 (benign reporting of variant by reputable sources) (**Sections 4.1.2 and 4.2.1**) (Richards et al., 2015).

Further, LR-WGS facilitates re-interpretation as new gene-disease associations emerge (Bamshad et al., 2019). For example, the lack of established disease associations for *ORAI1* and *NUP188* limited prior genetic testing in two individuals (**Sections 4.1.2 and 4.2.1**). Conversely, retention of comprehensive genome-wide data through LR-WGS would enable re-analysis as gene-disease links are established, potentially allowing unexplained cases to be resolved. In practice, systematic re-analysis would require infrastructure for secure long-term data storage, sufficient workforce capacity and funding and standardized guidelines for variant re-analysis, including regarding the use of automated pipelines to trigger manual review (Appelbaum et al., 2020; Best et al., 2024; Fehlberg et al., 2024; Welland et al., 2025). Genetic counsellors play a key role in this process, including assessing new data for variant interpretation, consenting patients for re-contact and communicating re-analyzed results (Appelbaum et al., 2020; Outram et al., 2024; Wain et al., 2020). As a contribution to ongoing efforts to improve variant interpretation and genetic diagnosis for other individuals, the new variant classifications made during this study, for example for the *NUP188* and *FLCN* variants, will also be submitted to ClinVar where appropriate.

5.4. LR-RNA-seq from disease-relevant tissue adds discovery and interpretive power to LR-WGS

Complementary LR-RNA-seq from disease-relevant skeletal muscle tissue enhanced the diagnostic utility of LR-WGS in this cohort by enabling tissue-specific assessment of transcriptional and splicing consequences. This is particularly important in light of growing evidence that disease mechanisms in NMDs are highly tissue-dependent, with strong expression of NMD genes in skeletal muscle and cardiac tissue compared with the broader genome (Martin et al., 2026). LR-RNA-seq supported diagnostic utility in this cohort through three primary mechanisms: gene discovery, functional variant interpretation, and disease characterization. First,

LR-RNA-seq increased discovery power of LR-WGS by identifying gene expression and/or splicing outliers using the IsoRanker framework (Cheng et al., 2025). This finding in our cohort highlights a key practical advantage of IsoRanker, that this tool can prioritize disease-relevant expression outliers for pathogenic variant interpretation even using a small sample size. This expands the accessibility and interpretability of LR-RNA-seq, as LRS-based reference datasets aren't always available. Using IsoRanker, LR-RNA-seq directly contributed to diagnosis in one participant (4% of the cohort) by identifying *NUP188* as a significantly down-regulated expression outlier compared to the cohort, thereby prioritizing the gene for the ultimate identification of biallelic pathogenic variants in LR-WGS data (**Section 4.1.5**). Importantly, *NUP188* is not a recognized NMD gene but is instead associated with a broader syndromic phenotype (Muir et al., 2020); thus, transcriptomic data enabled the expansion of variant analysis beyond predefined neuromuscular gene filters, using disease-relevant biological evidence. The diagnostic contribution of RNA-based outlier analyses in this study is comparable to that of previous studies, in which SR-RNA-seq outliers, complementing SR-WGS, directly contributed to diagnosis in 6-8% of rare disease cohorts (Yang et al., 2023a; Yépez et al., 2026).

Second, LR-RNA-seq complemented LR-WGS by strengthening the interpretation of functional variants. A homozygous *TTN* splice-region variant, previously classified as a VUS despite *in silico* predictions of splicing impact, was confirmed in this study as functionally disruptive through LR-RNA-seq (**Section 4.1.4**). As such, LR-RNA-seq provided functional evidence supporting the variant's pathogenicity, especially within the context of transcriptomic analysis in skeletal muscle, where *TTN* is highly expressed in physiologically relevant isoforms. LR-RNA-seq also supported interpretation of newly discovered variants, in which absence of the affected exons across the deleted region of a *DMD* CNV supported pathogenic interpretation

(Section 4.1.1). These findings illustrate the value of integrated multi-omics analysis to provide orthogonal evidence for pathogenicity. The present study is consistent with previous studies demonstrating the value of transcriptomics for resolving VUS and interrogating regions of the genome that remain difficult to interpret using DNA sequencing alone (Cummings et al., 2023). In summary, LR-RNA-seq from disease-relevant tissue can successfully complement genomic variant identification strategies by indicating functional impact of truly disease-causing variants, while improving diagnostic efficiency by reducing the number of likely benign variants for manual review.

Finally, LR-RNA-seq provided functional evidence relevant to disease characterization, with gene expression patterns from two individuals supporting diagnoses. In one case, an inflammatory transcriptomic signature in skeletal muscle supported the pathogenic interpretation of a mosaic D4Z4 region deletion associated with FSHD1 (Section 4.3.2). In another, LR-RNA-seq demonstrated expression patterns consistent with categorical diagnosis of mitochondrial myopathy even in the absence of clear genetic diagnosis (Section 4.3.1). These findings align with emerging multi-omic approaches for genetic diagnosis. For example, episignatures investigate large-scale DNA methylation pattern disruptions characteristic of affected samples for a given genetic or categorical diagnosis (Husson et al., 2024). Such functional information has been developed for NMDs such as DMD and myasthenia gravis, where genomic DNA methylation patterns in peripheral blood were associated with disease status (Lin et al., 2023; Schreyer et al., 2023). Importantly, LR-WGS can accurately call base modifications including DNA methylation, thus expanding the utility of this assay to include epigenetic analysis (Mortazavi et al., 2026).

5.5. Moving towards broader, clinical adoption of LR, multi-omics approaches

While the broader use of LR-WGS was previously limited by cost, advances in technology and the use of high-throughput methods have brought costs closer to those of SR-WGS and even SR-WES (Kernohan & Boycott, 2024; Yang et al., 2023). For example, the new LR-WGS PacBio method SPRQ-Nx is available at a cost of US\$345 per genome (PacBio, n.d.). Accuracy of LRS technologies has also advanced, with base-level accuracies comparable to those of SRS technologies (Espinosa et al., 2024); this was demonstrated in our study through the capture of almost all variants from prior SRS-based genetic testing. Alongside turn-around-times comparable to SRS methodologies, these factors now allow LR-WGS to be considered for use in the broader clinical setting (Hayeems et al., 2026). Indeed, a recent large-scale study took findings from LR-WGS in 1,000 diagnostic samples, and modelled the impact of the LR-WGS first approach on over 15,000 index patients, thereby determining that LR-WGS would result in a greater diagnostic yield than standard-of-care testing and supporting its use as a generic first-tier diagnostic test (Bitter et al., 2026.).

However, several other considerations remain which may impact implementation of this new technology and the comprehensive genomic data it generates. Importantly, the vast amount of information generated by LR-WGS requires innovative solutions to implement this technology at a clinical scale (Lionel et al., 2018). As demonstrated in our study, the integration of phenotypic information, advanced bioinformatic tools, and functional transcriptomic information can support effective, large-scale variant filtration. However, while many distinct tools exist for annotating LR-WGS, integrating this diverse information requires substantial manual effort (Tesi et al., 2024). The development of automated bioinformatic pipelines, machine learning and artificial intelligence (AI) tools for effective analysis of raw data is demonstrated (De La Vega et al., 2021;

Meng et al., 2023; Seo et al., 2022). The clinical utility, as well as the ethical and legal risks of such tools, are important considerations in clinical adoption (Adams et al., 2024). In moving towards clinical implementation, the co-development of comprehensive genomic tools and related guidelines with patient partners must also be considered to ensure that new infrastructure is acceptable and responsive to patient needs (Hartley et al., 2023).

Several factors also impact the broader use of LR-RNA-seq in genetic diagnostic workflow. Firstly, the elevated cost of LR-RNA-seq presents a barrier to more wide-spread usage, in part due to the complexity of steps required to construct libraries for bulk transcriptome profiling (Choi et al., 2024). However, while the current costs of LR-RNA-seq at 15 million reads/sample scale range from CA\$425 to CA\$2,250, prices should decrease with time following the advancement of high-throughput sequencing techniques, as happened with genomic sequencing technologies (Centre for Applied Genomics (TCAG), 2025). Secondly, disease-relevant tissue is not always available or easily accessible for all Mendelian diseases. For example, brain tissue for neurodevelopmental disorders is inaccessible and not all individuals with NMDs have a muscle biopsy performed, especially as genetic studies have taken an increasing role in the diagnostic pathway (Karunakaran et al., 2025). However, new methods have shown that reprogramming of fibroblasts into myogenic cells is an effective way to recapitulate the muscle transcriptome, thereby allowing functional assessment of variants that may not have been detectable in blood-derived DNA (Gonorazky et al., 2019). As IsoRanker can achieve adequate pilot results from a small sample size, this technology can be readily used from stem-cell-based models, including 2D models (e.g., *Ng2*-induced neurons) and 3D organoids (e.g., unpatterned cerebral organoids and patterned neural organoids and assembloids), where disease-relevant tissue is not readily available (Cheng et al., 2025). In addition, broader use of transcriptomics can be supported through workforce training in

interpreting multiple data sources, including genetic variants, transcriptomic information, and other omics technologies (Hasin et al., 2017). Lastly, there is currently a lack of LRS-derived datasets from healthy reference populations to support robust interpretation of outliers in LR-RNA-seq. However, a key finding of this study is that IsoRanker can effectively prioritise disease-relevant expression of outliers even without large, fully matched reference datasets.

5.6. Limitations

The strength of LRS approaches is currently constrained by the availability of appropriate tools and complementary datasets for variant filtration and interpretation. Most widely used population resources, such as gnomAD v4.0, are currently derived primarily from SRS and therefore incompletely represent SVs, CNVs and STRs. The lack of comprehensive population-scale data for complex variant classes reduces the accuracy of frequency-based filtering and limits interpretation of LRS-identified variants. Given the lack of large-scale, technology-matched reference datasets, LRS data from GA4K were used as a supplement. GA4K was the only LRS dataset available to us for this study and comprises approximately 500 individuals of European ancestry, thereby highlighting the need for additional LRS-based reference datasets to support the interpretive power of LRS assays (Pellecchia, 2024b). Moreover, current reference datasets are disproportionately enriched for European ancestry; this limits equitable variant interpretation for individuals of other ancestries, which ultimately leads to clinical disparities. Our cohort was comprised of 90% of individuals (19/21 reported ancestry) with some degree of European ancestry; consequently, cohorts with different ancestral backgrounds may achieve different diagnostic outcomes with LR-WGS using available reference datasets. Several large-scale initiatives are underway to improve the diversity of genomic reference resources, which may improve equity and accuracy of variant interpretation across populations. Lastly, while not LRS-specific, there are

fewer prediction tools for interpreting SVs and CNVs than for small variants. Thus, although LR-WGS technology is advanced, the lack of representative datasets limits the full diagnostic capacity of this resource. As more sequencing and functional studies are performed using LRS methods, which are increasingly affordable and accessible, we are hopeful that these limitations will decrease.

The pilot scope of this study also presents limitations. The sample size of this cohort is small, limiting the generalizability of LR-WGS with LR-RNA-seq to larger populations. However, as a pilot study assessing the use of these approaches in individuals with NMDs, our results are encouraging towards larger implementation of the technology, especially in the context of avoiding sequential testing. Further, the scope of this study also limited the variant filtration and review strategy; due to the vast number of variants generated by LR-WGS, a stringent filtering pipeline was required to reduce the number of variants to a manageable amount for manual review. This included only assessing variants in known NMD genes and did not allow for novel disease associations to be made. Due to the paucity of dedicated bioinformatic prediction tools to support interpretation and filtering of SVs, the filtering pipeline assessed only SVs in coding regions, potentially missing non-coding SVs that affect gene expression and through other mechanisms (Ferraro et al., 2020). Lastly, a strict maximum variant allele frequency of $< 0.1\%$ across all populations in gnomAD and GA4K (LRS-based cohort of ~ 500 individuals of European ancestry) was used to filter small variants; thus, pathogenic recessive variants with frequencies above this threshold may be missed. Moreover, this study used mostly singleton LR-WGS studies, as well as two duos, due to the elevated cost of LR-WGS at the time of recruitment and consenting (2017-2018). Performance of LR-WGS studies with trio studies (i.e., inclusion of parental samples for comparison) may improve pathogenic variant identification (Smedley et al., 2021). Altogether,

while the filtering strategy was devised with feasibility in mind for a pilot study and in line with established variant-filtration guidelines, relaxing the filtration strategy and dedicated efforts to uncover new gene-disease associations may be necessary for unresolved individuals.

Our multi-omic LR-RNA-seq approach also presented limitations, with opportunities for additional future analyses. Firstly, as previously mentioned, the lack of large-scale healthy population datasets using LR-RNA-seq may limit power to detect more subtle outliers. Second, the depth of LR-RNA-seq used in this study (15 million reads per sample) is improved compared to earlier transcriptomic datasets, for example PacBio Sequel II System with 4-5 million reads per sample (Underwood et al., 2020). However, read depth remains somewhat shallow compared to 40 to 100 million reads per sample used in comparable rare disease studies using SR-RNA-seq (Cummings et al., 2017; Gonorazky et al., 2019; Hong et al., 2022); improved read depth may improve detection of rare isoforms (Pardo-Palacios et al., 2024; Patterson et al., 2019). Thirdly, transcriptomics in this study was performed using bulk RNA-seq, which can lose patterns in cellular heterogeneity by blending cell populations during RNA isolation (Chapman & Lassmann, 2025). Single-cell RNA-seq has shown promise in capturing gene expression profiles of individual cells, thereby identifying cell-type-specific expression patterns and rare cell populations that may reveal additional diagnostic information (Chapman & Lassmann, 2025). Spatial omics may provide additional biological insight by preserving information about the location of gene expression patterns within the tissue, thereby relating gene expression to specific cellular environments and regions of pathology (Ren et al., 2024). Lastly, additional analyses are possible using IsoRanker analysis with LR-RNA-seq data, including allelic imbalance, which identifies unequal expression between alleles suggestive of regulatory or functional effects, and nonsense-

mediated decay after cycloheximide treatment, which can identify transcripts undergoing degradation as a consequence of truncating variants (Cheng et al., 2025).

5.7. Future Directions

Given the success of LR-WGS and complementary LR-RNA-seq from disease-relevant tissue in this cohort, future studies could expand these approaches to larger and more cohorts to improve statistical power, and support clinical validation and implementation of comprehensive, LRS technologies. These approaches could be applied across a broader range of rare Mendelian diseases to assess generalisability of the techniques and maximize diagnostic yield for other patient populations. In the Canadian context, integration with initiatives such as the Pan-Canadian Genome Library will support the development of population-scale reference resources and facilitate more equitable access to advanced genomic diagnostics across diverse populations (Loewens, 2023). Further, the Silent Genomes Project aims to improve genetic healthcare for First Nations, Inuit and Métis peoples through safe, Indigenous-led governance of genomic data and the development of an Indigenous Background Variant Library to improve variant interpretation (*Silent Genomes Project*, n.d.)

This study demonstrated the power of a single multi-omic strategy for the genetic diagnosis of rare diseases. A broad range of other multi-omics technologies and workflows continue to emerge, and research will have to pursue an understanding of the comparative diagnostic utility of these new approaches. For example, while LR-RNA-seq is expected to improve detection of full-length isoforms and complex alternative splicing events, direct comparisons between LR- and SR-RNA-seq in disease-relevant tissues have not yet been performed in the context of NMDs. Importantly, paired LR- and SR-RNA-seq has been generated by this research group, which will be used to directly compared the two transcriptomic approaches, and evaluate the distinct or

incremental diagnostic contribution of each modality. Moreover, this research will review the benefits of a combined LR- and SR-RNA-seq approach, which could leverage the strengths of both technologies to improve variant identification and interpretation. Additional omics modalities not covered in the present study are also emerging, including proteomics, which identifies post-translational modifications and pathway activation at the protein level, and metabolomics, which measures small-molecule metabolites to show metabolic changes and biochemical dysfunction signatures. Research of the benefits and applications of these distinct multi-omic technologies to the NMD field will be important for enhancing diagnostic pathways for NMDs and other rare diseases (Smirnov et al., 2023).

LR-WGS data from this cohort can be interrogated using additional analytical strategies beyond those employed in this study, potentially yielding genetic diagnoses in unresolved cases. As previously mentioned, LR-WGS datasets inherently capture epigenetic information, notably DNA methylation patterns (Hildonen et al., 2026). These reflect chemical modifications to DNA that regulate gene expression without altering the underlying sequence, and can be used in various diseases to enable diagnosis, prognosis or monitoring of disease (Wisman et al., 2024). However, the analysis of epigenetic data is currently limited by a lack of established analytical tools for epigenetic profiling. Thus, the development of tools to identify epigenetic outliers in LR-WGS represents an important future direction to enable more comprehensive use of LR-WGS data. This research group aims to use available LR-WGS-derived methylation data from the cohort to contribute towards tool development and application of these assays towards NMDs.

5.8. Conclusions

Integrated LR-WGS and LR-RNA-seq enhanced diagnostic yield in a cohort of previously unsolved NMDs by identifying pathogenic variants, supporting variant interpretation, and

revealing disease-relevant molecular signatures. The identification of high-interest variants in over one-quarter of this heavily pre-tested cohort demonstrates the value of LR multi-omic approaches for addressing persistent diagnostic challenges in neuromuscular genetics. Collectively, these findings demonstrate that LRS-based multi-omics can comprehensively and efficiently capture pathogenic variation, compared to standard stepwise diagnostic workflows. This supports use of the above methodology in broader, clinical settings and supports emerging evidence for LR-WGS to be used as an early-tier approach for streamlined, single-assay genomic evaluation.

Importantly, improved diagnostic yield has direct implications for individuals affected with NMDs. Recent work highlights the growing role of genetic diagnosis in guiding clinical management, particularly in the context of emerging disease-modifying therapies, clinical trials and precision medicine, which can prevent irreversible disease progression (Adams et al., 2024; Smedley et al., 2021). Within the field of genetic counselling, earlier and more definitive diagnosis enables more targeted risk assessment, reproductive planning, anticipatory guidance and psychosocial support (Efthymiou et al., 2016; Smedley et al., 2021). Moreover, timely diagnosis can reduce the psychosocial burden associated with a prolonged diagnostic odyssey for individuals and families affected by rare disease (Lionel et al., 2018). As the cost of LR-WGS continues to decline and complementary analytical tools and reference resources expand, its clinical applicability is expected to increase. Overall, this study supports integrating LR-WGS and complementary omics into clinical pathways for diagnosing NMDs and other genetically heterogeneous rare diseases.

Data availability statement

Participant-level genomic and transcriptomic data are not publicly available as they contain potentially identifiable human genetic information, and are stored on a secure server at the

University of Manitoba. Non-identifiable code developed for this study will be made publicly available through GitHub upon publication. This repository will include reproducible scripts for concurrent LR-WGS variant filtering using the study-specific filtering pipeline, as well as scripts for IsoRanker data processing and IsoRanker-guided LR-WGS variant filtering.

References

- About Tayoun, A. N., Pesaran, T., DiStefano, M. T., Oza, A., Rehm, H. L., Biesecker, L. G., & Harrison, S. M. (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Human Mutation*, 39(11), 1517–1524. <https://doi.org/10.1002/humu.23626>
- ACMG Board of Directors. (2015). Clinical utility of genetic and genomic services: A position statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 17(6), 505–507. <https://doi.org/10.1038/gim.2015.41>
- Adams, D. R., van Karnebeek, C. D. M., Agulló, S. B., Faundes, V., Jamuar, S. S., Lynch, S. A., Pintos-Morell, G., Puri, R. D., Shai, R., Steward, C. A., Tumiene, B., & Verloes, A. (2024). Addressing diagnostic gaps and priorities of the global rare diseases community: Recommendations from the IRDiRC diagnostics scientific committee. *European Journal of Medical Genetics*, 70, 104951. <https://doi.org/10.1016/j.ejmg.2024.104951>
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), 248–249. <https://doi.org/10.1038/nmeth0410-248>
- Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.]*, 0 7, Unit7.20. <https://doi.org/10.1002/0471142905.hg0720s76>
- Agilent Technologies. (2022). *Agilent 4200 TapeStation System System Manual*. https://www.agilent.com/cs/library/usermanuals/public/4200-TapeStation_SystemManual.pdf?srsId=AfmBOoqOUonERPzKMj9L5AeFXti2-FnFOKDgqWqneIG0EnYFpxEaPfg1
- AlAbdi, L., Shamseldin, H. E., Khouj, E., Helaby, R., Aljamal, B., Alqahtani, M., Almulhim, A., Hamid, H., Hashem, M. O., Abdulwahab, F., Abouyousef, O., Jaafar, A., Alshidi, T., Al-Owain, M.,

- Alhashem, A., Al Tala, S., Khan, A. O., Mardawi, E., Alkuraya, H., ... Alkuraya, F. S. (2023). Beyond the exome: Utility of long-read whole genome sequencing in exome-negative autosomal recessive diseases. *Genome Medicine*, 15, 114. <https://doi.org/10.1186/s13073-023-01270-8>
- Aleem, T., Rashid, M., Ahmad, N., Asif, M. F., Tariq, M., Malik, N. A., & Poulter, J. A. (2025). Exome sequencing reveals broad genetic heterogeneity for neuromuscular disorders in consanguineous Pakistani Families. *European Journal of Human Genetics*, 1–7. <https://doi.org/10.1038/s41431-025-01915-9>
- Alfares, A., Aloraini, T., Subaie, L. A., Alissa, A., Qudsi, A. A., Alahmad, A., Mutairi, F. A., Alswaid, A., Alothaim, A., Eyaid, W., Albalwi, M., Alturki, S., & Alfadhel, M. (2018). Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 20(11), 1328–1333. <https://doi.org/10.1038/gim.2018.41>
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21, 30. <https://doi.org/10.1186/s13059-020-1935-5>
- Amberger, J., Bocchini, C., Schiettecatte, F., Scott, A., & 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). <https://doi.org/10.1093/nar/gku1205>
- 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>
- Anna, A., & Monika, G. (2018). Splicing mutations in human genetic disorders: Examples, detection, and confirmation. *Journal of Applied Genetics*, 59(3), 253–268. <https://doi.org/10.1007/s13353-018-0444-7>

- Antoniadi, T., Buxton, C., Dennis, G., Forrester, N., Smith, D., Lunt, P., & Burton-Jones, S. (2015). Application of targeted multi-gene panel testing for the diagnosis of inherited peripheral neuropathy provides a high diagnostic yield with unexpected phenotype-genotype variability. *BMC Medical Genetics*, 16, 84. <https://doi.org/10.1186/s12881-015-0224-8>
- Appelbaum, P. S., Parens, E., Berger, S. M., Chung, W. K., & Burke, W. (2020). Is there a duty to reinterpret genetic data? The ethical dimensions. *Genetics in Medicine*, 22(3), 633–639. <https://doi.org/10.1038/s41436-019-0679-7>
- Arahata, K., Ishihara, T., Fukunaga, H., Orimo, S., Lee, J. H., Goto, K., & Nonaka, I. (1995). Inflammatory response in facioscapulohumeral muscular dystrophy (FSHD): Immunocytochemical and genetic analyses. *Muscle & Nerve. Supplement*, 2, S56-66.
- Aromolaran, K. A., Benzow, K. A., Cribbs, L., Koob, M. D., & Piedras-Renteria, E. S. (2009). Kelch-like 1 protein up-regulates T-type currents by an actin-F dependent increase in $\alpha 1H$ channels via the recycling endosome. *Channels*, 3(6), 402–412. <https://doi.org/10.4161/chan.3.6.9858>
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., ... National Eye Institute, N. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Baets, J., Deconinck, T., De Vriendt, E., Zimoń, M., Yperzeele, L., Van Hoorenbeeck, K., Peeters, K., Spiegel, R., Parman, Y., Ceulemans, B., Van Bogaert, P., Pou-Serradell, A., Bernert, G., Dinopoulos, A., Auer-Grumbach, M., Sallinen, S.-L., Fabrizi, G. M., Pauly, F., Van den Bergh, P., ... De Jonghe, P. (2011). Genetic spectrum of hereditary neuropathies with onset in the first year of life. *Brain*, 134(9), 2664–2676. <https://doi.org/10.1093/brain/awr184>
- Bagger, F. O., Borgwardt, L., Jespersen, A. S., Hansen, A. R., Bertelsen, B., Kodama, M., & Nielsen, F. C. (2024). Whole genome sequencing in clinical practice. *BMC Medical Genomics*, 17(1), 39. <https://doi.org/10.1186/s12920-024-01795-w>

- Bamshad, M. J., Nickerson, D. A., & Chong, J. X. (2019). Mendelian Gene Discovery: Fast and Furious with No End in Sight. *American Journal of Human Genetics*, 105(3), 448–455.
<https://doi.org/10.1016/j.ajhg.2019.07.011>
- Barbosa-Gouveia, S., Vázquez-Mosquera, M. E., González-Vioque, E., Hermida-Ameijeiras, Á., Sánchez-Pintos, P., de Castro, M. J., León, S. R., Gil-Fournier, B., Domínguez-González, C., Camacho Salas, A., Negrão, L., Fineza, I., Laranjeira, F., & Couce, M. L. (2022). Rapid Molecular Diagnosis of Genetically Inherited Neuromuscular Disorders Using Next-Generation Sequencing Technologies. *Journal of Clinical Medicine*, 11(10), 2750.
<https://doi.org/10.3390/jcm11102750>
- Barp, A., Mosca, L., & Sansone, V. A. (2021). Facilitations and Hurdles of Genetic Testing in Neuromuscular Disorders. *Diagnostics*, 11(4), Article 4.
<https://doi.org/10.3390/diagnostics11040701>
- Baskin, B., Banwell, B., Khater, R. A., Hawkins, C., & Ray, P. N. (2009). Becker muscular dystrophy caused by an intronic mutation reducing the efficiency of the splice donor site of intron 26 of the dystrophin gene. *Neuromuscular Disorders*, 19(3), 189–192.
<https://doi.org/10.1016/j.nmd.2008.11.003>
- Bchetnia, M., Bouchard, L., Mathieu, J., Campeau, P. M., Morin, C., Brisson, D., Laberge, A.-M., Vézina, H., Gaudet, D., & Laprise, C. (2021). Genetic burden linked to founder effects in Saguenay–Lac-Saint-Jean illustrates the importance of genetic screening test availability. *Journal of Medical Genetics*, 58(10), 653–665. <https://doi.org/10.1136/jmedgenet-2021-107809>
- Bean, L., Funke, B., Carlston, C. M., Gannon, J. L., Kantarci, S., Krock, B. L., Zhang, S., & Bayrak-Toydemir, P. (2020). Diagnostic gene sequencing panels: From design to report—a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 22(3), 453–461. <https://doi.org/10.1038/s41436-019-0666-z>

- Beck, M., & Hurt, E. (2017). The nuclear pore complex: Understanding its function through structural insight. *Nature Reviews. Molecular Cell Biology*, 18(2), 73–89.
<https://doi.org/10.1038/nrm.2016.147>
- Beecroft, S. J., Yau, K. S., Allcock, R. J. N., Mina, K., Gooding, R., Faiz, F., Atkinson, V. J., Wise, C., Sivadurai, P., Trajanoski, D., Kresoje, N., Ong, R., Duff, R. M., Cabrera-Serrano, M., Nowak, K. J., Pachter, N., Ravenscroft, G., Lamont, P. J., Davis, M. R., & Laing, N. G. (2020). Targeted gene panel use in 2249 neuromuscular patients: The Australasian referral center experience. *Annals of Clinical and Translational Neurology*, 7(3), 353–362. <https://doi.org/10.1002/acn3.51002>
- Begay, R. L., Tharp, C. A., Martin, A., Graw, S. L., Sinagra, G., Miani, D., Sweet, M. E., Slavov, D. B., Stafford, N., Zeller, M. J., Alnefaie, R., Rowland, T. J., Brun, F., Jones, K. L., Gowan, K., Mestroni, L., Garrity, D. M., & Taylor, M. R. G. (2016). FLNC Gene Splice Mutations Cause Dilated Cardiomyopathy. *JACC. Basic to Translational Science*, 1(5), 344–359.
<https://doi.org/10.1016/j.jacbts.2016.05.004>
- Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., Antipenko, A., Shang, L., Boisson, B., Casanova, J.-L., & Abel, L. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences of the United States of America*, 112(17), 5473–5478.
<https://doi.org/10.1073/pnas.1418631112>
- 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>
- Benito-Lozano, J., Arias-Merino, G., Gómez-Martínez, M., Ancochea-Díaz, A., Aparicio-García, A., Posada de la Paz, M., & Alonso-Ferreira, V. (2022). Diagnostic Process in Rare Diseases: Determinants Associated with Diagnostic Delay. *International Journal of Environmental Research and Public Health*, 19(11), 6456. <https://doi.org/10.3390/ijerph19116456>

- Berg, J. S., Khoury, M. J., & Evans, J. P. (2011). Deploying whole genome sequencing in clinical practice and public health: Meeting the challenge one bin at a time. *Genetics in Medicine*, 13(6), 499–504. <https://doi.org/10.1097/GIM.0b013e318220aaba>
- Best, S., Fehlberg, Z., Richards, C., Quinn, M. C. J., Lunke, S., Spurdle, A. B., Kassahn, K. S., Patel, C., Vears, D. F., Goranitis, I., Lynch, F., Robertson, A., Tudini, E., Christodoulou, J., Scott, H., McGaughan, J., & Stark, Z. (2024). Reanalysis of genomic data in rare disease: Current practice and attitudes among Australian clinical and laboratory genetics services. *European Journal of Human Genetics*, 32(11), 1428–1435. <https://doi.org/10.1038/s41431-024-01633-8>
- Bhatt, J. M. (2016). The Epidemiology of Neuromuscular Diseases. *Neurologic Clinics*, 34(4), 999–1021. <https://doi.org/10.1016/j.ncl.2016.06.017>
- Bick, D., Jones, M., Taylor, S. L., Taft, R. J., & Belmont, J. (2019). Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases. *Journal of Medical Genetics*, 56(12), 783–791. <https://doi.org/10.1136/jmedgenet-2019-106111>
- Biggar, W. D., Harris, V. A., Eliasoph, L., & Alman, B. (2006). Long-term benefits of deflazacort treatment for boys with Duchenne muscular dystrophy in their second decade. *Neuromuscular Disorders: NMD*, 16(4), 249–255. <https://doi.org/10.1016/j.nmd.2006.01.010>
- Bitter, T. J. J. de, Sanden, B. van der, Sagath, L., Höps, W., Arts, P., Groot, M. de, Weiss, M. M., Derks, R., Ouden, A. den, Heuvel, S. van den, Timmermans, R. G. J., Leeuwen, T. van, Galbany, J. C., Smits, J. G. A., Blok, L. S., Hofste, T., Steehouwer, M., Zomer, N., Sabbagh, Q., ... Vissers, L. E. L. M. (n.d.). Clinical Long-Read Genome Sequencing for Rare-Disease Diagnostics. *New England Journal of Medicine*, 0(0). <https://doi.org/10.1056/NEJMc2602512>
- 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., ... Lochmüller, H. (2015b). 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>

- Boerkoel, C. F., Takashima, H., Stankiewicz, P., Garcia, C. A., Leber, S. M., Rhee-Morris, L., & Lupski, J. R. (2001). Periaxin Mutations Cause Recessive Dejerine-Sottas Neuropathy. *American Journal of Human Genetics*, 68(2), 325–333. <https://doi.org/10.1086/318208>
- Bogaert, K. V. D., Geysens, M., Huremagic, B., Souche, E., Peeters, G., Olszewska, N., Meynants, S., Mancuso, G., Breckpot, J., Denayer, E., Devriendt, K., Peeters, H., Menten, B., Mortier, G., Vogels, A., Esch, H. V., & Vermeesch, J. (2026). *Closing the diagnostic gap in developmental disorders through comprehensive long-read sequencing*. In Review.
<https://doi.org/10.21203/rs.3.rs-9108777/v1>
- Bonne, G., Rivier, F., & Hamroun, D. (2026, May 18). *GeneTable of Neuromuscular Disorders*.
<https://www.musclegenetable.fr/index.html>
- Borràs, D. M., Vossen, R. H. A. M., Liem, M., Buermans, H. P. J., Dauwerse, H., van Heusden, D., Gansevoort, R. T., den Dunnen, J. T., Janssen, B., Peters, D. J. M., Losekoot, M., & Anvar, S. Y. (2017). Detecting PKD1 variants in polycystic kidney disease patients by single-molecule long-read sequencing. *Human Mutation*, 38(7), 870–879. <https://doi.org/10.1002/humu.23223>
- Bos, I., Kuks, J. B. M., & Wynia, K. (2015). Development and testing psychometric properties of an ICF-based health measure: The Neuromuscular Disease Impact Profile. *Journal of Rehabilitation Medicine*, 47(5), 445–453. <https://doi.org/10.2340/16501977-1938>
- Bos, I., Wynia, K., Almansa, J., Drost, G., Kremer, B., & Kuks, J. (2019). The prevalence and severity of disease-related disabilities and their impact on quality of life in neuromuscular diseases. *Disability and Rehabilitation*, 41(14), 1676–1681.
<https://doi.org/10.1080/09638288.2018.1446188>
- Boyce, F. M., Beggs, A. H., Feener, C., & Kunkel, L. M. (1991). Dystrophin is transcribed in brain from a distant upstream promoter. *Proceedings of the National Academy of Sciences of the United States of America*, 88(4), 1276–1280. <https://doi.org/10.1073/pnas.88.4.1276>
- Brodehl, A., Ferrier, R. A., Hamilton, S. J., Greenway, S. C., Brundler, M.-A., Yu, W., Gibson, W. T., McKinnon, M. L., McGillivray, B., Alvarez, N., Giuffre, M., Schwartzentruber, J., FORGE

- Canada Consortium, & Gerull, B. (2016). Mutations in FLNC are Associated with Familial Restrictive Cardiomyopathy. *Human Mutation*, 37(3), 269–279.
<https://doi.org/10.1002/humu.22942>
- 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., Maarel, S. M. van der, Laine, R., Ghosh, P. S., ... Kang, P. B. (2022). Diagnostic capabilities of nanopore long-read sequencing in muscular dystrophy. *Annals of Clinical and Translational Neurology*, 9(8), 1302. <https://doi.org/10.1002/acn3.51612>
- Brun, F., Gigli, M., Graw, S. L., Judge, D. P., Merlo, M., Murray, B., Calkins, H., Sinagra, G., Taylor, M. R., Mestroni, L., & James, C. A. (2020). FLNC truncations cause arrhythmogenic right ventricular cardiomyopathy. *Journal of Medical Genetics*, 57(4), 254–257.
<https://doi.org/10.1136/jmedgenet-2019-106394>
- Bushby, K., Finkel, R., Birnkrant, D. J., Case, L. E., Clemens, P. R., Cripe, L., Kaul, A., Kinnett, K., McDonald, C., Pandya, S., Poysky, J., Shapiro, F., Tomezsko, J., & Constantin, C. (2010). Diagnosis and management of Duchenne muscular dystrophy, part 1: Diagnosis, and pharmacological and psychosocial management. *The Lancet Neurology*, 9(1), 77–93.
[https://doi.org/10.1016/S1474-4422\(09\)70271-6](https://doi.org/10.1016/S1474-4422(09)70271-6)
- Canada, P. H. A. of. (2018, April 12). *Economic Burden of Illness in Canada, 2010* [Research].
<https://www.canada.ca/en/public-health/services/publications/science-research-data/economic-burden-illness-canada-2010.html>
- Cancer Genome Atlas Research Network. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, 513(7517), 202–209. <https://doi.org/10.1038/nature13480>
- Cao, Y., Tokita, M. J., Chen, E. S., Ghosh, R., Chen, T., Feng, Y., Gorman, E., Gibellini, F., Ward, P. A., Braxton, A., Wang, X., Meng, L., Xiao, R., Bi, W., Xia, F., Eng, C. M., Yang, Y., Gambin, T., Shaw, C., ... Stankiewicz, P. (2019). A clinical survey of mosaic single nucleotide variants in

- disease-causing genes detected by exome sequencing. *Genome Medicine*, 11(1), 48.
<https://doi.org/10.1186/s13073-019-0658-2>
- Carmichael, N., Tsipis, J., Windmueller, G., Mandel, L., & Estrella, E. (2015). “Is it Going to Hurt?”: The Impact of the Diagnostic Odyssey on Children and Their Families. *Journal of Genetic Counseling*, 24(2), 325–335. <https://doi.org/10.1007/s10897-014-9773-9>
- Caron, N. R., Chongo, M., Hudson, M., Arbour, L., Wasserman, W. W., Robertson, S., Correard, S., & Wilcox, P. (2020). Indigenous Genomic Databases: Pragmatic Considerations and Cultural Contexts. *Frontiers in Public Health*, 8, 111. <https://doi.org/10.3389/fpubh.2020.00111>
- Carvalho, C. M. B., & Lupski, J. R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nature Reviews. Genetics*, 17(4), 224–238.
<https://doi.org/10.1038/nrg.2015.25>
- Centre for Applied Genomics (TCAG). (2025, April). *TCAG NGS price list*. The Hospital for Sick Children. <https://www.tcag.ca/documents/facilities/TCAG-NGS-Price-List%20April%202025.pdf>
- Chae, J. H., Vasta, V., Cho, A., Lim, B. C., Zhang, Q., Eun, S. H., & Hahn, S. H. (2015). Utility of next generation sequencing in genetic diagnosis of early onset neuromuscular disorders. *Journal of Medical Genetics*, 52(3), 208–216. <https://doi.org/10.1136/jmedgenet-2014-102819>
- Chaisson, M. J. P., Huddleston, J., Dennis, M. Y., Sudmant, P. H., Malig, M., Hormozdiari, F., Antonacci, F., Surti, U., Sandstrom, R., Boitano, M., Landolin, J. M., Stamatoyannopoulos, J. A., Hunkapiller, M. W., Korlach, J., & Eichler, E. E. (2015). Resolving the complexity of the human genome using single-molecule sequencing. *Nature*, 517(7536), 608–611.
<https://doi.org/10.1038/nature13907>
- Chang, R. S.-K., & Wong, Y. K. (2018). Prognostic indicators of neuromuscular disorders for palliative care referral. *Annals of Palliative Medicine*, 7(3), 33538–33338.
<https://doi.org/10.21037/apm.2017.09.02>
- Chapman, T., & Lassmann, T. (2025). Single-cell data combined with phenotypes improves variant interpretation. *BMC Genomics*, 26, 540. <https://doi.org/10.1186/s12864-025-11711-w>

- Charles A Janeway, J., Travers, P., Walport, M., & Shlomchik, M. J. (2001). The distribution and functions of immunoglobulin isotypes. In *Immunobiology: The Immune System in Health and Disease. 5th edition*. Garland Science. <https://www.ncbi.nlm.nih.gov/books/NBK27162/>
- Chaudhuri, A., & Behan, P. O. (2004). Fatigue in neurological disorders. *Lancet (London, England)*, 363(9413), 978–988. [https://doi.org/10.1016/S0140-6736\(04\)15794-2](https://doi.org/10.1016/S0140-6736(04)15794-2)
- Chen, S., Francioli, L. C., Goodrich, J. K., Collins, R. L., Kanai, M., Wang, Q., Alföldi, J., Watts, N. A., Vittal, C., Gauthier, L. D., Poterba, T., Wilson, M. W., Tarasova, Y., Phu, W., Grant, R., Yohannes, M. T., Koenig, Z., Farjoun, Y., Banks, E., ... Karczewski, K. J. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. *Nature*, 625(7993), 92–100. <https://doi.org/10.1038/s41586-023-06045-0>
- Chen, X., Harting, J., Farrow, E., Thiffault, I., Kasperaviciute, D., Hoischen, A., Gilissen, C., Pastinen, T., & Eberle, M. A. (2023). Comprehensive SMN1 and SMN2 profiling for spinal muscular atrophy analysis using long-read PacBio HiFi sequencing. *American Journal of Human Genetics*, 110(2), 240–250. <https://doi.org/10.1016/j.ajhg.2023.01.001>
- Chen, X., Kronenberg, Z., Baker, D., Devaney, J., Noya, J., Berlyoung, A., Yusuff, S., Lynch, S., Brandon, R., Hruska, K., Lochovsky, L., Lindy, A., Dolzhenko, E., & Eberle, M. (2025). O26: Resolution of large complex VNTRs with HiFi sequencing: Applications in LPA Kringle IV-type 2 repeat and D4Z4 repeat. *Genetics in Medicine Open, 2025 ACMG Annual Clinical Genetics Meeting*, 3, 102102. <https://doi.org/10.1016/j.gimo.2025.102102>
- Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., Kruglyak, S., & Saunders, C. T. (2016). Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 32(8), 1220–1222. <https://doi.org/10.1093/bioinformatics/btv710>
- Chen, Z., Morris, H. R., Polke, J., Wood, N. W., Gandhi, S., Ryten, M., Houlden, H., & Tucci, A. (2025). Repeat expansion disorders. *Practical Neurology*, 25(3), 204–216. <https://doi.org/10.1136/pn-2023-003938>

- Cheng, Y.-H. H., Sedeño-Cortés, A. E., Ranchalis, J. E., Munson, K. M., Vollger, M. R., Balton, E., Genetti, C. A., Network, U. D., Consortium, G. R. to E. the G. of R. D., Research, U. of W. C. for R. D., Wojcik, M. H., Beggs, A. H., Bamshad, M. J., Wei, C.-L., Dipple, K. M., Kumar, R. D., Blue, E. E., Jarvik, G., Chong, J. X., ... Stergachis, A. B. (2025). *Long-read transcriptome analysis using IsoRanker for identifying pathogenic variants in Mendelian conditions* (p. 2025.11.07.25339764). medRxiv. <https://doi.org/10.1101/2025.11.07.25339764>
- Chintalaphani, S. R., Pineda, S. S., Deveson, I. W., & Kumar, K. R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathologica Communications*, 9, 98. <https://doi.org/10.1186/s40478-021-01201-x>
- 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, J., Hyun, J., Hyun, J., Kim, J.-H., Lee, J. H., & Bang, D. (2024). Cost and time-efficient construction of a 3'-end mRNA library from unpurified bulk RNA in a single tube. *Experimental & Molecular Medicine*, 56(2), 453–460. <https://doi.org/10.1038/s12276-024-01164-8>
- Chu, M. L., & Moran, E. (2018). The Limb–Girdle Muscular Dystrophies: Is Treatment on the Horizon? *Neurotherapeutics*, 15(4), 849–862. <https://doi.org/10.1007/s13311-018-0648-x>
- Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D., & Kingsmore, S. F. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genomic Medicine*, 3, 16. <https://doi.org/10.1038/s41525-018-0053-8>
- Cleary, J. D., Subramony, S. H., & Ranum, L. P. (2021). Spinocerebellar Ataxia Type 8. In M. P. Adam, S. Bick, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, & A. Amemiya (Eds.), *GeneReviews®*. University of Washington, Seattle. <http://www.ncbi.nlm.nih.gov/books/NBK1268/> (Original work published 2001)
- 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>
- Comi, G. P., Prella, A., Bresolin, N., Moggio, M., Bardoni, A., Gallanti, A., Vita, G., Toscano, A., Ferro, M. T., & Bordini, A. (1994). Clinical variability in Becker muscular dystrophy. Genetic, biochemical and immunohistochemical correlates. *Brain: A Journal of Neurology*, 117 (Pt 1), 1–14. <https://doi.org/10.1093/brain/117.1.1-a>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17, 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Córdoba, M., Rodríguez-Quiroga, S. A., Vega, P. A., Salinas, V., Perez-Maturo, J., Amartino, H., Vásquez-Dusefante, C., Medina, N., González-Morón, D., & Kauffman, M. A. (2018). Whole exome sequencing in neurogenetic odysseys: An effective, cost- and time-saving diagnostic approach. *PLoS ONE*, 13(2), e0191228. <https://doi.org/10.1371/journal.pone.0191228>
- 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., ... Ravenscroft, G. (2024). A CCG expansion in ABCD3 causes oculopharyngodistal myopathy in individuals of European ancestry. *Nature Communications*, 15(1), 6327. <https://doi.org/10.1038/s41467-024-49950-2>
- Court, F. A., Hewitt, J. E., Davies, K., Patton, B. L., Uncini, A., Wrabetz, L., & Feltri, M. L. (2009). A Laminin-2, Dystroglycan, Utrophin Axis Is Required for Compartmentalization and Elongation of Myelin Segments. *The Journal of Neuroscience*, 29(12), 3908–3919. <https://doi.org/10.1523/JNEUROSCI.5672-08.2009>

- Couthouis, J., Raphael, A. R., Siskind, C., Findlay, A. R., Buenrostro, J. D., Greenleaf, W. J., Vogel, H., Day, J. W., Flanigan, K. M., & Gitler, A. D. (2014). Exome sequencing identifies a DNAJB6 mutation in a family with dominantly-inherited limb-girdle muscular dystrophy. *Neuromuscular Disorders: NMD*, 24(5), 431–435. <https://doi.org/10.1016/j.nmd.2014.01.014>
- Crisafulli, S., Sultana, J., Fontana, A., Salvo, F., Messina, S., & Trifirò, G. (2020). Global epidemiology of Duchenne muscular dystrophy: An updated systematic review and meta-analysis. *Orphanet Journal of Rare Diseases*, 15(1), 141. <https://doi.org/10.1186/s13023-020-01430-8>
- Cumming, S. A., Hamilton, M. J., Robb, Y., Gregory, H., McWilliam, C., Cooper, A., Adam, B., McGhie, J., Hamilton, G., Herzyk, P., Tschannen, M. R., Worthey, E., Petty, R., Ballantyne, B., Warner, J., Farrugia, M. E., Longman, C., & Monckton, D. G. (2018). De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *European Journal of Human Genetics*, 26(11), 1635–1647. <https://doi.org/10.1038/s41431-018-0156-9>
- Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., Bolduc, V., Waddell, L. B., Sandaradura, S. A., O’Grady, G. L., Estrella, E., Reddy, H. M., Zhao, F., Weisburd, B., Karczewski, K. J., O’Donnell-Luria, A. H., Birnbaum, D., Sarkozy, A., Hu, Y., ... MacArthur, D. G. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Science Translational Medicine*, 9(386), eaal5209. <https://doi.org/10.1126/scitranslmed.aal5209>
- Dahlqvist, J. R., Andersen, G., Khawajazada, T., Vissing, C., Thomsen, C., & Vissing, J. (2019). Relationship between muscle inflammation and fat replacement assessed by MRI in facioscapulohumeral muscular dystrophy. *Journal of Neurology*, 266(5), 1127–1135. <https://doi.org/10.1007/s00415-019-09242-y>
- Dale, C. M., King, J., Nonoyama, M., Carbone, S., McKim, D., Road, J., Rose, L., Amin, R., Avendano, M., Dial, S., Fan, E., Fraser, I., Fowler, R., Goldstein, R., Katz, S., King, J., Leasa, D., Mawdsley, C., McKim, D., ... Rubinfeld, G. (2018). Transitions to Home Mechanical Ventilation. The

- Experiences of Canadian Ventilator-assisted Adults and Their Family Caregivers. *Annals of the American Thoracic Society*, 15(3), 357–364. <https://doi.org/10.1513/AnnalsATS.201708-663OC>
- Darras, B. T., Urion, D. K., & Ghosh, P. S. (2022). Dystrophinopathies. In *GeneReviews® [Internet]*. University of Washington, Seattle. <https://www.ncbi.nlm.nih.gov/books/NBK1119/>
- Datta, S., Kataria, S., & Govindarajan, R. (2019). A Case Report on Charcot-Marie-Tooth Disease with a Novel Periaxin Gene Mutation. *Cureus*, 11(7), e5111. <https://doi.org/10.7759/cureus.5111>
- De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., & Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature Genetics*, 49(6), 941–945. <https://doi.org/10.1038/ng.3858>
- De La Vega, F. M., Chowdhury, S., Moore, B., Frise, E., McCarthy, J., Hernandez, E. J., Wong, T., James, K., Guidugli, L., Agrawal, P. B., Genetti, C. A., Brownstein, C. A., Beggs, A. H., Löscher, B.-S., Franke, A., Boone, B., Levy, S. E., Öunap, K., Pajusalu, S., ... Kingsmore, S. F. (2021). Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. *Genome Medicine*, 13(1), 153. <https://doi.org/10.1186/s13073-021-00965-0>
- de Moura, M. C. D. S., Wutzki, H. C., Voos, M. C., Resende, M. B. D., Reed, U. C., & Hasue, R. H. (2015). Is functional dependence of Duchenne muscular dystrophy patients determinant of the quality of life and burden of their caregivers? *Arquivos De Neuro-Psiquiatria*, 73(1), 52–57. <https://doi.org/10.1590/0004-282X20140194>
- de Sainte Agathe, J.-M., Filser, M., Isidor, B., Besnard, T., Gueguen, P., Perrin, A., Van Goethem, C., Verebi, C., Masingue, M., Rendu, J., Cossée, M., Bergougnoux, A., Frobert, L., Buratti, J., Lejeune, É., Le Guern, É., Pasquier, F., Clot, F., Kalatzis, V., ... Baux, D. (2023). SpliceAI-visual: A free online tool to improve SpliceAI splicing variant interpretation. *Human Genomics*, 17(1), 7. <https://doi.org/10.1186/s40246-023-00451-1>

- Deenen, J. C., 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>
- Deenen, J. C., Verbeek, A. L., Verschuuren, J. J., van Engelen, B. G., & Voermans, N. C. (2025). Prevalence and incidence rates of 17 neuromuscular disorders: An updated review of the literature. *Journal of Neuromuscular Diseases*, 12(6), 713–722. <https://doi.org/10.1177/22143602241313118>
- Del Bigio, M. R., Chudley, A. E., Sarnat, H. B., Campbell, C., Goobie, S., Chodirker, B. N., & Selcen, D. (2011). Infantile Muscular Dystrophy in Canadian Aboriginals is an α B-Crystallinopathy. *Annals of Neurology*, 69(5), 866–871. <https://doi.org/10.1002/ana.22331>
- Delalande, O., Molza, A.-E., Dos Santos Morais, R., Chéron, A., Pollet, É., Raguenes-Nicol, C., Tascon, C., Giudice, E., Guilbaud, M., Nicolas, A., Bondon, A., Leturcq, F., Férey, N., Baaden, M., Perez, J., Roblin, P., Piétri-Rouxel, F., Hubert, J.-F., Czjzek, M., & Le Rumeur, E. (2018). Dystrophin's central domain forms a complex filament that becomes disorganized by in-frame deletions. *The Journal of Biological Chemistry*, 293(18), 6637–6646. <https://doi.org/10.1074/jbc.M117.809798>
- Dolzhenko, E., English, A., Dashnow, H., De Sena Brandine, G., Mokveld, T., Rowell, W. J., Karniski, C., Kronenberg, Z., Danzi, M. C., Cheung, W. A., Bi, C., Farrow, E., Wenger, A., Chua, K. P., Martínez-Cerdeño, V., Bartley, T. D., Jin, P., Nelson, D. L., Zuchner, S., ... Eberle, M. A. (2024). Characterization and visualization of tandem repeats at genome scale. *Nature Biotechnology*, 42(10), 1606–1614. <https://doi.org/10.1038/s41587-023-02057-3>
- Dolzhenko, E., van Vugt, J. J. F. A., Shaw, R. J., Bekritsky, M. A., van Blitterswijk, M., Narzisi, G., Ajay, S. S., Rajan, V., Lajoie, B. R., Johnson, N. H., Kingsbury, Z., Humphray, S. J., Schellevis, R. D., Brands, W. J., Baker, M., Rademakers, R., Kooyman, M., Tazelaar, G. H. P., van Es, M. A., ... Eberle, M. A. (2017). Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Research*, 27(11), 1895–1903. <https://doi.org/10.1101/gr.225672.117>

- Donner, K., Sandbacka, M., Lehtokari, V.-L., Wallgren-Pettersson, C., & Pelin, K. (2004). Complete genomic structure of the human nebulin gene and identification of alternatively spliced transcripts. *European Journal of Human Genetics: EJHG*, 12(9), 744–751.
<https://doi.org/10.1038/sj.ejhg.5201242>
- 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>
- Duan, D., Goemans, N., Takeda, S., Mercuri, E., & Aartsma-Rus, A. (2021). Duchenne muscular dystrophy. *Nature Reviews Disease Primers*, 7(1), 13. <https://doi.org/10.1038/s41572-021-00248-3>
- 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., Djinoić-Carugo, K., Penttilä, S., Raheem, O., Reardon, K., Malandrini, A., Gambelli, S., Villanova, M., Nowak, K. J., ... 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/10.1016/j.ajhg.2011.04.021>
- Dziewas, R., Allescher, H.-D., Aroyo, I., Bartolome, G., Beilenhoff, U., Bohlender, J., Breitbach-Snowdon, H., Fheodoroff, K., Glahn, J., Heppner, H.-J., Hörmann, K., Ledl, C., Lücking, C., Pokieser, P., Schefold, J. C., Schröter-Morasch, H., Schweikert, K., Sparing, R., Trapl-Grundschober, M., ... Pflug, C. (2021). Diagnosis and treatment of neurogenic dysphagia – S1 guideline of the German Society of Neurology. *Neurological Research and Practice*, 3(1), 23.
<https://doi.org/10.1186/s42466-021-00122-3>
- Ebbert, M. T. W., Farrugia, S. L., Sens, J. P., Jansen-West, K., Gendron, T. F., Prudencio, M., McLaughlin, I. J., Bowman, B., Seetin, M., DeJesus-Hernandez, M., Jackson, J., Brown, P. H., Dickson, D. W., van Blitterswijk, M., Rademakers, R., Petrucelli, L., & Fryer, J. D. (2018). Long-read sequencing across the C9orf72 ‘GGGGCC’ repeat expansion: Implications for clinical use and genetic

- discovery efforts in human disease. *Molecular Neurodegeneration*, 13, 46.
<https://doi.org/10.1186/s13024-018-0274-4>
- Ebbink, B. J., Aarsen, F. K., van Gelder, C. M., van den Hout, J. M. P., Weisglas-Kuperus, N., Jaeken, J., Lequin, M. H., Arts, W. F. M., & van der Ploeg, A. T. (2012). Cognitive outcome of patients with classic infantile Pompe disease receiving enzyme therapy. *Neurology*, 78(19), 1512–1518.
<https://doi.org/10.1212/WNL.0b013e3182553c11>
- Efthymiou, S., Manole, A., & Houlden, H. (2016). Next generation sequencing in neuromuscular diseases. *Current Opinion in Neurology*, 29(5), 527–536.
<https://doi.org/10.1097/WCO.0000000000000374>
- Eisenhofer, R., Nesme, J., Santos-Bay, L., Koziol, A., Sørensen, S. J., Alberdi, A., & Aizpurua, O. (2024). A comparison of short-read, HiFi long-read, and hybrid strategies for genome-resolved metagenomics. *Microbiology Spectrum*, 12(4), e03590-23.
<https://doi.org/10.1128/spectrum.03590-23>
- Endo, Y., Noguchi, S., Hara, Y., Hayashi, Y. K., Motomura, K., Miyatake, S., Murakami, N., Tanaka, S., Yamashita, S., Kizu, R., Bamba, M., Goto, Y.-I., Matsumoto, N., Nonaka, I., & Nishino, I. (2015). Dominant mutations in ORAI1 cause tubular aggregate myopathy with hypocalcemia via constitutive activation of store-operated Ca^{2+} channels. *Human Molecular Genetics*, 24(3), 637–648. <https://doi.org/10.1093/hmg/ddu477>
- Escobar-Huertas, J. F., Vaca-González, J. J., Guevara, J. M., Ramirez-Martinez, A. M., Trabelsi, O., & Garzón-Alvarado, D. A. (2024). Duchenne and Becker muscular dystrophy: Cellular mechanisms, image analysis, and computational models: A review. *Cytoskeleton*, 81(6–7), 269–286.
<https://doi.org/10.1002/cm.21826>
- Espinosa, E., Bautista, R., Larrosa, R., & Plata, O. (2024). Advancements in long-read genome sequencing technologies and algorithms. *Genomics*, 116(3), 110842.
<https://doi.org/10.1016/j.ygeno.2024.110842>

- Esquivel-Sada, D., & Nguyen, M. T. (2018). Diagnosis of rare diseases under focus: Impacts for Canadian patients. *Journal of Community Genetics*, 9(1), 37–50. <https://doi.org/10.1007/s12687-017-0320-x>
- Fakhro, K. A., Choi, M., Ware, S. M., Belmont, J. W., Towbin, J. A., Lifton, R. P., Khokha, M. K., & Brueckner, M. (2011). Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proceedings of the National Academy of Sciences*, 108(7), 2915–2920. <https://doi.org/10.1073/pnas.1019645108>
- Fearnley, L. G., Bennett, M. F., & Bahlo, M. (2022). Detection of repeat expansions in large next generation DNA and RNA sequencing data without alignment. *Scientific Reports*, 12(1), 13124. <https://doi.org/10.1038/s41598-022-17267-z>
- Fehlberg, Z., Stark, Z., & Best, S. (2024). Reanalysis of genomic data, how do we do it now and what if we automate it? A qualitative study. *European Journal of Human Genetics: EJHG*, 32(5), 521–528. <https://doi.org/10.1038/s41431-023-01532-4>
- Feingold, B., Mahle, W. T., Auerbach, S., Clemens, P., Domenighetti, A. A., Jefferies, J. L., Judge, D. P., Lal, A. K., Markham, L. W., Parks, W. J., Tsuda, T., Wang, P. J., Yoo, S.-J., & On behalf of the American Heart Association Pediatric Heart Failure Committee of the Council on Cardiovascular Disease in the Young; Council on Clinical Cardiology; Council on Cardiovascular Radiology and Intervention; Council on Functional Genomics and Translational Biology; and Stroke Council. (2017). Management of Cardiac Involvement Associated With Neuromuscular Diseases: A Scientific Statement From the American Heart Association. *Circulation*, 136(13), e200–e231. <https://doi.org/10.1161/CIR.0000000000000526>
- Ferraro, N. M., Strober, B. J., Einson, J., Abell, N. S., Aguet, F., Barbeira, A. N., Brandt, M., Bucan, M., Castel, S. E., Davis, J. R., Greenwald, E., Hess, G. T., Hilliard, A. T., Kember, R. L., Kotis, B., Park, Y., Peloso, G., Ramdas, S., Scott, A. J., ... Battle, A. (2020). Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science (New York, N.Y.)*, 369(6509), eaaz5900. <https://doi.org/10.1126/science.aaz5900>

- Feske, S., Gwack, Y., Prakriya, M., Srikanth, S., Puppel, S.-H., Tanasa, B., Hogan, P. G., Lewis, R. S., Daly, M., & Rao, A. (2006). A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature*, 441(7090), 179–185. <https://doi.org/10.1038/nature04702>
- Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews. Genetics*, 7(2), 85–97. <https://doi.org/10.1038/nrg1767>
- Finsterer, J., & Stöllberger, C. (2008). Cardiac involvement in Becker muscular dystrophy. *The Canadian Journal of Cardiology*, 24(10), 786–792. [https://doi.org/10.1016/s0828-282x\(08\)70686-x](https://doi.org/10.1016/s0828-282x(08)70686-x)
- Finsterer, J., & Stöllberger, C. (2016). Heart Disease in Disorders of Muscle, Neuromuscular Transmission, and the Nerves. *Korean Circulation Journal*, 46(2), 117–134. <https://doi.org/10.4070/kcj.2016.46.2.117>
- Finsterer, J., Stöllberger, C., & Wahbi, K. (2013). Cardiomyopathy in neurological disorders. *Cardiovascular Pathology: The Official Journal of the Society for Cardiovascular Pathology*, 22(5), 389–400. <https://doi.org/10.1016/j.carpath.2012.12.008>
- Flanigan, K. M. (2014). Duchenne and Becker muscular dystrophies. *Neurologic Clinics*, 32(3), 671–688, viii. <https://doi.org/10.1016/j.ncl.2014.05.002>
- Franklin. (n.d.). Retrieved September 6, 2025, from <https://franklin.genoox.com/clinical-db/home>
- Frischauf, I., Litviňuková, M., Schober, R., Zayats, V., Svobodová, B., Bonhenry, D., Lunz, V., Cappello, S., Tociu, L., Reha, D., Stallinger, A., Hochreiter, A., Pammer, T., Butorac, C., Muik, M., Groschner, K., Bogeski, I., Ettrich, R. H., Romanin, C., & Schindl, R. (2017). Transmembrane helix connectivity in Orai1 controls two gates for calcium-dependent transcription. *Science Signaling*, 10(507), eaao0358. <https://doi.org/10.1126/scisignal.aao0358>
- Frisullo, G., Frusciante, R., Nociti, V., Tasca, G., Renna, R., Iorio, R., Patanella, A. K., Iannaccone, E., Marti, A., Rossi, M., Bianco, A., Monforte, M., Tonali, P. A., Mirabella, M., Batocchi, A. P., & Ricci, E. (2011). CD8⁺ T Cells in Facioscapulohumeral Muscular Dystrophy Patients with Inflammatory Features at Muscle MRI. *Journal of Clinical Immunology*, 31(2), 155–166. <https://doi.org/10.1007/s10875-010-9474-6>

- Fromer, M., Moran, J. L., Chambert, K., Banks, E., Bergen, S. E., Ruderfer, D. M., Handsaker, R. E., McCarroll, S. A., O'Donovan, M. C., Owen, M. J., Kirov, G., Sullivan, P. F., Hultman, C. M., Sklar, P., & Purcell, S. M. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *American Journal of Human Genetics*, 91(4), 597–607. <https://doi.org/10.1016/j.ajhg.2012.08.005>
- Frosk, P., Del Bigio, M. R., Wrogemann, K., & Greenberg, C. R. (2005). Hutterite brothers both affected with two forms of limb girdle muscular dystrophy: LGMD2H and LGMD2I. *European Journal of Human Genetics*, 13(8), 978–982. <https://doi.org/10.1038/sj.ejhg.5201436>
- Frosk, P., Greenberg, C. R., Tennese, A. A. P., Lamont, R., Nylen, E., Hirst, C., Frappier, D., Roslin, N. M., Zaik, M., Bushby, K., Straub, V., Zatz, M., de Paula, F., Morgan, K., Fujiwara, T. M., & Wrogemann, K. (2005). The most common mutation in FKRP causing limb girdle muscular dystrophy type 2I (LGMD2I) may have occurred only once and is present in Hutterites and other populations. *Human Mutation*, 25(1), 38–44. <https://doi.org/10.1002/humu.20110>
- Fung, J. L. F., Yu, M. H. C., Huang, S., Chung, C. C. Y., Chan, M. C. Y., Pajusalu, S., Mak, C. C. Y., Hui, V. C. C., Tsang, M. H. Y., Yeung, K. S., Lek, M., & Chung, B. H. Y. (2020). A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis. *Npj Genomic Medicine*, 5(1), 37. <https://doi.org/10.1038/s41525-020-00144-x>
- García-González, J., & Gogolewski, K. (2026). AlphaGenome: A Swiss-army knife for exploring non-coding DNA. *Trends in Genetics : TIG*, 42(1), 4–6. <https://doi.org/10.1016/j.tig.2025.11.007>
- Gargano, M. A., Matentzoglou, N., Coleman, B., Addo-Lartey, E. B., Anagnostopoulos, A. V., Anderton, J., Avillach, P., Bagley, A. M., Bakštein, E., Balhoff, J. P., Baynam, G., Bello, S. M., Berk, M., Bertram, H., Bishop, S., Blau, H., Bodenstein, D. F., Botas, P., Boztug, K., ... Robinson, P. N. (2024). The Human Phenotype Ontology in 2024: Phenotypes around the world. *Nucleic Acids Research*, 52(D1), D1333–D1346. <https://doi.org/10.1093/nar/gkad1005>
- Garibaldi, M., Fattori, F., Riva, B., Labasse, C., Brochier, G., Ottaviani, P., Sacconi, S., Vizzaccaro, E., Laschena, F., Romero, N. B., Genazzani, A., Bertini, E., & Antonini, G. (2017). A novel gain-of-

- function mutation in ORAI1 causes late-onset tubular aggregate myopathy and congenital miosis. *Clinical Genetics*, 91(5), 780–786. <https://doi.org/10.1111/cge.12888>
- Gautier, B., Hajjar, H., Soares, S., Berthelot, J., Deck, M., Abbou, S., Campbell, G., Ceprian, M., Gonzalez, S., Fovet, C.-M., Schütza, V., Jouvenel, A., Rivat, C., Zerah, M., François, V., Le Guiner, C., Aubourg, P., Fledrich, R., & Tricaud, N. (2021). AAV2/9-mediated silencing of PMP22 prevents the development of pathological features in a rat model of Charcot-Marie-Tooth disease 1 A. *Nature Communications*, 12(1), 2356. <https://doi.org/10.1038/s41467-021-22593-3>
- Genetics and Metabolism*. (n.d.). Winnipeg Regional Health Authority. Retrieved May 21, 2026, from <https://wrha.mb.ca/genetics-and-metabolism/>
- 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>
- 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/10.1038/s41431-022-01190-y>
- Geng, L. N., Yao, Z., Snider, L., Fong, A. P., Cech, J. N., Young, J. M., van der Maarel, S. M., Ruzzo, W. L., Gentleman, R. C., Tawil, R., & Tapscott, S. J. (2012). DUX4 activates germline genes, retroelements and immune-mediators: Implications for facioscapulohumeral dystrophy. *Developmental Cell*, 22(1), 38–51. <https://doi.org/10.1016/j.devcel.2011.11.013>
- Ghaoui, R., Cooper, S. T., Lek, M., Jones, K., Corbett, A., Reddel, S. W., Needham, M., Liang, C., Waddell, L. B., Nicholson, G., O’Grady, G., Kaur, S., Ong, R., Davis, M., Sue, C. M., Laing, N. G., North, K. N., MacArthur, D. G., & Clarke, N. F. (2015). Use of Whole-Exome Sequencing for Diagnosis of Limb-Girdle Muscular Dystrophy: Outcomes and Lessons Learned. *JAMA Neurology*, 72(12), 1424–1432. <https://doi.org/10.1001/jamaneurol.2015.2274>

- Gibson, K. M., Nesbitt, A., Cao, K., Yu, Z., Denenberg, E., DeChene, E., Guan, Q., Bhoj, E., Zhou, X., Zhang, B., Wu, C., Dubbs, H., Wilkens, A., Medne, L., Bedoukian, E., White, P. S., Pennington, J., Lou, M., Conlin, L., ... Santani, A. (2018). Novel findings with reassessment of exome data: Implications for validation testing and interpretation of genomic data. *Genetics in Medicine*, 20(3), 329–336. <https://doi.org/10.1038/gim.2017.153>
- Gigante, S., Gouil, Q., Lucattini, A., Keniry, A., Beck, T., Tinning, M., Gordon, L., Woodruff, C., Speed, T. P., Blewitt, M. E., & Ritchie, M. E. (2019). Using long-read sequencing to detect imprinted DNA methylation. *Nucleic Acids Research*, 47(8), e46. <https://doi.org/10.1093/nar/gkz107>
- Gillespie, C. S., Sherman, D. L., Blair, G. E., & Brophy, P. J. (1994). Periaxin, a novel protein of myelinating Schwann cells with a possible role in axonal ensheathment. *Neuron*, 12(3), 497–508. [https://doi.org/10.1016/0896-6273\(94\)90208-9](https://doi.org/10.1016/0896-6273(94)90208-9)
- Gillespie, C. S., Sherman, D. L., Fleetwood-Walker, S. M., Cottrell, D. F., Tait, S., Garry, E. M., Wallace, V. C. J., Ure, J., Griffiths, I. R., Smith, A., & Brophy, P. J. (2000). Peripheral Demyelination and Neuropathic Pain Behavior in Periaxin-Deficient Mice. *Neuron*, 26(2), 523–531. [https://doi.org/10.1016/S0896-6273\(00\)81184-8](https://doi.org/10.1016/S0896-6273(00)81184-8)
- Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., Kao, D., Ohri, K., Viththyapaskaran, S., Tarnopolsky, M. A., Mathews, K. D., Moore, S. A., Osorio, A. N., Villanova, D., Kemaladewi, D. U., Cohn, R. D., Brudno, M., & Dowling, J. J. (2019). Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *The American Journal of Human Genetics*, 104(3), 466–483. <https://doi.org/10.1016/j.ajhg.2019.01.012>
- Gonzalez-Quereda, L., Rodriguez, M. J., Diaz-Manera, J., Alonso-Perez, J., Gallardo, E., Nascimento, A., Ortez, C., Natera-de Benito, D., Olive, M., Gonzalez-Mera, L., Lopez de Munain, A., Zulaica, M., Poza, J. J., Jerico, I., Torne, L., Riera, P., Milisenda, J., Sanchez, A., Garrabou, G., ... Gallano, P. (2020). Targeted Next-Generation Sequencing in a Large Cohort of Genetically Undiagnosed

- Patients with Neuromuscular Disorders in Spain. *Genes*, 11(5), 539.
<https://doi.org/10.3390/genes11050539>
- Government of Canada, S. C. (2021, September 17). *Population group of person*.
<https://www23.statcan.gc.ca/imdb/p3Var.pl?Function=DEC&Id=45159>
- Graham, C. D., Rose, M. R., Grunfeld, E. A., Kyle, S. D., & Weinman, J. (2011). A systematic review of quality of life in adults with muscle disease. *Journal of Neurology*, 258(9), 1581–1592.
<https://doi.org/10.1007/s00415-011-6062-5>
- GTEx Consortium. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509), 1318–1330. <https://doi.org/10.1126/science.aaz1776>
- Gudmundsson, S., Singer-Berk, M., Watts, N. A., Phu, W., Goodrich, J. K., Solomonson, M., Genome Aggregation Database Consortium, Rehm, H. L., MacArthur, D. G., & O'Donnell-Luria, A. (2022). Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation*, 43(8), 1012–1030. <https://doi.org/10.1002/humu.24309>
- Guizard, S., Miedzinska, K., Smith, J., Smith, J., Kuo, R. I., Davey, M., Archibald, A., & Watson, M. (2023). nf-core/isoseq: Simple gene and isoform annotation with PacBio Iso-Seq long-read sequencing. *Bioinformatics*, 39(5), btad150. <https://doi.org/10.1093/bioinformatics/btad150>
- Guo, Y., Long, J., He, J., Li, C.-I., Cai, Q., Shu, X.-O., Zheng, W., & Li, C. (2012). Exome sequencing generates high quality data in non-target regions. *BMC Genomics*, 13, 194.
<https://doi.org/10.1186/1471-2164-13-194>
- Gupta, A., & Jankovic, J. (2009). Spinocerebellar ataxia 8: Variable phenotype and unique pathogenesis. *Parkinsonism & Related Disorders*, 15(9), 621–626.
<https://doi.org/10.1016/j.parkreldis.2009.06.001>
- Han, Y., Jang, J.-H., & Chang, H. (2026). Targeted long-read sequencing for high-resolution repeat profiling in myotonic dystrophy type 1. *Experimental & Molecular Medicine*, 58(4), 1203–1215.
<https://doi.org/10.1038/s12276-026-01683-6>

- Hannan, A. J. (2012). Tandem repeat polymorphisms: Mediators of genetic plasticity, modulators of biological diversity and dynamic sources of disease susceptibility. *Advances in Experimental Medicine and Biology*, 769, 1–9.
- Hansen, L., Comyn, S., Mang, Y., Lind-Thomsen, A., Myhre, L., Jean, F., Eiberg, H., Tommerup, N., Rosenberg, T., & Pilgrim, D. (2014). The myosin chaperone UNC45B is involved in lens development and autosomal dominant juvenile cataract. *European Journal of Human Genetics*, 22(11), 1290–1297. <https://doi.org/10.1038/ejhg.2014.21>
- Hartley, T., Gillespie, M. K., Graham, I. D., Hayeems, R. Z., Li, S., Sampson, M., Boycott, K. M., & Potter, B. K. (2023). Exome and genome sequencing for rare genetic disease diagnosis: A scoping review and critical appraisal of clinical guidance documents produced by genetics professional organizations. *Genetics in Medicine*, 25(11), 100948. <https://doi.org/10.1016/j.gim.2023.100948>
- Hasin, Y., Seldin, M., & Lusi, A. (2017). Multi-omics approaches to disease. *Genome Biology*, 18, 83. <https://doi.org/10.1186/s13059-017-1215-1>
- Haskell, G. T., Adams, M. C., Fan, Z., Amin, K., Guzman Badillo, R. J., Zhou, L., Bizon, C., Chahin, N., Greenwood, R. S., Milko, L. V., Shiloh-Malawsky, Y., Crooks, K. R., Strande, N., Tennison, M., Tilley, C. R., Brandt, A., Wilhelmsen, K. C., Weck, K., Evans, J. P., & Berg, J. S. (2018a). Diagnostic utility of exome sequencing in the evaluation of neuromuscular disorders. *Neurology: Genetics*, 4(1), e212. <https://doi.org/10.1212/NXG.0000000000000212>
- Hayeems, R. Z., Ungar, W. J., Marshall, C. R., Gillespie, M. K., Szuto, A., Huang, L., Venkataramanan, V., Xiao, B., Chisholm, C., Stavropoulos, D. J., Bergeron, M. B., Lee, W., Costain, G., Jobling, R., Sawyer, S., Price, E. M., Lau, L., Mendoza, R., Somerville, M. J., & Boycott, K. M. (2026). Comparing the performance of exome and genome sequencing for rare disease diagnostics: A randomized implementation effectiveness trial. *Genetics in Medicine*, 28(1), 101605. <https://doi.org/10.1016/j.gim.2025.101605>
- He, Y., Zu, T., Benzow, K. A., Orr, H. T., Clark, H. B., & Koob, M. D. (2006). Targeted deletion of a single Sca8 ataxia locus allele in mice causes abnormal gait, progressive loss of motor

- coordination, and Purkinje cell dendritic deficits. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 26(39), 9975–9982.
<https://doi.org/10.1523/JNEUROSCI.2595-06.2006>
- Hekselman, I., & Yeger-Lotem, E. (2020). Mechanisms of tissue and cell-type specificity in heritable traits and diseases. *Nature Reviews. Genetics*, 21(3), 137–150. <https://doi.org/10.1038/s41576-019-0200-9>
- Hendrickson, P. G., Doráis, J. A., Grow, E. J., Whiddon, J. L., Lim, J.-W., Wike, C. L., Weaver, B. D., Pflueger, C., Emery, B. R., Wilcox, A. L., Nix, D. A., Peterson, C. M., Tapscott, S. J., Carrell, D. T., & Cairns, B. R. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERV1/HERV1 retrotransposons. *Nature Genetics*, 49(6), 925–934.
<https://doi.org/10.1038/ng.3844>
- Herman, I., Lopez, M. A., Marafi, D., Pehlivan, D., Calame, D. G., Abid, F., & Lotze, T. E. (2021). Clinical exome sequencing in the diagnosis of pediatric neuromuscular disease. *Muscle & Nerve*, 63(3), 304–310. <https://doi.org/10.1002/mus.27112>
- Hiatt, L., Weisburd, B., Dolzhenko, E., Rubinetti, V., Avvaru, A. K., VanNoy, G. E., Kurtas, N. E., Rehm, H. L., Quinlan, A. R., & Dashnow, H. (2025). STRchive: A dynamic resource detailing population-level and locus-specific insights at tandem repeat disease loci. *Genome Medicine*, 17(1), 29. <https://doi.org/10.1186/s13073-025-01454-4>
- Hilbert, J. E., Johnson, N. E., & Moxley, R. T. (2013). New Insights About the Incidence, Multisystem Manifestations, and Care of Patients with Congenital Myotonic Dystrophy. *The Journal of Pediatrics*, 163(1), 12–14. <https://doi.org/10.1016/j.jpeds.2013.02.022>
- Hildonen, M., Mariani, L., Dalsberg, J., Bak, M., Weksberg, R., Choufani, S., & Tümer, Z. (2026). Clinical Feasibility of Long-Read WGS for DNA Methylation Signature Analysis. *Clinical Genetics*, 109(4), 725–729. <https://doi.org/10.1111/cge.70108>
- Hoffman, E., Fischbeck, K., Brown, R., Johnson, M., Medori, R., Loire, J., Harris, J., Waterston, R., Brooke, M., Specht, L., Kupsky, W., Chamberlain, J., Caskey, T., Shapiro, F., & Kunkel, L.

- (1988). *Characterization of Dystrophin in Muscle-Biopsy Specimens from Patients with Duchenne's or Becker's Muscular Dystrophy*. 318(21), 1363–1368.
- Hoffman, E. P., Brown, R. H., & Kunkel, L. M. (1987). Dystrophin: The protein product of the Duchenne muscular dystrophy locus. *Cell*, 51(6), 919–928. [https://doi.org/10.1016/0092-8674\(87\)90579-4](https://doi.org/10.1016/0092-8674(87)90579-4)
- Hong, S. E., Kneissl, J., Cho, A., Kim, M. J., Park, S., Lee, J., Woo, S., Kim, S., Kim, J.-S., Kim, S. Y., Jung, S., Kim, J., Shin, J.-Y., Chae, J.-H., & Choi, M. (2022). Transcriptome-based variant calling and aberrant mRNA discovery enhance diagnostic efficiency for neuromuscular diseases. *Journal of Medical Genetics*, 59(11), 1075–1081. <https://doi.org/10.1136/jmedgenet-2021-108307>
- Höps, W., Weiss, M. M., Derks, R., Galbany, J. C., Ouden, A. den, van den Heuvel, S., Timmermans, R., Smits, J., Mokveld, T., Dolzhenko, E., Chen, X., van den Wijngaard, A., Eberle, M. A., Yntema, H. G., Hoischen, A., Gilissen, C., & Vissers, L. E. L. M. (2025). HiFi long-read genomes for difficult-to-detect, clinically relevant variants. *American Journal of Human Genetics*, 112(2), 450–456. <https://doi.org/10.1016/j.ajhg.2024.12.013>
- Hrdlickova, R., Toloue, M., & Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews. RNA*, 8(1). <https://doi.org/10.1002/wrna.1364>
- Hu, X., Li, N., Xu, Y., Li, G., Yu, T., Yao, R.-E., Fu, L., Wang, J., Yin, L., Yin, Y., Wang, Y., Jin, X., Wang, X., Wang, J., & Shen, Y. (2018). Proband-only medical exome sequencing as a cost-effective first-tier genetic diagnostic test for patients without prior molecular tests and clinical diagnosis in a developing country: The China experience. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 20(9), 1045–1053. <https://doi.org/10.1038/gim.2017.195>
- Huang, M., Zhang, Q., Jiao, J., Shi, J., Xu, Y., Zhang, C., Zhou, R., Liu, W., Liang, Y., Chen, H., Wang, Y., Xu, Z., & Hu, P. (2024). Comprehensive genetic analysis of facioscapulohumeral muscular dystrophy by Nanopore long-read whole-genome sequencing. *Journal of Translational Medicine*, 22(1), 451. <https://doi.org/10.1186/s12967-024-05259-8>
- Hubregtse, L., Bouman, K., Lama, C., Lassche, S., Graaf, N. de, Taglietti, V., Küsters, B., Periou, B., Relaix, F., Engelen, B. van, Authier, F.-J., Voermans, N. C., & Malfatti, E. (2024). An up-to-date

- myopathologic characterisation of facioscapulohumeral muscular dystrophy type 1 muscle biopsies shows sarcolemmal complement membrane attack complex deposits and increased skeletal muscle regeneration. *Neuromuscular Disorders*, 36, 6–15.
<https://doi.org/10.1016/j.nmd.2024.01.003>
- Husson, T., Lecoquierre, F., Nicolas, G., Richard, A.-C., Afenjar, A., Audebert-Bellanger, S., Badens, C., Bilan, F., Bizaoui, V., Boland, A., Bonnet-Dupeyron, M.-N., Brischoux-Boucher, E., Bonnet, C., Bournez, M., Boute, O., Brunelle, P., Caumes, R., Charles, P., Chassaing, N., ... Charbonnier, C. (2024). Episignatures in practice: Independent evaluation of published episignatures for the molecular diagnostics of ten neurodevelopmental disorders. *European Journal of Human Genetics*, 32(2), 190–199. <https://doi.org/10.1038/s41431-023-01474-x>
- Ikeda, Y., Dalton, J. C., Moseley, M. L., Gardner, K. L., Bird, T. D., Ashizawa, T., Seltzer, W. K., Pandolfo, M., Milunsky, A., Potter, N. T., Shoji, M., Vincent, J. B., Day, J. W., & Ranum, L. P. W. (2004). Spinocerebellar Ataxia Type 8: Molecular Genetic Comparisons and Haplotype Analysis of 37 Families with Ataxia. *American Journal of Human Genetics*, 75(1), 3–16.
<https://doi.org/10.1086/422014>
- Ikeda, Y., Shizuka, M., Watanabe, M., Okamoto, K., & Shoji, M. (2000). Molecular and clinical analyses of spinocerebellar ataxia type 8 in Japan. *Neurology*, 54(4), 950–955.
<https://doi.org/10.1212/wnl.54.4.950>
- 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., ... 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>
- Ishiura, H., Doi, K., Mitsui, J., Yoshimura, J., Matsukawa, M. K., Fujiyama, A., Toyoshima, Y., Kakita, A., Takahashi, H., Suzuki, Y., Sugano, S., Qu, W., Ichikawa, K., Yurino, H., Higasa, K., Shibata,

- S., Mitsue, A., Tanaka, M., Ichikawa, Y., ... Tsuji, S. (2018). Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nature Genetics*, 50(4), 581–590.
<https://doi.org/10.1038/s41588-018-0067-2>
- Itoh, G., Sugino, S., Ikeda, M., Mizuguchi, M., Kanno, S., Amin, M. A., Iemura, K., Yasui, A., Hirota, T., & Tanaka, K. (2013). Nucleoporin Nup188 is required for chromosome alignment in mitosis. *Cancer Science*, 104(7), 871–879. <https://doi.org/10.1111/cas.12159>
- Ivashchenko, V., de Groot, M., Derks, R., den Ouden, A., Khazeeva, G., van den Heuvel, S., Timmermans, R., Corominas Galbany, J., Pfundt, R., Hofste, T., Yntema, H., Vissers, L., Hoischen, A., Hampstead, J., & Gilissen, C. (2026). Genome-wide methylation detection and episinature analysis using PacBio long-read sequencing. *Genome Medicine*, 18(1), 11.
<https://doi.org/10.1186/s13073-025-01506-9>
- Jadhav, T., Aruta, M., Diaz-miranda, M. A., Zucco, A., Conlin, L. K., Rajagopalan, R., & Wang, J. (2026). Evaluation of PacBio Long-Read and PCR-Based Short-Read Sequencing for Mitochondrial DNA (mtDNA) Variant Detection, with an Emphasis on Detection and Quantification of mtDNA Deletion. *International Journal of Molecular Sciences*, 27(8), 3562.
<https://doi.org/10.3390/ijms27083562>
- Jaganathan, K., Ersaro, N., Novakovsky, G., Wang, Y., James, T., Schwartzentruber, J., Fiziev, P., Kassam, I., Cao, F., Hawe, J., Cavanagh, H., Lim, A., Png, G., McRae, J., Banerjee, A., Kumar, A., Ulirsch, J., Zhang, Y., Aguet, F., ... Farh, K. K.-H. (2025). Predicting expression-altering promoter mutations with deep learning. *Science*, 389(6760), eads7373.
<https://doi.org/10.1126/science.ads7373>
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O’Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., ... Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, 36(4), 338–345.
<https://doi.org/10.1038/nbt.4060>

- Jana, U., Rodriguez, O. L., Lees, W., Engelbrecht, E., Vanwinkle, Z., Peres, A., Gibson, W. S., Shields, K., Schultze, S., Dorgham, A., Emery, M., Deikus, G., Sebra, R., Eichler, E. E., Yaari, G., Smith, M. L., & Watson, C. T. (2026). The human IG heavy chain constant gene locus is enriched for large structural variants and coding polymorphisms that vary among human populations. *Cell Genomics*, 6(1), 101058. <https://doi.org/10.1016/j.xgen.2025.101058>
- Jin, H., Fu, H., Wang, J., Wang, Z., Liu, J., Han, F., Zheng, H., & Jiang, Y. (2024). Generation of a DMD loss-of-function mutant human embryonic stem cell lines by CRISPR base editing. *Stem Cell Research*, 76, 103343. <https://doi.org/10.1016/j.scr.2024.103343>
- Johnson, K., Schneider, T., Babaian, N., Azage, M., & McIntosh, P. (2024). 475P A novel clinic to resolve variants of uncertain significance in neuromuscular patients. *Neuromuscular Disorders*, *Neuromuscular Disorders WMS Congress Issue - Abstracts 2024*, 43, 104441.312. <https://doi.org/10.1016/j.nmd.2024.07.321>
- Jones, K., & Moore, S. (2020). NEW GENES AND DISEASES / NGS & RELATED TECHNIQUES: P.328 Neuromuscular disease variant of unknown significance (VUS) resolution using muscle biopsy evaluation: The Iowa experience. *Neuromuscular Disorders*, 30, S142. <https://doi.org/10.1016/j.nmd.2020.08.324>
- Kagiava, A., Karaikos, C., Richter, J., Tryfonos, C., Jennings, M. J., Heslegrave, A. J., Sargiannidou, I., Stavrou, M., Zetterberg, H., Reilly, M. M., Christodoulou, C., Horvath, R., & Kleopa, K. A. (2021). AAV9-mediated Schwann cell-targeted gene therapy rescues a model of demyelinating neuropathy. *Gene Therapy*, 28(10), 659–675. <https://doi.org/10.1038/s41434-021-00250-0>
- Kamolvisit, W., Cheawsamoot, C., Chetruengchai, W., Kor-anantakul, P., Thangpong, R., Srichomthong, C., Assawapitaksakul, A., Syananondh, K., Kontun, S., Buasong, A., Od-Ek, P., Suphapeetiporn, K., & Shotelersuk, V. (2025). Singleton rapid long-read genome sequencing as first tier genetic test for critically ill children with suspected genetic diseases. *European Journal of Human Genetics*, 33(6), 768–774. <https://doi.org/10.1038/s41431-025-01818-9>

- Karachunski, P., Stark, M., McGrattan, K., Weigel, C., Ihinger, J., & Kang, P. (2022). P.48 The impact of newborn screening and early therapy on the course of spinal muscular atrophy: A retrospective analysis of a single center experience. *Neuromuscular Disorders*, 32, S61.
<https://doi.org/10.1016/j.nmd.2022.07.084>
- 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>
- Karnebeek, I. E. A., Boon, H. T. M., Huis, A. M. P., Cup, E. H. C., Eggink, C. A., Schouten, M. I., van der Looij, H. J., van Engelen, B. G. M., Smulders, F. H. P., & Voermans, N. C. (2022). Having an eye for myotonic dystrophy: A qualitative study on experiences and support needs in myotonic dystrophy type 1 patients with a diagnostic delay after early-onset cataract. *Neuromuscular Disorders*, 32(10), 829–835. <https://doi.org/10.1016/j.nmd.2022.09.003>
- Karunakaran, S., Kuruvilla, A., Nair, M., Nair, S. S., & Narasimhaiah, D. (2025). A Clinical Study on the Utility of Muscle Biopsy in Patients with Suspected Myopathy. *Annals of Indian Academy of Neurology*, 28(3), 363. https://doi.org/10.4103/aian.aian_934_24
- Kernohan, K. D., & Boycott, K. M. (2024). The expanding diagnostic toolbox for rare genetic diseases. *Nature Reviews Genetics*, 25(6), 401–415. <https://doi.org/10.1038/s41576-023-00683-w>
- Kevelam, S. H., Taube, J. R., van Spaendonk, R. M. L., Bertini, E., Sperle, K., Tarnopolsky, M., Tonduti, D., Valente, E. M., Travaglini, L., Sistermans, E. A., Bernard, G., Catsman-Berrevoets, C. E., van Karnebeek, C. D. M., Østergaard, J. R., Friederich, R. L., Fawzi Elsaid, M., Schieving, J. H., Tarailo-Graovac, M., Orcesi, S., ... Wolf, N. I. (2015). Altered PLP1 splicing causes hypomyelination of early myelinating structures. *Annals of Clinical and Translational Neurology*, 2(6), 648–661. <https://doi.org/10.1002/acn3.203>

- Khadilkar, S. V., Yadav, R. S., & Patel, B. A. (2024). *Neuromuscular Disorders: A Comprehensive Review with Illustrative Cases*. Springer Nature. <https://doi.org/10.1007/978-981-97-9010-4>
- Khan, A., Frazer-Green, L., Amin, R., Wolfe, L., Faulkner, G., Casey, K., Sharma, G., Selim, B., Zielinski, D., Aboussouan, L. S., McKim, D., & Gay, P. (2023). Respiratory Management of Patients With Neuromuscular Weakness: An American College of Chest Physicians Clinical Practice Guideline and Expert Panel Report. *CHEST*, *164*(2), 394–413. <https://doi.org/10.1016/j.chest.2023.03.011>
- Kieny, P., Chollet, S., Delalande, P., Le Fort, M., Magot, A., Pereon, Y., & Perrouin Verbe, B. (2013). Evolution of life expectancy of patients with Duchenne muscular dystrophy at AFM Yolaine de Kepper centre between 1981 and 2011. *Annals of Physical and Rehabilitation Medicine*, *56*(6), 443–454. <https://doi.org/10.1016/j.rehab.2013.06.002>
- Kiiski, K., Lehtokari, V.-L., Löytynoja, A., Ahlstén, L., Laitila, J., Wallgren-Pettersson, C., & Pelin, K. (2016). A recurrent copy number variation of the NEB triplicate region: Only revealed by the targeted nemaline myopathy CGH array. *European Journal of Human Genetics*, *24*(4), 574–580. <https://doi.org/10.1038/ejhg.2015.166>
- Kim, J., Ahmed, U. A., Pagan, A., Margolin, R. E., Rohan, A., Tantivit, N., Boskey, E. R., & Renthal, N. E. (2026). Moving beyond puberty: Listening to lived experience to expand sexual and reproductive healthcare for adults with muscular dystrophy. *Journal of Neuromuscular Diseases*, 22143602261445253. <https://doi.org/10.1177/22143602261445253>
- Kircher, M., Witten, D. M., Jain, P., O’Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. <https://doi.org/10.1038/ng.2892>
- Kitamura, Y., Kondo, E., Urano, M., Aoki, R., & Saito, K. (2016). Target resequencing of neuromuscular disease-related genes using next-generation sequencing for patients with undiagnosed early-onset neuromuscular disorders. *Journal of Human Genetics*, *61*(11), 931–942. <https://doi.org/10.1038/jhg.2016.79>

- Knierim, E., Lucke, B., Schwarz, J. M., Schuelke, M., & Seelow, D. (2011). Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PloS One*, 6(11), e28240. <https://doi.org/10.1371/journal.pone.0028240>
- Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J. O. B., Danis, D., Gourdine, J.-P., Gargano, M., Harris, N. L., Matentzoglou, N., McMurry, J. A., Osumi-Sutherland, D., Cipriani, V., Balhoff, J. P., Conlin, T., Blau, H., Baynam, G., Palmer, R., Gratian, D., Dawkins, H., ... Robinson, P. N. (2019). Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research*, 47(D1), D1018–D1027. <https://doi.org/10.1093/nar/gky1105>
- Koob, M. D., Moseley, M. L., Schut, L. J., Benzow, K. A., Bird, T. D., Day, J. W., & Ranum, L. P. W. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nature Genetics*, 21(4), 379–384. <https://doi.org/10.1038/7710>
- Kreiman, B. L., & Boles, R. G. (2020). State of the Art of Genetic Testing for Patients With Autism: A Practical Guide for Clinicians. *Seminars in Pediatric Neurology, Advances in the Diagnosis and Treatment of Autism Spectrum Disorder-1*, 34, 100804. <https://doi.org/10.1016/j.spen.2020.100804>
- Kremer, L. S., Bader, D. M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., Haack, T. B., Graf, E., Schwarzmayer, T., Terrile, C., Koňáříková, E., Repp, B., Kastenmüller, G., Adamski, J., Lichtner, P., Leonhardt, C., Funalot, B., Donati, A., Tiranti, V., ... Prokisch, H. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nature Communications*, 8(1), 15824. <https://doi.org/10.1038/ncomms15824>
- Ku, C.-S., Wu, M., Cooper, D. N., Naidoo, N., Pawitan, Y., Pang, B., Iacopetta, B., & Soong, R. (2012). Exome versus transcriptome sequencing in identifying coding region variants. *Expert Review of Molecular Diagnostics*, 12(3), 241–251. <https://doi.org/10.1586/erm.12.10>
- Kuhn, M., Gläser, D., Joshi, P. R., Zierz, S., Wenninger, S., Schoser, B., & Deschauer, M. (2016). Utility of a next-generation sequencing-based gene panel investigation in German patients with

- genetically unclassified limb-girdle muscular dystrophy. *Journal of Neurology*, 263(4), 743–750.
<https://doi.org/10.1007/s00415-016-8036-0>
- Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, 2015(11), 951–969. <https://doi.org/10.1101/pdb.top084970>
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1), 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- Labade, A. S., Karmodiya, K., & Sengupta, K. (2016). HOXA repression is mediated by nucleoporin Nup93 assisted by its interactors Nup188 and Nup205. *Epigenetics & Chromatin*, 9, 54.
<https://doi.org/10.1186/s13072-016-0106-0>
- Landfeldt, E., Lindgren, P., Bell, C. F., Guglieri, M., Straub, V., Lochmüller, H., & Bushby, K. (2016). Quantifying the burden of caregiving in Duchenne muscular dystrophy. *Journal of Neurology*, 263, 906–915. <https://doi.org/10.1007/s00415-016-8080-9>
- Landfeldt, E., Lindgren, P., Bell, C. F., Schmitt, C., Guglieri, M., Straub, V., Lochmüller, H., & Bushby, K. (2014). The burden of Duchenne muscular dystrophy. *Neurology*, 83(6), 529–536.
<https://doi.org/10.1212/WNL.0000000000000669>
- 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(Database issue), D980–D985. <https://doi.org/10.1093/nar/gkt1113>
- Lappalainen, I., Lopez, J., Skipper, L., Hefferon, T., Spalding, J. D., Garner, J., Chen, C., Maguire, M., Corbett, M., Zhou, G., Paschall, J., Ananiev, V., Flicek, P., & Church, D. M. (2013). dbVar and DGVa: Public archives for genomic structural variation. *Nucleic Acids Research*, 41(Database issue), D936–D941. <https://doi.org/10.1093/nar/gks1213>
- Law, M., King, G., King, S., Kertoy, M., Hurley, P., Rosenbaum, P., Young, N., & Hanna, S. (2006). Patterns of participation in recreational and leisure activities among children with complex physical disabilities. *Developmental Medicine and Child Neurology*, 48(5), 337–342.
<https://doi.org/10.1017/S0012162206000740>

- Leckie, J. N., Joel, M. M., Martens, K., King, A., King, M., Korngut, L. W., de Koning, A. P. J., Pfeffer, G., & Schellenberg, K. L. (2021). Highly Elevated Prevalence of Spinobulbar Muscular Atrophy in Indigenous Communities in Canada Due to a Founder Effect. *Neurology Genetics*, 7(4), e607. <https://doi.org/10.1212/NXG.0000000000000607>
- Ledergerber, C., & Dessimoz, C. (2011). Base-calling for next-generation sequencing platforms. *Briefings in Bioinformatics*, 12(5), 489–497. <https://doi.org/10.1093/bib/bbq077>
- Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., Das, K., Toy, T., Harry, B., Yourshaw, M., Fox, M., Fogel, B. L., Martinez-Agosto, J. A., Wong, D. A., Chang, V. Y., Shieh, P. B., Palmer, C. G. S., Dipple, K. M., Grody, W. W., ... Nelson, S. F. (2014). Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *JAMA*, 312(18), 1880–1887. <https://doi.org/10.1001/jama.2014.14604>
- Lee, H.-F., Chi, C.-S., & Tsai, C.-R. (2021). Diagnostic yield and treatment impact of whole-genome sequencing in paediatric neurological disorders. *Developmental Medicine & Child Neurology*, 63(8), 934–938. <https://doi.org/10.1111/dmcn.14722>
- Lee, J.-S., Lee, J. Y., Song, D. W., Bae, H. S., Doo, H. M., Yu, H. S., Lee, K. J., Kim, H. K., Hwang, H., Kwak, G., Kim, D., Kim, S., Hong, Y. B., Lee, J. M., & Choi, B.-O. (2020). Targeted PMP22 TATA-box editing by CRISPR/Cas9 reduces demyelinating neuropathy of Charcot-Marie-Tooth disease type 1A in mice. *Nucleic Acids Research*, 48(1), 130–140. <https://doi.org/10.1093/nar/gkz1070>
- Lek, M., & MacArthur, D. (2014). The Challenge of Next Generation Sequencing in the Context of Neuromuscular Diseases. *Journal of Neuromuscular Diseases*, 1(2), 135–149.
- Lemmers, R. J. L. F., Tawil, R., Petek, L. M., Balog, J., Block, G. J., Santen, G. W. E., Amell, A. M., van der Vliet, P. J., Almomani, R., Straasheijm, K. R., Krom, Y. D., Klooster, R., Sun, Y., den Dunnen, J. T., Helmer, Q., Donlin-Smith, C. M., Padberg, G. W., van Engelen, B. G. M., de Greef, J. C., ... van der Maarel, S. M. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-

- permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nature Genetics*, 44(12), 1370–1374. <https://doi.org/10.1038/ng.2454>
- Lemmers, R. J. L. F., van der Vliet, P. J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J. G., Snider, L., Straasheijm, K. R., van Ommen, G. J., Padberg, G. W., Miller, D. G., Tapscott, S. J., Tawil, R., Frants, R. R., & van der Maarel, S. M. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science*, 329(5999), 1650–1653. <https://doi.org/10.1126/science.1189044>
- Lewis, R. G., O’Shea, J. M., Pizzo, L., Wen, T., Fulmer, M. L., Zhao, J., Verheijen, J., Zhang, C., Velinder, M., Nicholas, T. J., Boyden, S. E., Ward, A., Baldwin, E. E., Andrews, A., Ruiz, J. H., Marchetti, M., Viskochil, D., Carey, J. C., Bleyl, S. B., ... Bayrak-Toydemir, P. (2025). RNA sequencing provides functional insights and diagnostic resolution in previously unsolved rare disease cases. *BMC Medical Genomics*, 18, 182. <https://doi.org/10.1186/s12920-025-02227-z>
- Lexell, E. M., Langdell, I., & Lexell, J. (2017). Vocational situation and experiences from the work environment among individuals with neuromuscular diseases. *Work (Reading, Mass.)*, 56(4), 519–530. <https://doi.org/10.3233/WOR-172527>
- Li, C., Vandersluis, S., Holubowich, C., Ungar, W. J., Goh, E. S., Boycott, K. M., Sikich, N., Dhalla, I., & Ng, V. (2021). Cost-effectiveness of genome-wide sequencing for unexplained developmental disabilities and multiple congenital anomalies. *Genetics in Medicine*, 23(3), 451–460. <https://doi.org/10.1038/s41436-020-01012-w>
- Li, H. (2026). *Lh3/minimap2* [C]. <https://github.com/lh3/minimap2> (Original work published 2017)
- Li, S., Han, X., Zhang, L., Xu, Y., Chang, C., Gao, L., Zhan, J., Hua, R., Mao, A., & Wang, Y. (2023). An Effective and Universal Long-Read Sequencing-Based Approach for SMN1 2 + 0 Carrier Screening through Family Trio Analysis. *Clinical Chemistry*, 69(11), 1295–1306. <https://doi.org/10.1093/clinchem/hvad152>
- Lim, K. R. Q., Nguyen, Q., & Yokota, T. (2020). DUX4 Signalling in the Pathogenesis of Facioscapulohumeral Muscular Dystrophy. *International Journal of Molecular Sciences*, 21(3), 729. <https://doi.org/10.3390/ijms21030729>

- Lim, W. F., & Rinaldi, C. (2023). RNA Transcript Diversity in Neuromuscular Research. *Journal of Neuromuscular Diseases*, 10(4), 473–482. <https://doi.org/10.3233/JND-221601>
- Lin, J., Tao, L., Deng, L., Zhou, R., Lou, S., Chen, S., Chen, X., Lu, C., Li, P., & Hu, B. (2023). Epigenome-wide DNA methylation analysis of myasthenia gravis. *FEBS Open Bio*, 13(7), 1375–1389. <https://doi.org/10.1002/2211-5463.13656>
- Lionel, A. C., Costain, G., Monfared, N., Walker, S., Reuter, M. S., Hosseini, S. M., Thiruvahindrapuram, B., Merico, D., Jobling, R., Nalpathamkalam, T., Pellicchia, G., Sung, W. W. L., Wang, Z., Bikangaga, P., Boelman, C., Carter, M. T., Cordeiro, D., Cytrynbaum, C., Dell, S. D., ... Marshall, C. R. (2018). Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genetics in Medicine*, 20(4), 435–443. <https://doi.org/10.1038/gim.2017.119>
- Liu, X., Wu, C., Li, C., & Boerwinkle, E. (2016). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Non-synonymous and Splice Site SNVs. *Human Mutation*, 37(3), 235–241. <https://doi.org/10.1002/humu.22932>
- Loewens, C. (2023). *Government of Canada invests \$15M in first-of-its-kind pan-Canadian genome library*. McGill University Newsroom. <https://www.mcgill.ca/newsroom/channels/news/government-canada-invests-15m-first-its-kind-pan-canadian-genome-library-352299>
- Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nature Reviews Genetics*, 21(10), 597–615. <https://doi.org/10.1038/s41576-020-0236-x>
- Long, J., Cui, D., Yu, C., & Meng, W. (2024). Evaluating the clinical efficacy of a long-read sequencing-based approach for carrier screening of spinal muscular atrophy. *Human Genomics*, 18, 110. <https://doi.org/10.1186/s40246-024-00676-8>
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J.,

- Magazine, H., Syron, J., ... Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6), 580–585. <https://doi.org/10.1038/ng.2653>
- Lu, D., Liu, C., Ji, W., Xia, R., Li, S., Liu, Y., Liu, N., Liu, Y., Deng, X. W., & Li, B. (2024). Nanopore ultra-long sequencing and adaptive sampling spur plant complete telomere-to-telomere genome assembly. *Molecular Plant*, 17(11), 1773–1786. <https://doi.org/10.1016/j.molp.2024.10.008>
- Luglio, A., Maggi, E., Riviello, F. N., Conforti, A., Sorrentino, U., & Zuccarello, D. (2024). Hereditary Neuromuscular Disorders in Reproductive Medicine. *Genes*, 15(11), 1409. <https://doi.org/10.3390/genes15111409>
- Lupski, J., & Garcia, C. (2001). Charcot-Marie-Tooth Peripheral Neuropathies and Related Disorders. In *The metabolic and molecular bases of inherited disease*. (pp. 5759–5788). McGraw Hill Medical.
- Lyle, R., Wright, T. J., Clark, L. N., & Hewitt, J. E. (1995). The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics*, 28(3), 389–397. <https://doi.org/10.1006/geno.1995.1166>
- Maestri, S., Maturo, M. G., Cosentino, E., Marcolungo, L., Iadarola, B., Fortunati, E., Rossato, M., & Delledonne, M. (2020). A Long-Read Sequencing Approach for Direct Haplotype Phasing in Clinical Settings. *International Journal of Molecular Sciences*, 21(23), 9177. <https://doi.org/10.3390/ijms21239177>
- Magliano, L., Patalano, M., Sagliocchi, A., Scutifero, M., Zaccaro, A., D’Angelo, M. G., Civati, F., Brighina, E., Vita, G., Vita, G. L., Messina, S., Sframeli, M., Pane, M., Lombardo, M. E., Scalise, R., D’Amico, A., Colia, G., Catteruccia, M., Balottin, U., ... Politano, L. (2014). “I have got something positive out of this situation”: Psychological benefits of caregiving in relatives of young people with muscular dystrophy. *Journal of Neurology*, 261(1), 188–195. <https://doi.org/10.1007/s00415-013-7176-8>

- Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. *Genome Biology*, 20, 246.
<https://doi.org/10.1186/s13059-019-1828-7>
- Mangin, A., de Pontual, L., Tsai, Y.-C., Monteil, L., Nizon, M., Boisseau, P., Mercier, S., Ziegler, J., Harting, J., Heiner, C., Gourdon, G., & Tomé, S. (2021). Robust Detection of Somatic Mosaicism and Repeat Interruptions by Long-Read Targeted Sequencing in Myotonic Dystrophy Type 1. *International Journal of Molecular Sciences*, 22(5), 2616. <https://doi.org/10.3390/ijms22052616>
- Mansfield, C., Boeri, M., Coulter, J., Baranowski, E., Sparks, S., An Haack, K., & Hamed, A. (2024). The value of knowing: Preferences for genetic testing to diagnose rare muscle diseases. *Orphanet Journal of Rare Diseases*, 19, 173. <https://doi.org/10.1186/s13023-024-03160-7>
- 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>
- Manzo, R., Annunziata, A., Calabrese, C., Marotta, A., & Fiorentino, G. (2025). Psychological aspects in neuromuscular patients: Case series. *Acta Myologica*, 44(2), 67–72.
<https://doi.org/10.36185/2532-1900-1168>
- Marchant, R. G., Bryen, S. J., Bahlo, M., Cairns, A., Chao, K. R., Corbett, A., Davis, M. R., Ganesh, V. S., Ghaoui, R., Jones, K. J., Kornberg, A. J., Lek, M., Liang, C., MacArthur, D. G., Oates, E. C., O'Donnell-Luria, A., O'Grady, G. L., Osei-Owusu, I. A., Rafehi, H., ... Cooper, S. T. (2024). Genome and RNA sequencing boost neuromuscular diagnoses to 62% from 34% with exome sequencing alone. *Annals of Clinical and Translational Neurology*, 11(5), 1250–1266.
<https://doi.org/10.1002/acn3.52041>
- Marian, A. J. (2014). Sequencing Your Genome: What Does It Mean? *Methodist DeBakey Cardiovascular Journal*, 10(1), 3–6. <https://doi.org/10.14797/mdcj-10-1-3>
- Martin, A. B. (2025). *Using long-read sequencing to diagnose rare neuromuscular disorders: A pilot study*. <http://hdl.handle.net/1993/39077>

- Martin, A., Llanes-Cuesta, M. A., Hartley, J. N., Frosk, P., Drögemöller, B. I., & Wright, G. E. B. (2026). *Identifying disease-causing mechanisms and fundamental biology of neuromuscular disorder genes through genomic feature analysis* (p. 2026.04.21.719902). bioRxiv.
<https://doi.org/10.64898/2026.04.21.719902>
- 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), NA-NA. <https://doi.org/10.1186/s13073-022-01026-w>
- Mary, P., Servais, L., & Vialle, R. (2018). Neuromuscular diseases: Diagnosis and management. *Orthopaedics & Traumatology: Surgery & Research, 2017 Instructional Course Lectures (SoFCOT)*, 104(1, Supplement), S89–S95. <https://doi.org/10.1016/j.otsr.2017.04.019>
- Mathieu, J., De Braekeleer, M., & Prévost, C. (1990). Genealogical reconstruction of myotonic dystrophy in the Saguenay-Lac-Saint-Jean area (Quebec, Canada). *Neurology*, 40(5), 839–842.
<https://doi.org/10.1212/wnl.40.5.839>
- Matthews, E., Brassington, R., Kuntzer, T., Jichi, F., & Manzur, A. Y. (2016). Corticosteroids for the treatment of Duchenne muscular dystrophy. *The Cochrane Database of Systematic Reviews*, 2016(5), CD003725. <https://doi.org/10.1002/14651858.CD003725.pub4>
- McCarl, C.-A., Picard, C., Khalil, S., Kawasaki, T., Röther, J., Papolos, A., Kutok, J., Hivroz, C., LeDeist, F., Plogmann, K., Ehl, S., Notheis, G., Albert, M. H., Belohradsky, B. H., Kirschner, J., Rao, A., Fischer, A., & Feske, S. (2009). ORAI1 deficiency and lack of store-operated Ca²⁺ entry cause immunodeficiency, myopathy and ectodermal dysplasia. *The Journal of Allergy and Clinical Immunology*, 124(6), 1311-1318.e7. <https://doi.org/10.1016/j.jaci.2009.10.007>
- McElwee, F., Sansom, S. L., Buchanan, J., Taylor, J. C., & Wordsworth, S. (2026). The cost and cost-effectiveness of whole-exome and whole-genome sequencing: A systematic literature review. *European Journal of Human Genetics*, 1–9. <https://doi.org/10.1038/s41431-026-02146-2>
- McMillan, H. J., Gerber, B., Cowling, T., Khuu, W., Mayer, M., Wu, J. W., Maturi, B., Klein-Panneton, K., Cabalteja, C., & Lochmüller, H. (2021). Burden of Spinal Muscular Atrophy (SMA) on

- Patients and Caregivers in Canada. *Journal of Neuromuscular Diseases*, 8(4), 553–568.
<https://doi.org/10.3233/JND-200610>
- Meade, O., Buchanan, H., & Coulson, N. (2018). The use of an online support group for neuromuscular disorders: A thematic analysis of message postings. *Disability and Rehabilitation*, 40(19), 2300–2310. <https://doi.org/10.1080/09638288.2017.1334239>
- Meienberg, J., Bruggmann, R., Oexle, K., & Matyas, G. (2016). Clinical sequencing: Is WGS the better WES? *Human Genetics*, 135, 359–362. <https://doi.org/10.1007/s00439-015-1631-9>
- Mekhuri, S., Dale, C., Buchanan, F., Hammash, N., Ambreen, M., Saha, S., Rose, L., & Amin, R. (2025). Family Caregivers of Individuals With Neuromuscular Disease Participating in a Randomized Controlled Trial of a Digital Peer Support Program: Nested Qualitative Study. *Journal of Medical Internet Research*, 27, e72141. <https://doi.org/10.2196/72141>
- Meng, L., Attali, R., Talmy, T., Regev, Y., Mizrahi, N., Smirin-Yosef, P., Vossaert, L., Taborda, C., Santana, M., Machol, I., Xiao, R., Dai, H., Eng, C., Xia, F., & Tzur, S. (2023). Evaluation of an automated genome interpretation model for rare disease routinely used in a clinical genetic laboratory. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 25(6), 100830. <https://doi.org/10.1016/j.gim.2023.100830>
- 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 Running title: Long-read WGS identifies causal SV 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>
- Miller, B. R., Powers, M., Park, M., Fischer, W., & Forbes, D. J. (2000a). Identification of a New Vertebrate Nucleoporin, Nup188, with the Use of a Novel Organelle Trap Assay. *Molecular Biology of the Cell*, 11(10), 3381–3396. <https://doi.org/10.1091/mbc.11.10.3381>

- Miller, B. R., Powers, M., Park, M., Fischer, W., & Forbes, D. J. (2000b). Identification of a New Vertebrate Nucleoporin, Nup188, with the Use of a Novel Organelle Trap Assay. *Molecular Biology of the Cell*, 11(10), 3381–3396. <https://doi.org/10.1091/mbc.11.10.3381>
- Miller, D. E., Sulovari, A., Wang, T., Loucks, H., Hoekzema, K., Munson, K. M., Lewis, A. P., Fuerte, E. P. A., Paschal, C. R., Walsh, T., Thies, J., Bennett, J. T., Glass, I., Dipple, K. M., Patterson, K., Bonkowski, E. S., Nelson, Z., Squire, A., Sikes, M., ... Eichler, E. E. (2021). Targeted long-read sequencing identifies missing disease-causing variation. *American Journal of Human Genetics*, 108(8), 1436–1449. <https://doi.org/10.1016/j.ajhg.2021.06.006>
- Mitsuhashi, S., & Matsumoto, N. (2020). Long-read sequencing for rare human genetic diseases. *Journal of Human Genetics*, 65(1), 11–19. <https://doi.org/10.1038/s10038-019-0671-8>
- Mohren, L., Erdlenbruch, F., Leitão, E., Kilpert, F., Hönes, G. S., Kaya, S., Schröder, C., Thieme, A., Sturm, M., Park, J., Schlüter, A., Ruiz, M., Morales de la Prida, M., Casasnovas, C., Becker, K., Roggenbuck, U., Pechlivanis, S., Kaiser, F. J., Synofzik, M., ... Depienne, C. (2024). Identification and characterisation of pathogenic and non-pathogenic FGF14 repeat expansions. *Nature Communications*, 15, 7665. <https://doi.org/10.1038/s41467-024-52148-1>
- Monaco, A. P., Bertelson, C. J., Liechti-Gallati, S., Moser, H., & Kunkel, L. M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics*, 2(1), 90–95. [https://doi.org/10.1016/0888-7543\(88\)90113-9](https://doi.org/10.1016/0888-7543(88)90113-9)
- Moon, D., & Seo, G. H. (2025). The importance and clinical utility of reanalysis of exome and genome sequencing data. *Kidney Research and Clinical Practice*. <https://doi.org/10.23876/j.krcp.25.210>
- Morin, G., Biancalana, V., Echaniz-Laguna, A., Noury, J.-B., Lornage, X., Moggio, M., Ripolone, M., Violano, R., Marcorelles, P., Maréchal, D., Renaud, F., Maurage, C.-A., Tard, C., Cuisset, J.-M., Laporte, J., & Böhm, J. (2020). Tubular aggregate myopathy and Stormorken syndrome: Mutation spectrum and genotype/phenotype correlation. *Human Mutation*, 41(1), 17–37. <https://doi.org/10.1002/humu.23899>

- Morin, G., Bruechle, N. O., Singh, A. R., Knopp, C., Jedraszak, G., Elbracht, M., Brémond-Gignac, D., Hartmann, K., Sevestre, H., Deutz, P., Hérent, D., Nürnberg, P., Roméo, B., Konrad, K., Mathieu-Dramard, M., Oldenburg, J., Bourges-Petit, E., Shen, Y., Zerres, K., ... Rochette, J. (2014). Gain-of-Function Mutation in STIM1 (P.R304W) Is Associated with Stormorken Syndrome. *Human Mutation*, 35(10), 1221–1232. <https://doi.org/10.1002/humu.22621>
- Mori-Yoshimura, M., Mizuno, Y., Yoshida, S., Ishihara, N., Minami, N., Morimoto, E., Maruo, K., Nonaka, I., Komaki, H., Nishino, I., Sekiguchi, M., Sato, N., Takeda, S., & Takahashi, Y. (2019). Psychiatric and neurodevelopmental aspects of Becker muscular dystrophy. *Neuromuscular Disorders: NMD*, 29(12), 930–939. <https://doi.org/10.1016/j.nmd.2019.09.006>
- Mortazavi, M., Guevara, J., Diaz, J., Tran, S., Ziaei Jam, H., Reeves, C., Batalov, S., Jepsen, K., Bainbridge, M., Besterman, A. D., Gymrek, M., Palmer, A. A., & Sebat, J. (2026). Long-read genome sequencing improves detection and functional interpretation of structural and repeat variants in autism. *Cell Genomics*, 6(5), 101186. <https://doi.org/10.1016/j.xgen.2026.101186>
- Moseley, M. L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A. K., Daughters, R. S., Chen, G., Weatherspoon, M. R., Clark, H. B., Ebner, T. J., Day, J. W., & Ranum, L. P. W. (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nature Genetics*, 38(7), 758-. Gale OneFile: Health and Medicine.
- Moshirfar, M., Webster, C. R., Seitz, T. S., Ronquillo, Y. C., & Hoopes, P. C. (2022). Ocular Features and Clinical Approach to Cataract and Corneal Refractive Surgery in Patients with Myotonic Dystrophy. *Clinical Ophthalmology (Auckland, N.Z.)*, 16, 2837–2842. <https://doi.org/10.2147/OPHTH.S372633>
- Moustakli, E., Christopoulos, P., Potiris, A., Zikopoulos, A., Mavrogianni, D., Karampas, G., Kathopoulis, N., Anagnostaki, I., Domali, E., Tzallas, A. T., Drakakis, P., & Stavros, S. (2025). Long-Read Sequencing and Structural Variant Detection: Unlocking the Hidden Genome in Rare Genetic Disorders. *Diagnostics*, 15(14), 1803. <https://doi.org/10.3390/diagnostics15141803>

- Muir, A. M., Cohen, J. L., Sheppard, S. E., Guttipatti, P., Lo, T. Y., Weed, N., Doherty, D., DeMarzo, D., Fagerberg, C. R., Kjærsgaard, L., Larsen, M. J., Rump, P., Löhner, K., Hirsch, Y., Zeevi, D. A., Zackai, E. H., Bhoj, E., Song, Y., & Mefford, H. C. (2020). Bi-allelic Loss-of-Function Variants in NUP188 Cause a Recognizable Syndrome Characterized by Neurologic, Ocular, and Cardiac Abnormalities. *American Journal of Human Genetics*, 106(5), 623–631.
<https://doi.org/10.1016/j.ajhg.2020.03.009>
- Müller, K., Oh, K.-W., Nordin, A., Panthi, S., Kim, S. H., Nordin, F., Freischmidt, A., Ludolph, A. C., Ki, C. S., Forsberg, K., Weishaupt, J., Kim, Y.-E., & Andersen, P. M. (2022). De novo mutations in SOD1 are a cause of ALS. *Journal of Neurology, Neurosurgery, and Psychiatry*, 93(2), 201–206.
<https://doi.org/10.1136/jnnp-2021-327520>
- Müller, U. (2021). Spinocerebellar ataxias (SCAs) caused by common mutations. *Neurogenetics*, 22(4), 235–250. <https://doi.org/10.1007/s10048-021-00662-5>
- Nakamura, R., Misawa, K., Tohnai, G., Nakatochi, M., Furuhashi, S., Atsuta, N., Hayashi, N., Yokoi, D., Watanabe, H., Watanabe, H., Katsuno, M., Izumi, Y., Kanai, K., Hattori, N., Morita, M., Taniguchi, A., Kano, O., Oda, M., Shibuya, K., ... Sobue, G. (2020). A multi-ethnic meta-analysis identifies novel genes, including ACSL5, associated with amyotrophic lateral sclerosis. *Communications Biology*, 3(1), 1–9. <https://doi.org/10.1038/s42003-020-01251-2>
- Narayanaswami, P., Weiss, M., Selcen, D., David, W., Raynor, E., Carter, G., Wicklund, M., Barohn, R. J., Ensrud, E., Griggs, R. C., Gronseth, G., & Amato, A. A. (2014). Evidence-based guideline summary: Diagnosis and treatment of limb-girdle and distal dystrophies. *Neurology*, 83(16), 1453–1463. <https://doi.org/10.1212/WNL.0000000000000892>
- Nemes, J. P., Benzow, K. A., & Koob, M. D. (2000). The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). *Human Molecular Genetics*, 9(10), 1543–1551. <https://doi.org/10.1093/hmg/9.10.1543>
- Nesin, V., Wiley, G., Kousi, M., Ong, E.-C., Lehmann, T., Nicholl, D. J., Suri, M., Shahrizaila, N., Katsanis, N., Gaffney, P. M., Wierenga, K. J., & Tsiokas, L. (2014). Activating mutations in

- STIM1 and ORAI1 cause overlapping syndromes of tubular myopathy and congenital miosis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(11), 4197–4202. <https://doi.org/10.1073/pnas.1312520111>
- Nettleton, S. (2006). “I just want permission to be ill”: Towards a sociology of medically unexplained symptoms. *Social Science & Medicine* (1982), 62(5), 1167–1178. <https://doi.org/10.1016/j.socscimed.2005.07.030>
- Ng, K. W. P., Chin, H.-L., Chin, A. X. Y., & Goh, D. L.-M. (2022). Using gene panels in the diagnosis of neuromuscular disorders: A mini-review. *Frontiers in Neurology*, 13, 997551. <https://doi.org/10.3389/fneur.2022.997551>
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., ... Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, 376(6588), 44–53. <https://doi.org/10.1126/science.abj6987>
- Orsini, M., Carolina, A., Ferreira, A. de F., de Assis, A. C. D., Magalhães, T., Teixeira, S., Bastos, V. H., Marinho, V., Oliveira, T., Fiorelli, R., Oliveira, A. B., & de Freitas, M. R. G. (2018). Cognitive impairment in neuromuscular diseases: A systematic review. *Neurology International*, 10(2), 7473. <https://doi.org/10.4081/ni.2018.7473>
- Osman, H., Adamji, Z., Pfeffer, G., Warman-Chardon, J., Varma, P., Keindel, J., & Lintern, S. (2025). From Doubt to Diagnosis: Canadian Patient Perspectives on a Limb-Girdle Muscular Dystrophy Diagnosis. *Health Expectations : An International Journal of Public Participation in Health Care and Health Policy*, 28(3), e70271. <https://doi.org/10.1111/hex.70271>
- Outram, S. M., Rego, S., Norstad, M., & Ackerman, S. (2024). The Need to Standardize the Reanalysis of Genomic Sequencing Results: Findings from Interviews with Underserved Families in Genomic Research. *Journal of Bioethical Inquiry*, 21(1), 95–104. <https://doi.org/10.1007/s11673-023-10267-2>

- PacBio. (n.d.). *Long-read sequencing at scale*. PacBio. Retrieved June 12, 2026, from <https://www.pacb.com/revio/>
- PacBio. (2026). *PacificBiosciences/paraphase* [Rust]. <https://github.com/PacificBiosciences/paraphase> (Original work published 2022)
- PacBio. (2026). *PacificBiosciences/pbmm2* [Raku]. <https://github.com/PacificBiosciences/pbmm2> (Original work published 2017)
- Pace, A., Roda, W., Poon, C., McMillan, H. J., Oskoui, M., MacKenzie, A., Chakraborty, P., & Round, J. (2025). A cost-utility analysis of newborn screening for spinal muscular atrophy in Canada. *Orphanet Journal of Rare Diseases*, 20, 428. <https://doi.org/10.1186/s13023-025-03927-6>
- Paganoni, S., Nicholson, K., Leigh, F., Swoboda, K., Chad, D., Drake, K., Haley, K., Cudkowicz, M., & Berry, J. D. (2017). Developing multidisciplinary clinics for neuromuscular care and research. *Muscle & Nerve*, 56(5), 848–858. <https://doi.org/10.1002/mus.25725>
- Pangalila, R. F., van den Bos, G. A. M., Stam, H. J., van Exel, N. J. A., Brouwer, W. B. F., & Roebroek, M. E. (2012). Subjective caregiver burden of parents of adults with Duchenne muscular dystrophy. *Disability and Rehabilitation*, 34(12), 988–996. <https://doi.org/10.3109/09638288.2011.628738>
- Paradis, C. M., Friedman, S., Lazar, R. M., & Kula, R. W. (1993). Anxiety disorders in a neuromuscular clinic. *The American Journal of Psychiatry*, 150(7), 1102–1104. <https://doi.org/10.1176/ajp.150.7.1102>
- Pardo-Palacios, F. J., Wang, D., Reese, F., Diekhans, M., Carbonell-Sala, S., Williams, B., Loveland, J. E., De María, M., Adams, M. S., Balderrama-Gutierrez, G., Behera, A. K., Gonzalez Martinez, J. M., Hunt, T., Lagarde, J., Liang, C. E., Li, H., Meade, M. J., Moraga Amador, D. A., Prjibelski, A. D., ... Brooks, A. N. (2024). Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nature Methods*, 21(7), 1349–1363. <https://doi.org/10.1038/s41592-024-02298-3>

- Park, J., Oh, H. M., Park, H. J., Cho, A.-R., Lee, D.-W., Jang, J.-H., & Jang, D.-H. (2019). Usefulness of comprehensive targeted multigene panel sequencing for neuromuscular disorders in Korean patients. *Molecular Genetics & Genomic Medicine*, 7(10), e00947. <https://doi.org/10.1002/mgg3.947>
- Patterson, J., Carpenter, E. J., Zhu, Z., An, D., Liang, X., Geng, C., Drmanac, R., & Wong, G. K.-S. (2019). Impact of sequencing depth and technology on de novo RNA-Seq assembly. *BMC Genomics*, 20, 604. <https://doi.org/10.1186/s12864-019-5965-x>
- Peddareddygar, L. R., Sobol, I., Pillai, H. B., Ito, M., Batish, S. D., & Grewal, R. P. (2013). Analysis of Association of Deletion in the Repeat Region of the Periaxin Gene With Late Onset Motor Neuropathy. *Journal of Neurology Research*, 2(6), 235–243. <https://doi.org/10.4021/jnr.v2i6.154>
- Pedersen, B. S., Brown, J. M., Dashnow, H., Wallace, A. D., Velinder, M., Tristani-Firouzi, M., Schiffman, J. D., Tvrdik, T., Mao, R., Best, D. H., Bayrak-Toydemir, P., & Quinlan, A. R. (2021). Effective variant filtering and expected candidate variant yield in studies of rare human disease. *NPJ Genomic Medicine*, 6, 60. <https://doi.org/10.1038/s41525-021-00227-3>
- 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. (2024b). *TCAG CNV/SV annotation pipeline* [Unpublished manuscript].
- Pellecchia, G. (2024a). *TCAG small variant annotation pipeline* [Unpublished manuscript].
- Pellerin, D., Danzi, M., Renaud, M., Houlden, H., Synofzik, M., Zuchner, S., & Brais, B. (2024). GAA-FGF14-Related Ataxia. In M. P. Adam, S. Bick, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, & A. Amemiya (Eds.), *GeneReviews®*. University of Washington, Seattle. <http://www.ncbi.nlm.nih.gov/books/NBK599589/>

- Pellerin, D., Del Gobbo, G., Couse, M., Dolzhenko, E., Dicaire, M.-J., Rebelo, A., Roth, V., Wandzel, M., Bonnet, C., Ashton, C., Lamont, P. J., Laing, N. G., Renaud, M., Ravenscroft, G., Houlden, H., Synofzik, M., Eberle, M. A., Boycott, K. M., Pastinen, T., ... Danzi, M. C. (2023). *A common flanking variant is associated with enhanced meiotic stability of the FGF14-SCA27B locus* (p. 2023.05.11.540430). bioRxiv. <https://doi.org/10.1101/2023.05.11.540430>
- Pera, M. C., Coratti, G., Berti, B., D'Amico, A., Sframeli, M., Albamonte, E., de Sanctis, R., Messina, S., Catteruccia, M., Brigati, G., Antonaci, L., Lucibello, S., Bruno, C., Sansone, V. A., Bertini, E., Tiziano, D., Pane, M., & Mercuri, E. (2020). Diagnostic journey in Spinal Muscular Atrophy: Is it still an odyssey? *PLoS ONE*, *15*(3), e0230677. <https://doi.org/10.1371/journal.pone.0230677>
- Perez, B. A., Shorrock, H. K., Banez-Coronel, M., Zu, T., Romano, L. E., Laboissonniere, L. A., Reid, T., Ikeda, Y., Reddy, K., Gomez, C. M., Bird, T., Ashizawa, T., Schut, L. J., Brusco, A., Berglund, J. A., Hasholt, L. F., Nielsen, J. E., Subramony, S., & Ranum, L. P. (2021). CCG•CGG interruptions in high-penetrance SCA8 families increase RAN translation and protein toxicity. *EMBO Molecular Medicine*, *13*(11), e14095. <https://doi.org/10.15252/emmm.202114095>
- 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>
- Perez, P. G. (2024). Neuromuscular Notes: Genetic Testing for Neuromuscular Diseases. *Practical Neurology*. <https://practicalneurology.com/diseases-diagnoses/neuromuscular/neuromuscular-notes-genetic-testing-for-neuromuscular-diseases/32181/>
- Peric', S., Rakocevic-Stojanovic, V., Stevic, Z., Basta, I., Pavlovic, S., Vujanac, V., Marjanovic, L., & Lavrnic, D. (2010). Health-related quality of life in patients with myotonic dystrophy type 1 and amyotrophic lateral sclerosis. *Acta Neurologica Belgica*, *110*(1), 71–77.

- Perrin, A., Goethem, C. V., Thèze, C., Puechberty, J., Guignard, T., Lecardonnel, B., Lacourt, D., Métay, C., Isapof, A., Whalen, S., Ferreira, A., Arne-Bes, M.-C., Quijano-Roy, S., Nectoux, J., Leturcq, F., Richard, P., Larrieux, M., Bergougnoux, A., Pellestor, F., ... 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>
- Perrin, C., Unterborn, J. N., Ambrosio, C. D., & Hill, N. S. (2004). Pulmonary complications of chronic neuromuscular diseases and their management. *Muscle & Nerve*, 29(1), 5–27. <https://doi.org/10.1002/mus.10487>
- Piñeros-Fernández, M. C., Morte, B., & García-Giménez, J. L. (2024). Utility of exome sequencing for the diagnosis of pediatric-onset neuromuscular diseases beyond diagnostic yield: A narrative review. *Neurological Sciences*, 45(4), 1455–1464. <https://doi.org/10.1007/s10072-023-07210-z>
- 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>
- Pommié, C., Levadoux, S., Sabatier, R., Lefranc, G., & Lefranc, M.-P. (2004). IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *Journal of Molecular Recognition: JMR*, 17(1), 17–32. <https://doi.org/10.1002/jmr.647>
- Posey, J. E., Rosenfeld, J. A., James, R. A., Bainbridge, M., Niu, Z., Wang, X., Dhar, S., Wiszniewski, W., Akdemir, Z. H. C., Gambin, T., Xia, F., Person, R. E., Walkiewicz, M., Shaw, C. A., Sutton, V. R., Beaudet, A. L., Muzny, D., Eng, C. M., Yang, Y., ... Plon, S. E. (2016). Molecular Diagnostic Experience of Whole-Exome Sequencing in Adult Patients. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, 18(7), 678–685. <https://doi.org/10.1038/gim.2015.142>
- Preston, M. K., & Wang, L. H. (1993). Facioscapulohumeral Muscular Dystrophy. In M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, & A. Amemiya (Eds.), *GeneReviews®*. University of Washington, Seattle. <http://www.ncbi.nlm.nih.gov/books/NBK1443/>

- 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>
- Price, T. R., Hodgkinson, V., Westbury, G., Korngut, L., Innes, M. A., Marshall, C. R., Nelson, T. N., Huang, L., Parboosingh, J., & Mah, J. K. (2024). A Study on the Incidence and Prevalence of 5q Spinal Muscular Atrophy in Canada Using Multiple Data Sources. *The Canadian Journal of Neurological Sciences. Le Journal Canadien Des Sciences Neurologiques*, 51(5), 660–671. <https://doi.org/10.1017/cjn.2024.1>
- Qin, D. (2019). Next-generation sequencing and its clinical application. *Cancer Biology & Medicine*, 16(1), 4–10. <https://doi.org/10.20892/j.issn.2095-3941.2018.0055>
- Qin, L., Wang, J., Tian, X., Yu, H., Truong, C., Mitchell, J. J., Wierenga, K. J., Craigen, W. J., Zhang, V. W., & Wong, L.-J. C. (2016). Detection and Quantification of Mosaic Mutations in Disease Genes by Next-Generation Sequencing. *The Journal of Molecular Diagnostics: JMD*, 18(3), 446–453. <https://doi.org/10.1016/j.jmoldx.2016.01.002>
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. (Version v4.3.3) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Regier, D. A., Loewen, R., Chan, B., Ehman, M., Pollard, S., Friedman, J. M., Stockler-Ipsiroglu, S., Karnebeek, C. van, Race, S., Elliott, A. M., Dragojlovic, N., Lynd, L. D., & Weymann, D. (2024). Real-world diagnostic outcomes and cost-effectiveness of genome-wide sequencing for developmental and seizure disorders: Evidence from Canada. *Genetics in Medicine*, 26(4). <https://doi.org/10.1016/j.gim.2024.101069>
- Reimann, J., & Kornblum, C. (2020). Towards Central Nervous System Involvement in Adults with Hereditary Myopathies. *Journal of Neuromuscular Diseases*, 7(4), 367–393. <https://doi.org/10.3233/JND-200507>

- Ren, J., Luo, S., Shi, H., & Wang, X. (2024). Spatial omics advances for in situ RNA biology. *Molecular Cell*, 84(19), 3737–3757. <https://doi.org/10.1016/j.molcel.2024.08.002>
- 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(Database issue), D886–D894. <https://doi.org/10.1093/nar/gky1016>
- Ricci, G., Govoni, A., Torri, F., Astrea, G., Buchignani, B., Marinella, G., Battini, R., Manca, M. L., Castiglione, V., Giannoni, A., Emdin, M., & Siciliano, G. (2024). Characterization of Phenotypic Variability in Becker Muscular Dystrophy for Clinical Practice and Towards Trial Readiness: A Two-Years Follow up Study. *Journal of Neuromuscular Diseases*, 11(2), 375–387. <https://doi.org/10.3233/JND-221513>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehman, H. L. (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>
- Riggs, E. R., Andersen, E. F., Cherry, A. M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D. I., South, S. T., Thorland, E. C., Pineda-Alvarez, D., Aradhya, S., & Martin, C. L. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genetics in Medicine*, 22(2), 245–257. <https://doi.org/10.1038/s41436-019-0686-8>
- Rizzo, F., Bono, S., Ruepp, M. D., Salani, S., Ottoboni, L., Abati, E., Melzi, V., Cordiglieri, C., Pagliarini, S., De Gioia, R., Anastasia, A., Taiana, M., Garbellini, M., Lodato, S., Kunderfranco, P., Cazzato, D., Cartelli, D., Lonati, C., Bresolin, N., ... Corti, S. (2023). Combined RNA interference and gene replacement therapy targeting MFN2 as proof of principle for the treatment of Charcot–

- Marie–Tooth type 2A. *Cellular and Molecular Life Sciences: CMLS*, 80(12), 373.
<https://doi.org/10.1007/s00018-023-05018-w>
- Roberts, A. M., Ware, J. S., Herman, D. S., Schafer, S., Baksi, J., Bick, A. G., Buchan, R. J., Walsh, R., John, S., Wilkinson, S., Mazzarotto, F., Felkin, L. E., Gong, S., L. MacArthur, J. A., Cunningham, F., Flannick, J., Gabriel, S. B., Altshuler, D. M., Macdonald, P. S., ... Cook, S. A. (2015). Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Science Translational Medicine*, 7(270), 270ra6-270ra6.
<https://doi.org/10.1126/scitranslmed.3010134>
- 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. *The American Journal of Human Genetics*, 83(5), 610–615.
<https://doi.org/10.1016/j.ajhg.2008.09.017>
- Roggenbuck, J., Eubank, B. H. F., Wright, J., Harms, M. B., & Kolb, S. J. (2023). Evidence-based consensus guidelines for ALS genetic testing and counseling. *Annals of Clinical and Translational Neurology*, 10(11), 2074–2091. <https://doi.org/10.1002/acn3.51895>
- Rohlin, A., Wernersson, J., Engwall, Y., Wiklund, L., Björk, J., & Nordling, M. (2009). Parallel sequencing used in detection of mosaic mutations: Comparison with four diagnostic DNA screening techniques. *Human Mutation*, 30(6), 1012–1020. <https://doi.org/10.1002/humu.20980>
- Rose, L., Fowler, R. A., Fan, E., Fraser, I., Leasa, D., Mawdsley, C., Pedersen, C., Rubinfeld, G., & CANuVENT group. (2015). Prolonged mechanical ventilation in Canadian intensive care units: A national survey. *Journal of Critical Care*, 30(1), 25–31. <https://doi.org/10.1016/j.jcrc.2014.07.023>
- Rose, L., McKim, D., Leasa, D., Nonoyama, M., Tandon, A., Bai, Y. Q., Amin, R., Katz, S., Goldstein, R., & Gershon, A. (2019). Trends in incidence, prevalence, and mortality of neuromuscular disease in

- Ontario, Canada: A population-based retrospective cohort study (2003-2014). *PLoS ONE*, 14(3), e0210574. <https://doi.org/10.1371/journal.pone.0210574>
- Rosenberg, A., Tian, C., He, H., Ulm, E., Collins Ruff, K., & B. Nagaraj, C. (2023). An evaluation of clinical presentation and genetic testing approaches for patients with neuromuscular disorders. *American Journal of Medical Genetics Part A*, 191(11), 2679–2692. <https://doi.org/10.1002/ajmg.a.63356>
- Ross, J. P., Dion, P. A., & Rouleau, G. A. (2020). Exome sequencing in genetic disease: Recent advances and considerations. *F1000Research*, 9, F1000 Faculty Rev-336. <https://doi.org/10.12688/f1000research.19444.1>
- Rudaks, L. I., Stevanovski, I., Yeow, D., Reis, A. L. M., Chintalaphani, S. R., Cheong, P. L., Gamaarachchi, H., Worgan, L., Ahmad, K., Hayes, M., Hannaford, A., Kim, S., Fung, V. S. C., Halmagyi, G. M., Martin, A., Manser, D., Tchan, M., Ng, K., Kennerson, M. L., ... Kumar, K. R. (2025). Targeted Long-Read Sequencing as a Single Assay Improves the Diagnosis of Spastic-Ataxia Disorders. *Annals of Clinical and Translational Neurology*, 12(4), 832–841. <https://doi.org/10.1002/acn3.70008>
- Ruggieri, A., Ramachandran, N., Wang, P., Haan, E., Kneebone, C., Manavis, J., Morandi, L., Moroni, I., Blumbergs, P., Mora, M., & Minassian, B. A. (2015). Non-coding VMA21 deletions cause X-linked myopathy with excessive autophagy. *Neuromuscular Disorders: NMD*, 25(3), 207–211. <https://doi.org/10.1016/j.nmd.2014.11.014>
- Sagath, L., Kiiski, K., Naidu, K., Patel, K., Jonson, P. H., Laarne, M., Djordjevic, D., Yoon, G., LaGroon, A., Rogers, C., Galindo, M. K., Scherer, K., Kunstmann, E., Koparir, E., Ho, D., Davis, M., Joshi, P., Zygmunt, A., Orbach, R., ... Lehtokari, V.-L. (2025). Structural variation in nebulin and its impact on phenotype and inheritance: Establishing a dominant distal phenotype caused by large deletions. *European Journal of Human Genetics*, 33(9), 1153–1162. <https://doi.org/10.1038/s41431-025-01891-0>

- Salmon, K., Kiernan, M. C., Kim, S. H., Andersen, P. M., Chio, A., van den Berg, L. H., Van Damme, P., Al-Chalabi, A., Lillo, P., Andrews, J. A., & Genge, A. (2022). The importance of offering early genetic testing in everyone with amyotrophic lateral sclerosis. *Brain*, *145*(4), 1207–1210.
<https://doi.org/10.1093/brain/awab472>
- Sandestig, A., Engström, K., Pepler, A., Danielsson, I., Odelberg-Johnsson, P., Biskup, S., Holz, A., & Stefanova, M. (2020). NUP188 Biallelic Loss of Function May Underlie a New Syndrome: Nucleoporin 188 Insufficiency Syndrome? *Molecular Syndromology*, *10*(6), 313–319.
<https://doi.org/10.1159/000504818>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Sarparanta, J., Jonson, P. H., Golzio, C., Sandell, S., Luque, H., Screen, M., McDonald, K., Stajich, J. M., Mahjneh, I., Vihola, A., Raheem, O., Penttilä, S., Lehtinen, S., Huovinen, S., Palmio, J., Tasca, G., Ricci, E., Hackman, P., Hauser, M., ... Udd, B. (2012). Mutations affecting the cytoplasmic functions of the co-chaperone DNAJB6 cause limb-girdle muscular dystrophy. *Nature Genetics*, *44*(4), 450–455, S1-2. <https://doi.org/10.1038/ng.1103>
- Saunders, C. J., Miller, N. A., Soden, S. E., Dinwiddie, D. L., Noll, A., Alnadi, N. A., Andrews, N., Patterson, M. L., Krivohlavek, L. A., Fellis, J., Humphray, S., Saffrey, P., Kingsbury, Z., Weir, J. C., Betley, J., Grocock, R. J., Margulies, E. H., Farrow, E. G., Artman, M., ... Kingsmore, S. F. (2012). Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Science Translational Medicine*, *4*(154), 154ra135.
<https://doi.org/10.1126/scitranslmed.3004041>
- Savarese, M., Maggi, L., Vihola, A., Jonson, P. H., Tasca, G., Ruggiero, L., Bello, L., Magri, F., Giugliano, T., Torella, A., Evilä, A., Di Fruscio, G., Vanakker, O., Gibertini, S., Vercelli, L., Ruggieri, A., Antozzi, C., Luque, H., Janssens, S., ... Nigro, V. (2018). Interpreting Genetic

- Variants in Titin in Patients With Muscle Disorders. *JAMA Neurology*, 75(5), 557–565.
<https://doi.org/10.1001/jamaneurol.2017.4899>
- Scherer, S. S., Xu, Y. T., Bannerman, P. G., Sherman, D. L., & Brophy, P. J. (1995). Periaxin expression in myelinating Schwann cells: Modulation by axon-glial interactions and polarized localization during development. *Development*, 121(12), 4265–4273. <https://doi.org/10.1242/dev.121.12.4265>
- Schiava, M., Lofra, R. M., Bourke, J. P., Díaz-Manera, J., James, M. K., Elseed, M. A., Malinova, M., Michel-Sodhi, J., Moat, D., Ghimenton, E., Mccallum, M., Díaz, C. F. B., Mayhew, A., Wong, K., Richardson, M., Tasca, G., Eglon, G., Eagle, M., Turner, C., ... Guglieri, M. (2024). Functional abilities, respiratory and cardiac function in a large cohort of adults with Duchenne muscular dystrophy treated with glucocorticoids. *European Journal of Neurology*, 31(6), e16267.
<https://doi.org/10.1111/ene.16267>
- Schobers, G., Derks, R., den Ouden, A., Swinkels, H., van Reeuwijk, J., Bosgoed, E., Lugtenberg, D., Sun, S. M., Corominas Galbany, J., Weiss, M., Blok, M. J., Olde Keizer, R. A. C. M., Hofste, T., Hellebrekers, D., de Leeuw, N., Stegmann, A., Kamsteeg, E.-J., Paulussen, A. D. C., Ligtenberg, M. J. L., ... Vissers, L. E. L. M. (2024). Genome sequencing as a generic diagnostic strategy for rare disease. *Genome Medicine*, 16(1), 32. <https://doi.org/10.1186/s13073-024-01301-y>
- Schobers, G., Schieving, J. H., Yntema, H. G., Pennings, M., Pfundt, R., Derks, R., Hofste, T., de Wijs, I., Wieskamp, N., van den Heuvel, S., Galbany, J. C., Gilissen, C., Nelen, M., Brunner, H. G., Kleefstra, T., Kamsteeg, E.-J., Willemsen, M. A. A. P., & Vissers, L. E. L. M. (2022). Reanalysis of exome negative patients with rare disease: A pragmatic workflow for diagnostic applications. *Genome Medicine*, 14(1), 66. <https://doi.org/10.1186/s13073-022-01069-z>
- Schofield, D., Alam, K., Douglas, L., Shrestha, R., MacArthur, D. G., Davis, M., Laing, N. G., Clarke, N. F., Burns, J., Cooper, S. T., North, K. N., Sandaradura, S. A., & O'Grady, G. L. (2017). Cost-effectiveness of massively parallel sequencing for diagnosis of paediatric muscle diseases. *NPJ Genomic Medicine*, 2, 4. <https://doi.org/10.1038/s41525-017-0006-7>

- Schön, U., Holzer, A., Laner, A., Kleinle, S., Scharf, F., Benet-Pagès, A., Peschel, O., Holinski-Feder, E., & Diebold, I. (2021). HPO-driven virtual gene panel: A new efficient approach in molecular autopsy of sudden unexplained death. *BMC Medical Genomics*, *14*, 94.
<https://doi.org/10.1186/s12920-021-00946-7>
- Schreyer, L., Reilly, J., McConkey, H., Kerkhof, J., Levy, M. A., Hu, J., Hnaini, M., Sadikovic, B., & Campbell, C. (2023). The discovery of the DNA methylation epsignature for Duchenne muscular dystrophy. *Neuromuscular Disorders*, *33*(1), 5–14. <https://doi.org/10.1016/j.nmd.2022.12.003>
- Scott, M. O., Sylvester, J. E., Heiman-Patterson, T., Shi, Y. J., Fieles, W., Stedman, H., Burghes, A., Ray, P., Worton, R., & Fischbeck, K. H. (1988). Duchenne muscular dystrophy gene expression in normal and diseased human muscle. *Science*, *239*(4846), 1418–1420.
<https://doi.org/10.1126/science.2450401>
- Sedlazeck, F. J., Lee, H., Darby, C. A., & Schatz, M. C. (2018). Piercing the dark matter: Bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics*, *19*(6), 329–346.
<https://doi.org/10.1038/s41576-018-0003-4>
- Segarra-Casas, A., Yépez, V. A., Demidov, G., Laurie, S., Esteve-Codina, A., Gagneur, J., Parkhurst, Y., Muni-Lofra, R., Harris, E., Marini-Bettolo, C., Straub, V., & Töpf, A. (2024). An Integrated Transcriptomics and Genomics Approach Detects an X/Autosome Translocation in a Female with Duchenne Muscular Dystrophy. *International Journal of Molecular Sciences*, *25*(14), 7793.
<https://doi.org/10.3390/ijms25147793>
- Sen, S., Handler, H. P., Victorsen, A., Flaten, Z., Ellison, A., Knutson, T. P., Munro, S. A., Martinez, R. J., Billington, C. J., Laffin, J. J., Bray, S., Mroz, P., Yohe, S., Nelson, A. C., Bower, M., & Thyagarajan, B. (2025). Validation of a comprehensive long-read sequencing platform for broad clinical genetic diagnosis. *Frontiers in Genetics*, *16*, 1499456.
<https://doi.org/10.3389/fgene.2025.1499456>
- Seo, G. H., Lee, H., Lee, J., Han, H., Cho, Y. K., Kim, M., Choi, Y., Choi, J., Choi, I. H., Rhie, S., Chae, K. Y., Kim, Y.-M., Cheon, C. K., Kim, S. J., Lee, J., Kang, E., Byeon, J. H., Yu, H. J., Shin, Y.-L.,

- ... Eun, B.-L. (2022). Diagnostic performance of automated, streamlined, daily updated exome analysis in patients with neurodevelopmental delay. *Molecular Medicine*, 28(1), 38.
<https://doi.org/10.1186/s10020-022-00464-x>
- Serano, M., Fiore, F., Sorrentino, V., & Rossi, D. (2026). Tubular Aggregate Myopathies: Genetic Heterogeneity and Diverse Clinical Features Converging on Calcium Dysregulation. *Cells*, 15(7), 635. <https://doi.org/10.3390/cells15070635>
- Shatunov, A., Olivé, M., Odgerel, Z., Stadelmann-Nessler, C., Irlbacher, K., van Landeghem, F., Bayarsaikhan, M., Lee, H.-S., Goudeau, B., Chinnery, P. F., Straub, V., Hilton-Jones, D., Damian, M. S., Kaminska, A., Vicart, P., Bushby, K., Dalakas, M. C., Sambuughin, N., Ferrer, I., ... Goldfarb, L. G. (2009). In-frame deletion in the seventh immunoglobulin-like repeat of filamin C in a family with myofibrillar myopathy. *European Journal of Human Genetics*, 17(5), 656–663.
<https://doi.org/10.1038/ejhg.2008.226>
- Sherman, D. L., Fabrizi, C., Gillespie, C. S., & Brophy, P. J. (2001). Specific disruption of a schwann cell dystrophin-related protein complex in a demyelinating neuropathy. *Neuron*, 30(3), 677–687.
[https://doi.org/10.1016/s0896-6273\(01\)00327-0](https://doi.org/10.1016/s0896-6273(01)00327-0)
- Shortt, J. A., Ruggiero, R. P., Cox, C., Wacholder, A. C., & Pollock, D. D. (2020). Finding and extending ancient simple sequence repeat-derived regions in the human genome. *Mobile DNA*, 11(1), 11.
<https://doi.org/10.1186/s13100-020-00206-y>
- Sidlauskaite, E., Le Gall, L., Mariot, V., & Dumonceaux, J. (2020). DUX4 Expression in FSHD Muscles: Focus on Its mRNA Regulation. *Journal of Personalized Medicine*, 10(3), 73.
<https://doi.org/10.3390/jpm10030073>
- Silent Genomes Project*. (n.d.). B.C Children's Hospital Research Institute. Retrieved July 30, 2026, from <https://www.bcchr.ca/silent-genomes-project>
- Silveira, I., Alonso, I., Guimarães, L., Mendonça, P., Santos, C., Maciel, P., Fidalgo de Matos, J. M., Costa, M., Barbot, C., Tuna, A., Barros, J., Jardim, L., Coutinho, P., & Sequeiros, J. (2000). High

- Germinal Instability of the (CTG)_n at the SCA8 Locus of Both Expanded and Normal Alleles. *American Journal of Human Genetics*, 66(3), 830–840. <https://doi.org/10.1086/302827>
- Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40(Web Server issue), W452–457. <https://doi.org/10.1093/nar/gks539>
- Skalsky, A., & Scott-Wyrd, P. (2023). Physical Therapy, Bracing and Surgical Treatment in Muscular Dystrophies. In P. Narayanaswami & T. Liewluck (Eds.), *Principles and Practice of the Muscular Dystrophies* (pp. 303–329). Springer International Publishing. https://doi.org/10.1007/978-3-031-44009-0_19
- Smedley, D., Smith, K. R., Martin, A., Thomas, E. A., McDonagh, E. M., Cipriani, V., Ellingford, J. M., Arno, G., Tucci, A., Vandrovcova, J., Chan, G., Williams, H. J., Ratnaike, T., Wei, W., Stirrups, K., Ibanez, K., Moutsianas, L., Wielscher, M., Need, A., ... Caulfield, M. (2021). 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care—Preliminary Report. *The New England Journal of Medicine*, 385(20), 1868–1880. <https://doi.org/10.1056/NEJMoa2035790>
- Smirnov, D., Konstantinovskiy, N., & Prokisch, H. (2023). Integrative omics approaches to advance rare disease diagnostics. *Journal of Inherited Metabolic Disease*, 46(5), 824–838. <https://doi.org/10.1002/jimd.12663>
- Smith, I. C., Abusetah, Y., Osman, H., Garg, A., Grant, A., Lochmuller, H., McMillan, H., Pfeffer, G., Korngut, L., Gagnon, C., Lintern, S., Wojtal, D., Selby, K., Thavorn, K., & Warman-Chardon, J. (2025). Assessing the socio-economic burden of inherited and inflammatory neuromuscular diseases (BIND study): A study protocol. *Orphanet Journal of Rare Diseases*, 20(1), 404. <https://doi.org/10.1186/s13023-025-03904-z>
- Sone, J., Mitsuhashi, S., Fujita, A., Mizuguchi, T., Hamanaka, K., Mori, K., & Koike, H. (2019). Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nature Genetics*, 51(8), 1215–1215. <https://doi.org/10.1038/s41588-019-0459-y>

- Srivastava, S., Love-Nichols, J. A., Dies, K. A., Ledbetter, D. H., Martin, C. L., Chung, W. K., Firth, H. V., Frazier, T., Hansen, R. L., Prock, L., Brunner, H., Hoang, N., Scherer, S. W., Sahin, M., & Miller, D. T. (2019). Meta-analysis and multidisciplinary consensus statement: Exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genetics in Medicine*, 21(11), 2413–2421. <https://doi.org/10.1038/s41436-019-0554-6>
- Stavropoulos, D. J., Merico, D., Jobling, R., Bowdin, S., Monfared, N., Thiruvahindrapuram, B., Nalpathamkalam, T., Pellecchia, G., Yuen, R. K. C., Szego, M. J., Hayeems, R. Z., Shaul, R. Z., Brudno, M., Girdea, M., Frey, B., Alipanahi, B., Ahmed, S., Babul-Hirji, R., Porras, R. B., ... Marshall, C. R. (2016). Whole Genome Sequencing Expands Diagnostic Utility and Improves Clinical Management in Pediatric Medicine. *NPJ Genomic Medicine*, 1, 15012-. <https://doi.org/10.1038/npjgenmed.2015.12>
- Stavrou, M., Kagiava, A., Choudury, S. G., Jennings, M. J., Wallace, L. M., Fowler, A. M., Heslegrave, A., Richter, J., Tryfonos, C., Christodoulou, C., Zetterberg, H., Horvath, R., Harper, S. Q., & Kleopa, K. A. (2022). A translatable RNAi-driven gene therapy silences PMP22/Pmp22 genes and improves neuropathy in CMT1A mice. *The Journal of Clinical Investigation*, 132(13), e159814. <https://doi.org/10.1172/JCI159814>
- Stavrou, M., Kagiava, A., Sargiannidou, I., Georgiou, E., & Kleopa, K. A. (2023). Charcot–Marie–Tooth neuropathies: Current gene therapy advances and the route toward translation. *Journal of the Peripheral Nervous System*, 28(2), 150–168. <https://doi.org/10.1111/jns.12543>
- 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>
- Stevanovski, I., Chintalaphani, S. R., Gamaarachchi, H., Ferguson, J. M., Pineda, S. S., Scriba, C. K., Tchan, M., Fung, V., Ng, K., Cortese, A., Houlden, H., Dobson-Stone, C., Fitzpatrick, L.,

- Halliday, G., Ravenscroft, G., Davis, M. R., Laing, N. G., Fellner, A., Kennerson, M., ... Deveson, I. W. (2022). Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Science Advances*, 8(9), eabm5386. <https://doi.org/10.1126/sciadv.abm5386>
- Steyaert, W., Sagath, L., Demidov, G., Yépez, V. A., Esteve-Codina, A., Gagneur, J., Ellwanger, K., Derks, R., Weiss, M., den Ouden, A., van den Heuvel, S., Swinkels, H., Zomer, N., Steehouwer, M., O’Gorman, L., Astuti, G., Neveling, K., Schüle, R., Xu, J., ... Solve-RD consortium. (2025). Unraveling undiagnosed rare disease cases by HiFi long-read genome sequencing. *Genome Research*, 35(4), 755–768. <https://doi.org/10.1101/gr.279414.124>
- Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Hsi-Yang Fritz, M., Konkel, M. K., Malhotra, A., Stütz, A. M., Shi, X., Paolo Casale, F., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., ... Korbel, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571), 75–81. <https://doi.org/10.1038/nature15394>
- Tagimacruz, T., Seeger, T. A., Degeling, K., Fooks, K., Venkataramanan, V., Bernier, F. P., Boycott, K. M., Mendoza-Londono, R., Hartley, T., Hayeems, R. Z., & Marshall, D. A. (2026). Clinical utility of exome sequencing: Post-exome testing decision changes in the management of children with suspected rare genetic disease. *Genetics in Medicine*, 28(9). <https://doi.org/10.1016/j.gim.2026.102626>
- Tardy, C., Trani, J. P., Pienkowski, V. M., Morin, L., Castro, C., Souville, L., Humbert, C., Gérard, L., Eudes, N., Assoumani, A., Bertaux, K., Verebi, C., Nectoux, J., Campana, E. S., Jacquemont, M. L., Toutain, A., Mallaret, M., Tard, C., Fradin, M., ... Magdinier, F. (2025). *Comparison of long-read sequencing strategies for resolving complex genotypes at Facioscapulohumeral dystrophy-associated loci* (p. 2025.08.08.25333030). medRxiv. <https://doi.org/10.1101/2025.08.08.25333030>

- Taylor, J. C., Martin, H. C., Lise, S., Broxholme, J., Cazier, J.-B., Rimmer, A., Kanapin, A., Lunter, G., Fiddy, S., Allan, C., Aricescu, A. R., Attar, M., Babbs, C., Becq, J., Beeson, D., Bento, C., Bignell, P., Blair, E., Buckle, V. J., ... McVean, G. (2015). Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nature Genetics*, 47(7), 717–726. <https://doi.org/10.1038/ng.3304>
- Tesi, N., Salazar, A., Zhang, Y., van der Lee, S., Hulsman, M., Knoop, L., Wijesekera, S., Krizova, J., Schneider, A.-F., Pennings, M., Slegers, K., Kamsteeg, E.-J., Reinders, M., & Holstege, H. (2024). Characterizing tandem repeat complexities across long-read sequencing platforms with TREAT and otter. *Genome Research*, 34(11), 1942–1953. <https://doi.org/10.1101/gr.279351.124>
- Tewhey, R., Bansal, V., Torkamani, A., Topol, E. J., & Schork, N. J. (2011). The importance of phase information for human genomics. *Nature Reviews. Genetics*, 12(3), 215–223. <https://doi.org/10.1038/nrg2950>
- Thada, P. K., Bhandari, J., Forshaw, K. C., & Umapathi, K. K. (2024). Becker Muscular Dystrophy. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK556092/>
- 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](https://doi.org/10.1016/S1474-4422(20)30028-4)
- Tian, X., Liang, W.-C., Feng, Y., Wang, J., Zhang, V. W., Chou, C.-H., Huang, H.-D., Lam, C. W., Hsu, Y.-Y., Lin, T.-S., Chen, W.-T., Wong, L.-J., & Jong, Y.-J. (2015). Expanding genotype/phenotype of neuromuscular diseases by comprehensive target capture/NGS. *Neurology: Genetics*, 1(2), e14. <https://doi.org/10.1212/NXG.0000000000000015>
- Torri, F., Lopriore, P., Montano, V., Siciliano, G., Mancuso, M., & Ricci, G. (2023). Pathophysiology and Management of Fatigue in Neuromuscular Diseases. *International Journal of Molecular Sciences*, 24(5), 5005. <https://doi.org/10.3390/ijms24055005>

- Toscano, A., Emmanuele, V., Savarese, M., Musumeci, O., Torella, A., Conca, E., Moggio, M., Nigro, V., & Rodolico, C. (2017). Pseudo-dominant inheritance of a novel homozygous HACD1 mutation associated with congenital myopathy: The first caucasian family. *Neuromuscular Disorders*, 27, S173. <https://doi.org/10.1016/j.nmd.2017.06.290>
- Trollet, C., Boulinguez, A., Roth, F., Stojkovic, T., Butler-Browne, G., Evangelista, T., Lacau St Guily, J., & Richard, P. (1993). Oculopharyngeal Muscular Dystrophy. In M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, & A. Amemiya (Eds.), *GeneReviews®*. University of Washington, Seattle. <http://www.ncbi.nlm.nih.gov/books/NBK1126/>
- Tsiplova, K., Zur, R. M., Marshall, C. R., Stavropoulos, D. J., Pereira, S. L., Merico, D., Young, E. J., Sung, W. W. L., Scherer, S. W., & Ungar, W. J. (2017). A microcosting and cost-consequence analysis of clinical genomic testing strategies in autism spectrum disorder. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 19(11), 1268–1275. <https://doi.org/10.1038/gim.2017.47>
- Tupler, R., Berardinelli, A., Barbierato, L., Frants, R., Hewitt, J. E., Lanzi, G., Maraschio, P., & Tiepolo, L. (1996). Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. *Journal of Medical Genetics*, 33(5), 366–370. <https://doi.org/10.1136/jmg.33.5.366>
- Underwood, J., Galvin, B., & Hon, T. (2020). A complete solution for full-length transcript sequencing using the PacBio Sequel II System. *PacBio*. <https://www.pacb.com/proceedings/a-complete-solution-for-full-length-transcript-sequencing-using-the-pacbio-sequel-ii-system/>
- Ungar, W. J., Wu, V., Marshall, C. R., Hwang, J., Hayeems, R. Z., Tsiplova, K., Gillespie, M. K., Szuto, A., Chisholm, C., Stavropoulos, D. J., Venkataramanan, V., Xiao, B., Costain, G., Bergeron, M. B., Sawyer, S., Lau, L., Huang, L., Mendoza-Londono, R., Somerville, M. J., ... Zaidi, S. H. (2026). A microcosting and cost consequence analysis from a randomized controlled trial comparing genome sequencing with exome sequencing for genetic diagnosis. *Genetics in Medicine*, 28(2). <https://doi.org/10.1016/j.gim.2025.101561>

- Valdés-Mas, R., Gutiérrez-Fernández, A., Gómez, J., Coto, E., Astudillo, A., Puente, D. A., Reguero, J. R., Álvarez, V., Morís, C., León, D., Martín, M., Puente, X. S., & López-Otín, C. (2014). Mutations in filamin C cause a new form of familial hypertrophic cardiomyopathy. *Nature Communications*, 5(1), 5326. <https://doi.org/10.1038/ncomms6326>
- van den Bogaart, F. J. A., 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: NMD*, 27(1), 73–77. <https://doi.org/10.1016/j.nmd.2016.09.017>
- van Overveld, P. G. M., Lemmers, R. J. F. L., Sandkuijl, L. A., Enthoven, L., Winokur, S. T., Bakels, F., Padberg, G. W., van Ommen, G.-J. B., Frants, R. R., & van der Maarel, S. M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nature Genetics*, 35(4), 315–317. <https://doi.org/10.1038/ng1262>
- Vassar, M., & Holzmann, M. (2013). The retrospective chart review: Important methodological considerations. *Journal of Educational Evaluation for Health Professions*, 10, 12. <https://doi.org/10.3352/jeehp.2013.10.12>
- Verbiest, M., Maksimov, M., Jin, Y., Anisimova, M., Gymrek, M., & Bilgin Sonay, T. (2023). Mutation and selection processes regulating short tandem repeats give rise to genetic and phenotypic diversity across species. *Journal of Evolutionary Biology*, 36(2), 321–336. <https://doi.org/10.1111/jeb.14106>
- Verhaart, I. E. C., & Aartsma-Rus, A. (2019). Therapeutic developments for Duchenne muscular dystrophy. *Nature Reviews. Neurology*, 15(7), 373–386. <https://doi.org/10.1038/s41582-019-0203-3>
- Volodarsky, M., Kerkhof, J., Stuart, A., Levy, M., Brady, L. I., Tarnopolsky, M., Lin, H., Ainsworth, P., & Sadikovic, B. (2021). Comprehensive genetic sequence and copy number analysis for Charcot-Marie-Tooth disease in a Canadian cohort of 2517 patients. *Journal of Medical Genetics*, 58(4), 284–288. <https://doi.org/10.1136/jmedgenet-2019-106641>

- Vorgerd, M., van der Ven, P. F. M., Bruchertseifer, V., Löwe, T., Kley, R. A., Schröder, R., Lochmüller, H., Himmel, M., Koehler, K., Fürst, D. O., & Huebner, A. (2005). A mutation in the dimerization domain of filamin c causes a novel type of autosomal dominant myofibrillar myopathy. *American Journal of Human Genetics*, 77(2), 297–304. <https://doi.org/10.1086/431959>
- Voulgaris, A., Antoniadou, M., Agraftiotis, M., & Steiropoulos, P. (2019). Respiratory Involvement in Patients with Neuromuscular Diseases: A Narrative Review. *Pulmonary Medicine*, 2019, 2734054. <https://doi.org/10.1155/2019/2734054>
- Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences = Theorie in Den Biowissenschaften*, 131(4), 281–285. <https://doi.org/10.1007/s12064-012-0162-3>
- Wain, K. E., Azzariti, D. R., Goldstein, J. L., Johnson, A. K., Krautscheid, P., Lepore, B., O’Daniel, J. M., Ritter, D., Savatt, J. M., Riggs, E. R., & Martin, C. L. (2020). Variant Interpretation is a Component of Clinical Practice Among Genetic Counselors in Multiple Specialties. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, 22(4), 785–792. <https://doi.org/10.1038/s41436-019-0705-9>
- Waldrop, M. A., Pastore, M., Schrader, R., Sites, E., Bartholomew, D., Tsao, C.-Y., & Flanigan, K. M. (2019). Diagnostic Utility of Whole Exome Sequencing in the Neuromuscular Clinic. *Neuropediatrics*, 50(2), 96–102. <https://doi.org/10.1055/s-0039-1677734>
- Wallace, S. E., & Bean, L. J. (2023a). Resources for Genetics Professionals—Genetic Disorders Associated with Founder Variants Common in the Cree and/or Ojibway Population. In *GeneReviews® [Internet]*. University of Washington, Seattle. <https://www.ncbi.nlm.nih.gov/books/NBK541231/>
- Wallace, S. E., & Bean, L. J. (2023b). Resources for Genetics Professionals—Genetic Disorders Associated with Founder Variants Common in the Inuit Population. In *GeneReviews® [Internet]*. University of Washington, Seattle. <https://www.ncbi.nlm.nih.gov/books/NBK535572/>

- Wallace, S. E., & Mirzaa, G. M. (2022). Resources for Genetics Professionals—Genetic Disorders Associated with Founder Variants Common in the Métis Population. In *GeneReviews® [Internet]*. University of Washington, Seattle.
https://www.ncbi.nlm.nih.gov/books/NBK581291/table/founder_metis.T.genetic_disorders_associated_with_founder_variants_common_in_the_m%C3%A9tis_population/
- Wang, C.-H., Hsu, T.-R., Liu, M.-Y., Wang, L.-Y., Chou, I.-J., Lee, W.-T., Liang, W.-C., Lee, I.-C., Chen, H.-J., Kao, S.-M., Ho, H.-C., Niu, D.-M., Hsiao, K.-J., Chang, M.-Y., Hsieh, H.-M., & Jong, Y.-J. (2025). Newborn screening facilitates early theranostics and improved spinal muscular atrophy outcome: Five-year real-world evidence from Taiwan. *Orphanet Journal of Rare Diseases*, 20(1), 197. <https://doi.org/10.1186/s13023-025-03697-1>
- Wang, K., Kim, C., Bradfield, J., Guo, Y., Toskala, E., Otieno, F. G., Hou, C., Thomas, K., Cardinale, C., Lyon, G. J., Golhar, R., & Hakonarson, H. (2013). Whole-genome DNA/RNA sequencing identifies truncating mutations in RBCK1 in a novel Mendelian disease with neuromuscular and cardiac involvement. *Genome Medicine*, 5(7), 67. <https://doi.org/10.1186/gm471>
- 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>
- Wang, Q., Shashikant, C. S., Jensen, M., Altman, N. S., & Girirajan, S. (2017). Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Scientific Reports*, 7, 885. <https://doi.org/10.1038/s41598-017-01005-x>
- Warburton, P. E., & Sebra, R. P. (2023). Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. *Annual Review of Genomics and Human Genetics*, 24(1), 109–132. <https://doi.org/10.1146/annurev-genom-101722-103045>
- Weiler, T. (2001). *Limb Girdle Muscular Dystrophy in Unique Manitoba Populations*. <https://doi.org/10.13140/RG.2.1.1585.5760>

- Weiler, T., Greenberg, C. R., Nylen, E., Halliday, W., Morgan, K., Eggertson, D., & Wrogemann, K. (1996). Limb-girdle muscular dystrophy and Miyoshi myopathy in an aboriginal Canadian kindred map to LGMD2B and segregate with the same haplotype. *American Journal of Human Genetics*, 59(4), 872–878.
- Welland, M. J., Ahlquist, K. D., De Fazio, P., Austin-Tse, C., Pais, L., Wedd, L., Bryen, S., Rius, R., Franklin, M., Morrison, C., Hall, G., Gauthier, L., Bloemendal, A., Francis, D. I., Mallett, A. J., Mallawaarachchi, A., Lockhart, P. J., Leventer, R., Scheffer, I. E., ... Stark, Z. (2025). *Scalable automated reanalysis of genomic data in research and clinical rare disease cohorts* (2025.05.19.25327921). medRxiv: The Preprint Server for Health Sciences.
<https://doi.org/10.1101/2025.05.19.25327921>
- Wenger, A. M., Guturu, H., Bernstein, J. A., & Bejerano, G. (2017). Systematic reanalysis of clinical exome data yields additional diagnoses: Implications for providers. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 19(2), 209–214.
<https://doi.org/10.1038/gim.2016.88>
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., Ebler, J., Functammasan, A., Kolesnikov, A., Olson, N. D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A. M., Schatz, M. C., Myers, G., DePristo, M. A., ... Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, 37(10), 1155–1162.
<https://doi.org/10.1038/s41587-019-0217-9>
- Westra, D., Schouten, M. I., Stunnenberg, B. C., Kusters, B., Saris, C. G. J., Erasmus, C. E., van Engelen, B. G., Bulk, S., Verschuuren-Bemelmans, C. C., Gerkes, E. H., de Geus, C., van der Zwaag, P. A., Chan, S., Chung, B., Barge-Schaapveld, D. Q. C. M., Kriek, M., Sznajer, Y., van Spaendonck-Zwarts, K., van der Kooi, A. J., ... Voermans, N. C. (2019). Panel-Based Exome Sequencing for Neuromuscular Disorders as a Diagnostic Service. *Journal of Neuromuscular Diseases*, 6(2), 241–258. <https://doi.org/10.3233/JND-180376>

- Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A. H., Karczewski, K., Ing, A. Y., Barton, P. J. R., Funke, B., Cook, S. A., MacArthur, D., & Ware, J. S. (2017). Using high-resolution variant frequencies to empower clinical genome interpretation. *Genetics in Medicine*, 19(10), 1151–1158. <https://doi.org/10.1038/gim.2017.26>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Winblad, S., Jensen, C., Månsson, J.-E., Samuelsson, L., & Lindberg, C. (2010). Depression in Myotonic Dystrophy type 1: Clinical and neuronal correlates. *Behavioral and Brain Functions: BBF*, 6, 25. <https://doi.org/10.1186/1744-9081-6-25>
- Winder, T. L., Tan, C. A., Klemm, S., White, H., Westbrook, J. M., Wang, J. Z., Entezam, A., Truty, R., Nussbaum, R. L., McNally, E. M., & Aradhya, S. (2020). Clinical utility of multigene analysis in over 25,000 patients with neuromuscular disorders. *Neurology Genetics*, 6(2), e412. <https://doi.org/10.1212/nxg.0000000000000412>
- Wisman, G. B. A., Wojdacz, T. K., Altucci, L., Rots, M. G., DeMeo, D. L., & Snieder, H. (2024). Clinical promise and applications of epigenetic biomarkers. *Clinical Epigenetics*, 16, 192. <https://doi.org/10.1186/s13148-024-01806-8>
- Wojcik, M. H., Lemire, G., Berger, E., Zaki, M. S., Wissmann, M., Win, W., White, S. M., Weisburd, B., Wieczorek, D., Waddell, L. B., Verboon, J. M., VanNoy, G. E., Töpf, A., Tan, T. Y., Syrbe, S., Strehlow, V., Straub, V., Stenton, S. L., Snow, H., ... O'Donnell-Luria, A. (2024). Genome Sequencing for Diagnosing Rare Diseases. *New England Journal of Medicine*, 390(21), 1985–1997. <https://doi.org/10.1056/NEJMoa2314761>
- Wong, C.-J., Wang, L., Holers, V. M., Frazer-Abel, A., van der Maarel, S. M., Tawil, R., Statland, J. M., Tapscott, S. J., & ReSolve Network. (2022). Elevated plasma complement components in

- facioscapulohumeral dystrophy. *Human Molecular Genetics*, 31(11), 1821–1829.
<https://doi.org/10.1093/hmg/ddab364>
- Wong, K. N., McIntyre, M., Cook, S., Hart, K., Wilson, A., Moldt, S., Rohrwasser, A., & Butterfield, R. J. (2024). A Five-Year Review of Newborn Screening for Spinal Muscular Atrophy in the State of Utah: Lessons Learned. *International Journal of Neonatal Screening*, 10(3), 54.
<https://doi.org/10.3390/ijns10030054>
- Wright, C. F., McRae, J. F., Clayton, S., Gallone, G., Aitken, S., FitzGerald, T. W., Jones, P., Prigmore, E., Rajan, D., Lord, J., Sifrim, A., Kelsell, R., Parker, M. J., Barrett, J. C., Hurles, M. E., FitzPatrick, D. R., & Firth, H. V. (2018). Making new genetic diagnoses with old data: Iterative reanalysis and reporting from genome-wide data in 1133 families with developmental disorders. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, 20(10), 1216–1223.
<https://doi.org/10.1038/gim.2017.246>
- Wu, Y.-F., Chen, J.-A., & Jong, Y.-J. (2025). Treating neuromuscular diseases: Unveiling gene therapy breakthroughs and pioneering future applications. *Journal of Biomedical Science*, 32(1), 30.
<https://doi.org/10.1186/s12929-025-01123-z>
- 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/10.1002/acn3.51201>
- Xuan, J., Yu, Y., Qing, T., Guo, L., & Shi, L. (2013). Next-generation sequencing in the clinic: Promises and challenges. *Cancer Letters*, 340(2), 284–295. <https://doi.org/10.1016/j.canlet.2012.11.025>
- Yang, Z., Yang, X., Chen, Y., Wang, Z., Fu, X., Song, L., Yang, X., Peng, Z., & Dai, Y. (2023a). *Whole genome and transcriptome sequencing in neuromuscular disorders: A diagnostic and health economic analysis* (p. 2023.12.21.23300182). medRxiv.
<https://doi.org/10.1101/2023.12.21.23300182>

- Yang, Z., Yang, X., Chen, Y., Wang, Z., Fu, X., Song, L., Yang, X., Peng, Z., & Dai, Y. (2023b). *Whole genome and transcriptome sequencing in neuromuscular disorders: A diagnostic and health economic analysis* (p. 2023.12.21.23300182). medRxiv.
<https://doi.org/10.1101/2023.12.21.23300182>
- 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/10.1016/j.nmd.2023.05.004>
- Yépez, V. A., Luknárová, R., Beijer, D., Estévez-Arias, B., Mei, D., Morsy, H., Mueller, J. S., Polavarapu, K., Demidov, G., Doornbos, C., Ellwanger, K., Kraß, L., Laurie, S., Matalonga, L., Abdelrazek, I. M., Astuti, G., Bisulli, F., Brechtmann, F., Dabad, M., ... Gagneur, J. (2026). *Standardized transcriptome analysis improves rare disease diagnosis in the pan-European Solve-RD consortium* (p. 2026.02.10.26345954). medRxiv. <https://doi.org/10.64898/2026.02.10.26345954>
- Yeung, P. S.-W., Yamashita, M., & Prakriya, M. (2020). Molecular basis of allosteric Orai1 channel activation by STIM1. *The Journal of Physiology*, 598(9), 1707–1723.
<https://doi.org/10.1113/JP276550>
- Yeung, P. S.-W., Yamashita, M., & Prakriya, M. (2023). A pathogenic human Orai1 mutation unmasks STIM1-independent rapid inactivation of Orai1 channels. *eLife*, 12, e82281.
<https://doi.org/10.7554/eLife.82281>
- 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., ... Wang, Z. (2021). The GGC repeat expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy type 3. *Brain*, 144(6), 1819–1832. <https://doi.org/10.1093/brain/awab077>
- Yu, J., Shan, J., Yu, M., Di, L., Xie, Z., Zhang, W., Lv, H., Meng, L., Zheng, Y., Zhao, Y., Gang, Q., Guo, X., Wang, Y., Xi, J., Zhu, W., Da, Y., Hong, D., Yuan, Y., Yan, C., ... Deng, J. (2022). The CGG

- repeat expansion in RILPL1 is associated with oculopharyngodistal myopathy type 4. *American Journal of Human Genetics*, 109(3), 533–541. <https://doi.org/10.1016/j.ajhg.2022.01.012>
- 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>
- 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>
- Zeng, J., Huang, X., Li, Y., Yan, T., Lou, J., Lyu, G., & Li, Y. (2025). Long-read sequencing transforms the diagnosis of congenital adrenal hyperplasia: Resolving pseudogene interference and structural variations. *Frontiers in Pediatrics*, 13, 1603819. <https://doi.org/10.3389/fped.2025.1603819>
- Zeng, S., Zhang, M., Wang, X., Hu, Z., Li, J., Li, N., Wang, J., Liang, F., Yang, Q., Liu, Q., Fang, L., Hao, J., Shi, F., Ding, X., Teng, J., Yin, X., Jiang, H., Liao, W., Liu, J., ... Tang, B. (2019). Long-read sequencing identified intronic repeat expansions in SAMD12 from Chinese pedigrees affected with familial cortical myoclonic tremor with epilepsy. *Journal of Medical Genetics*, 56(4), 265–270. <https://doi.org/10.1136/jmedgenet-2018-105484>
- Zhao, L., Hu, C., Pan, S., Wang, D., Wang, Y., & Li, X. (2024). Two novel deep intronic variants cause Duchenne muscular dystrophy by splice-altering mechanism. *Neuromuscular Disorders: NMD*, 45, 104470. <https://doi.org/10.1016/j.nmd.2024.104470>
- Zhao, Y., Tsuiko, O., Jatsenko, T., Peeters, G., Souche, E., Geysens, M., Dimitriadou, E., Vanhie, A., Peeraer, K., Debrock, S., Van Esch, H., & Vermeesch, J. R. (2025). Long-read whole-genome sequencing-based concurrent haplotyping and aneuploidy profiling of single cells. *Nucleic Acids Research*, 53(6), gkaf247. <https://doi.org/10.1093/nar/gkaf247>

Zhu, M., Need, A. C., Han, Y., Ge, D., Maia, J. M., Zhu, Q., Heinzen, E. L., Cirulli, E. T., Pelak, K., He, M., Ruzzo, E. K., Gumbs, C., Singh, A., Feng, S., Shianna, K. V., & Goldstein, D. B. (2012). Using ERDS to Infer Copy-Number Variants in High-Coverage Genomes. *American Journal of Human Genetics*, 91(3), 408–421. <https://doi.org/10.1016/j.ajhg.2012.07.004>

Appendix

Appendix A

Data Capture Sheet for patient phenotyping (Objective 1).

Data Capture Sheet

Participant ID (NM#):

Year/month of birth:

Age:

Participant deceased (Y/N):

Sex/gender:

Relation to any other participants?

Suspected inheritance pattern?

Initial Data Capture:

Aggregate data:

Age:

Sex/gender:

Race/ethnicity:

Age referred to genetics:

Participant Description:

Minimal phenotype description:

- Other features:
- Elevated CK?

Genetics intake/follow-up information:

When referred to genetics:

- By whom:
- Referring symptoms:

When first seen in genetics:

When last seen in genetics:

- Reason for lack of follow-up, if applicable:

Prenatal Information:

- Maternal age at delivery:
- Primigravida:
- Prenatal complications: none significant.
- Gestational age at delivery:
 - Prematurity (Y/N):
- Type of delivery (Caesarean, vaginal, etc.):
- Weight at birth (g):
- Birth Length:
- Complications with delivery:

- Neonatal complications:

Family History:

- Pedigree included in patient file (Y/N):
 - Date of most recently updated pedigree:
- Parental consanguinity (Y/N):
 - Degree of relatedness:
- Paternal ethnicity:
- Maternal ethnicity:
- Significant/contributory family history:
- **Any suspected inheritance pattern?**

Relevant Developmental History:

Onset:

- Referring symptoms:

Phenotype:

- Review of Systems:
 - Growth (e.g., height, weight, head circumference, etc.):
 - Neurological:
 - Facial features:
 - Cardiovascular:
 - Respiratory:
 - Hepatic:
 - Gastrointestinal:
 - Urogenital:
 - Hair/skin:

- Endocrine:
- Immunological:
- Musculoskeletal:

Muscular:

Strength:

- **Description** (normal, slight weakness, weakness):
- **Location** (proximal, distal, generalized):
- **Predominance** (upper, lower):
- **Bulbar involvement** (Y/N):
- **Anatomic distribution** (symmetrical, asymmetrical):
- **Associated symptoms** (e.g., cramping, stiffness, pain):
- **Symptom onset** (acute, chronic, episodic):
- **Symptom progression** (stable, improving, worsening):
 - **Rate of progression:**
- **Related exposures to toxins, infectious agents, medications:**
- **Contributing or aggravating factors** (e.g., exercise vs. rest):

Function:

- **Gait characteristics:**
- **Gower manoeuvre** (Y/N):
- **Posture:**
- **Passive range of motion:**
- **Tip toe:**
- **Heel walk:**

Tone:

- **Description** (normal, spasticity, rigidity, etc.):
- **Location** (proximal, distal, generalized):
- **Predominance** (upper, lower):
- **Anatomic distribution** (symmetrical, asymmetrical):

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):

Reflexes:

- **Deep tendon reflexes** (normal, increased, decreased):
- **Infantile reflexes** (persistent, absent):
- **Babinski sign** (+/-, absent):

Additional notes:

- **Skeletal:**

- Description (e.g., arthrogryposis/contractures, scoliosis, scapular winging, dislocations, etc.):

Investigations:

a. Genetic testing findings:

Test performed	Process date	Test technology	Where test performed	Findings	Other (if panel, list genes) (if related familial testing, include)

b. Significant laboratory findings (in biochemical testing):

Muscle function	Result/Interpretation	When performed (*collection date*)
-----------------	-----------------------	------------------------------------

Creatine kinase (CK)		
Myoglobin		
Pompe Disease Screen (dried blood spot)		
<i>General testing</i>		
Plasma serum chemistry (including sodium, potassium, chloride, CO ₂ , anion gap, lactate, glucose, urea, creatinine, ammonia, CK, AST, ALT, GGT, Alk Phos)		
Sodium		
Magnesium		
Chloride		
Calcium		
Phosphorus		
HCO₃ (Bicarbonate)		
CO₂		
O₂		
pH		
<i>Metabolism</i>		
ALT, AST, GGT		
Blood glucose		
Ammonia		
Lactate		
Acylcarnitine		
Serum carnitine		
Urine organic acids		
Urine amino acids		
Urine metabolic screen		

Plasma amino acids		
VLCFA		
Palmitate oxidation		

c. Imaging and biopsy findings

	Date performed	Age of participant	Findings	Provider/notes
EMG/NCS			EMG: NCS:	
MRI				
Muscle biopsy				
Skin biopsy				

Complications:

a. Additional confounding diagnoses (Y/N):

- If yes, describe:

b. Death (if patient is deceased): *not applicable*

Other Relevant Medical Assessment History:

Other Relevant Medical History and Notes

Appendix B

Annotation terms and descriptions of variant call files, divided by variant class (SNVs and CNVs/SV).

Table A1. VCF annotations and annotation descriptions for SNVs, as per SickKids TCAG. Annotations ordered from top to bottom in order provided by SickKids TCAG.

Annotation	TCAG Description
CHROM	chromosome
start	start position (1-positional system)
end	end position (1-positional system)
ref_allele	Reference allele
alt_allele	Alternate allele
var_type	snp (single nucleotide variation, can be ref-alt), ins (insertion), del (deletion), mnp (multiple bases substitutions), complex (could be block substitution). alleles are phased in relation to one another.
Original_VCFKEY	unique variant identifier for original vcf file.
multi_allelic	Flag to represent multi-allelic sites (ie 1=multi-allelic).
Filter	Filter status: PASS if this position has passed all filters.
END	End position (for use with symbolic alleles).
zygosity	heterozygous-reference (ref-alt), homozygous-alternate (hom-alt) and heterozygous-alternate(alt-alt)
Genotype	Genotype, represented as “reference allele alternate allele” (ref-alt) or “alternate allele alternate allele” (hom-alt) or “alternate allele 1 alternate allele 2” (alt-alt). This is based on pre-normalized raw genotype (eg: A/C).
Ad_ref	Allelic depth for the ref allele.
Ad_alt	Allelic depth for the alt allele.
Gt_post_norm	Raw genotype after left-normalization (eg : 0/1)
Gt_pre_norm	Raw genotype before left-normalization (eg : 1/2)
Gq	Genotype quality.
Dp	Filtered basecall depth used for site genotyping.

min_dp	Minimum DP observed within the GVCF block.
vaf	Variant allele fractions.
pl	Normalized, Phred-scaled likelihoods for genotypes as defined in the VCF specification .
ps	Phase set identifier.
typeseq_priority	type of sequence overlapped, with respect to known genes/transcripts and their coding / noncoding status: (a) “exonic” represents coding exons, (b) “splicing” represents core splicing site (by default, 15 bp on the intron side of intron-exon and exon-intron junctions), (c) "ncRNA _ exonic" represents exons of non-coding RNA genes, (d) "ncRNA_splicing" represents core splicing sites of non-coding RNA genes , (e) “UTR5” represents 5’ untranslated region, (f) “UTR3” represents 3’ untranslated region, (g) “upstream” represents 1kb upstream of TSS, (h) “downstream” represents 1kb downstream of TSS and (i) “intergenic” represents intergenic regions (beyond upstream/downstream threshold(1kb)). For variants with multiple sequence overlaps (eg, exonic for one transcript and intronic for other), all possible typeseq values will be listed in semicolon-delimited format (eg: exonic;intronic). For multi-sequence overlap variants, the value is prioritized based on Annovar prioritization dvhrmr
typeseq_refseq_select	Type of sequence overlapped, with respect to RefSeq Select transcript (https://www.ncbi.nlm.nih.gov/refseq/refseq_select/).
refseq_id	combined Annovar output on coding sequence mapping and effect, composed of: (a) for coding exotic changes ("typeset" exonic): gene official symbol : RefSeq ID : position in the coding sequence : amino acid change, [idem]; (b) for core splice site changes (typeseq "exonic"): gene official symbol (RefSeq ID : exon number : coding sequence position and change, [idem])
effect_priority	type of effect on the coding sequence: (a) “synonymous SNV” , (b)"nonsynonymous SNV", (c) "stopgain SNV", (d) "frameshift deletion", (e) "frameshift insertion", (f) "frameshift substitution", (g) "nonframeshift deletion", (h) "nonframeshift insertion", (i) "nonframeshift substitution", (j) “stoploss SNV”. For variants with multiple effects, all possible values will be represented in comma-separated fashion.
effect_refseq_select	Effect on the coding sequence with respect to RefSeq Select transcript (https://www.ncbi.nlm.nih.gov/refseq/refseq_select/).
distance_splice_junction	A positive integer represents the distance from the nearest exon boundary, in genomic level.

gene_symbol	official gene symbol.
entrez_id	entrez_gene id.
gene_desc	full gene name.
pac_bio_darkgene_panel	Twist alliance Dark Genes panel.
gene_type	ncbiRefseqCurated
omim_id	OMIM gene accession id.
omim_phenotype	OMIM disorder/disease description when available for the corresponding omim gene accession.
omim_inheritance	OMIM inheritance.
mpo	(array of) MPO (Mammalian Phenotype Ontology) top level phenotype(s), imported from MGI and mapped from an orthologous mouse gene; the genotype-phenotype association is typically supported by a heterozygous/homozygous knock-out or other transgenic experiment, sometimes involving more than one gene: these details are exported by TCAG from MGI, but not included in this annotation field, so they should be looked up on the mgi website.
hpo	(array of) HPO (Human Phenotype Ontology) top level phenotype(s), imported from HPO; the genotype-phenotype association is typically supported by an OMIM entry; modes of inheritance are also exported by TCAG from HPO, but not included in this annotation field, so they should be looked up on the HPO website.
cgd_disease	the Clinical Genomics Database is compiled by curators and maintained by the NHGRI (National Human Genome Research Institute); for every gene in the database, the CGD provides a list of one or more genetic disorders and a mode of inheritance; this field reports the genetic disorder(s) .
cgd_inheritance	the Clinical Genomics Database is compiled by curators and maintained by the NHGRI (National Human Genome Research Institute); for every gene in the database, the CGD provides a list of one or more genetic disorders and a mode of inheritance; this field reports the mode of inheritance (AD, AR, AD/AR, XL, more complex modes); since the CGD mode of inheritance is directly added by a curator and it's tied to specific genetic disorder(s), it could be considered more accurate than the mode of inheritance for top level HPO phenotypes.
gnom_ad_lof_oe	Observed over expected ratio for high confidence pLoF variants in transcript (lof.obs divided by lof.exp).
gnom_ad_lof_oe_ci_upper	LOEUF - upper bound of 90% confidence interval for o/e ratio for high confidence pLoF variants (lower values indicate more constrained).

gnom_ad_mis_oe	Observed over expected ratio for missense variants in transcript (mis.obs divided by mis.exp).
gnom_ad_mis_oe_ci_upper	Upper bound of 90% confidence interval for o/e ratio for missense variants
gnom_ad_lof_p_li	Probability of loss-of-function intolerance; probability that transcript falls into distribution of haploinsufficient genes (~21% o/e pLoF ratio; computed from high confidence pLoF gnomAD data).
gnom_ad_lof_p_rec	Probability that transcript falls into distribution of recessive genes (~71% o/e pLoF ratio; computed from high confidence pLoF gnomAD data).
gnom_ad_mis_z_score	Z-score for missense variants in transcript. Higher (more positive) Z scores indicate that the transcript is more intolerant of variation (more constrained).
acmg_disease	Any (exonic, intronic or splice) variants in genes in ACMG (SF v3.2) published recommendations for reporting incidental findings. (https://www.nature.com/articles/s41436-021-01172-3)
clinvar_sig	Overall ClinVar significance code; “pathogenic” is the code of interest for rare disorders. Clinical significance values for all the individual submissions (SCVs) aggregated for the RCV record in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/)
clinvar_clnref	References includes PubMed ID
clinvar_disease	Clinvar interpreted condition.
clinvar_allele_id	ClinVar alleleID. A unique integer identifier assigned to each individual variant in ClinVar.
clinvar_review_status	The level of review supporting the assertion of clinical significance(https://www.ncbi.nlm.nih.gov/clinvar/docs/details/#review_status)
clinvar_sig_simple	expected values 0,1 or -1; 0 = no current value of pathogenic; 1 = at least one record submitted with pathogenic/likely pathogenic; -1 = no values for clinical significance at all for this variant or set of variants. Used for the "included" variants that are only in ClinVar because they are included in a haplotype or genotype with an interpretation.
cosmic	exact match (position, allele) to the Cosmic database of somatic variants; useful only for cancer projects.
gnomad_v4_exome_ac	Alternate allele count from gnomAD exome dataset.
gnomad_v4_exome_ac_xx	Alternate allele count in female samples from gnomAD exome dataset.
gnomad_v4_exome_ac_xy	Alternate allele count in male samples from gnomAD exome dataset.
gnomad_v4_exome_an	Total number of alleles from gnomAD exome dataset.

gnomad_v4_exome_an_xx	Total number of alleles in female samples from gnomAD exome dataset.
gnomad_v4_exome_an_xy	Total number of alleles in male samples from gnomAD exome dataset.
gnomad_v4_exome_af	Alternate allele frequency from gnomAD exome dataset.
gnomad_v4_exome_af_xx	Alternate allele frequency in female samples from gnomAD exome dataset.
gnomad_v4_exome_af_xy	Alternate allele frequency in male samples from gnomAD exome dataset.
gnomad_v4_exome_af_afr	Alternate allele frequency in samples of African/African-American ancestry from gnomAD exome dataset.
gnomad_v4_exome_af_amr	Alternate allele frequency in samples of Latino ancestry from gnomAD exome dataset.
gnomad_v4_exome_af_asj	Alternate allele frequency in samples of Ashkenazi Jewish ancestry from gnomAD exome dataset.
gnomad_v4_exome_af_eas	Alternate allele frequency in samples of East Asian ancestry from gnomAD exome dataset.
gnomad_v4_exome_af_fin	Alternate allele frequency in samples of Finnish ancestry from gnomAD exome dataset.
gnomad_v4_exome_af_grpmax	Maximum allele frequency across genetic ancestry groups from gnomAD exome dataset.
gnomad_v4_exome_af_mid	Alternate allele frequency in samples of Middle Eastern ancestry from gnomAD exome dataset.
gnomad_v4_exome_af_nfe	Alternate allele frequency in samples of Non-Finnish European ancestry from gnomAD exome dataset.
gnomad_v4_exome_af_raw	Alternate allele frequency in samples, before removing low-confidence genotypes from gnomAD exome dataset.
gnomad_v4_exome_af_remaining	Alternate allele frequency in samples of remaining individuals ancestry from gnomAD exome dataset.
gnomad_v4_exome_af_sas	Alternate allele frequency in samples of South Asian ancestry from gnomAD exome dataset.
gnomad_v4_exome_faf95	Filtering allele frequency (using Poisson 95% CI) from gnomAD exome dataset.
gnomad_v4_exome_faf95_xx	Filtering allele frequency (using Poisson 95% CI) for female samples in non-PAR regions of sex chromosomes only, from gnomAD exome dataset.
gnomad_v4_exome_faf95_xy	Filtering allele frequency (using Poisson 95% CI) for male samples in non-PAR regions of sex chromosomes only, from gnomAD exome dataset.
gnomad_v4_exome_faf95_afr	Filtering allele frequency (using Poisson 95% CI) for samples of African/African-American ancestry from gnomAD exome dataset.
gnomad_v4_exome_faf95_amr	Filtering allele frequency (using Poisson 95% CI) for samples of Latino ancestry from gnomAD exome dataset.

gnomad_v4_exome_faf95_eas	Filtering allele frequency (using Poisson 95% CI) for samples of East Asian ancestry from gnomAD exome dataset.
gnomad_v4_exome_faf95_nfe	Filtering allele frequency (using Poisson 95% CI) for samples of Non-Finnish European ancestry from gnomAD exome dataset.
gnomad_v4_exome_faf95_sas	Filtering allele frequency (using Poisson 95% CI) for samples of South Asian ancestry from gnomAD exome dataset.
gnomad_v4_exome_fafmax_faf95_max	Maximum filtering allele frequency (using Poisson 95% CI) across genetic ancestry groups, from gnomAD exome dataset.
gnomad_v4_exome_fafmax_faf95_max_gen_anc	Genetic ancestry group with maximum filtering allele frequency (using Poisson 95% CI), from gnomAD exome dataset.
gnomad_v4_exome_grpmax	Genetic ancestry group with the maximum allele frequency, from gnomAD exome dataset.
gnomad_v4_exome_lcr	Flag represents whether the variant falls within a low complexity region or not
gnomad_v4_exome_nhomalt	Count of homozygous individuals, from gnomAD exome dataset.
gnomad_v4_exome_nhomalt_xx	Count of homozygous individuals in female samples from gnomAD exome dataset.
gnomad_v4_exome_nhomalt_xy	Count of homozygous individuals in male samples, from gnomAD exome dataset.
gnomad_v4_exome_pangolin_large_st_ds	Pangolin's largest delta score across 2 splicing consequences, which reflects the probability of the variant being splice-altering.
gnomad_v4_exome_spliceai_ds_max	Illumina's SpliceAI max delta score; interpreted as the probability of the variant being splice-altering.
gnomad_v4_genome_ac	Alternate allele count from gnomAD exome dataset.
gnomad_v4_genome_ac_xx	Alternate allele count in female samples from gnomAD exome dataset.
gnomad_v4_genome_ac_xy	Alternate allele count in male samples from gnomAD exome dataset.
gnomad_v4_genome_an	Total number of alleles from gnomAD exome dataset.
gnomad_v4_genome_an_xx	Total number of alleles in female samples from gnomAD exome dataset.
gnomad_v4_genome_an_xy	Total number of alleles in male samples from gnomAD exome dataset.
gnomad_v4_genome_af	Alternate allele frequency from gnomAD exome dataset.
gnomad_v4_genome_af_afr	Alternate allele frequency in samples of African/African-American ancestry from gnomAD exome dataset.
gnomad_v4_genome_af_ami	Alternate allele frequency in samples of Amish ancestry from gnomAD genome dataset.
gnomad_v4_genome_af_amr	Alternate allele frequency in samples of Latino ancestry from gnomAD exome dataset.

gnomad_v4_genome_af_asj	Alternate allele frequency in samples of Ashkenazi Jewish ancestry from gnomAD exome dataset.
-------------------------	---

gnomad_v4_genome_af_eas	Alternate allele frequency in samples of East Asian ancestry from gnomAD exome dataset.
gnomad_v4_genome_af_fin	Alternate allele frequency in samples of Finnish ancestry from gnomAD exome dataset.
gnomad_v4_genome_af_grpmax	Maximum allele frequency across genetic ancestry groups from gnomAD exome dataset
gnomad_v4_genome_af_mid	Alternate allele frequency in samples of Middle Eastern ancestry from gnomAD exome dataset.
gnomad_v4_genome_af_nfe	Alternate allele frequency in samples of Non-Finnish European ancestry from gnomAD exome dataset
gnomad_v4_genome_af_raw	Alternate allele frequency in samples, before removing low-confidence genotypes from gnomAD exome dataset.
gnomad_v4_genome_af_remaining	Alternate allele frequency in samples of remaining individuals ancestry from gnomAD exome dataset.
gnomad_v4_genome_af_sas	Alternate allele frequency in samples of South Asian ancestry from gnomAD genome dataset.
gnomad_v4_genome_faf95	Filtering allele frequency (using Poisson 95% CI) from gnomAD genome dataset.
gnomad_v4_genome_faf95_afr	Filtering allele frequency (using Poisson 95% CI) for samples of African/African-American ancestry from gnomAD genome dataset.
gnomad_v4_genome_faf95_amr	Filtering allele frequency (using Poisson 95% CI) for samples of Latino ancestry from gnomAD genome dataset.
gnomad_v4_genome_faf95_eas	Filtering allele frequency (using Poisson 95% CI) for samples of East Asian ancestry from gnomAD genome dataset.
gnomad_v4_genome_faf95_nfe	Filtering allele frequency (using Poisson 95% CI) for samples of Non-Finnish European ancestry from gnomAD genome dataset.
gnomad_v4_genome_faf95_sas	Filtering allele frequency (using Poisson 95% CI) for samples of South Asian ancestry from gnomAD genome dataset.
gnomad_v4_genome_faf95_xx	Filtering allele frequency (using Poisson 95% CI) for female samples in non-PAR regions of sex chromosomes only, from gnomAD genome dataset.
gnomad_v4_genome_faf95_xy	Filtering allele frequency (using Poisson 95% CI) for male samples in non-PAR regions of sex chromosomes only, from gnomAD genome dataset.
gnomad_v4_genome_fafmax_faf95_max	Maximum filtering allele frequency (using Poisson 95% CI) across genetic_ancestry groups, from gnomAD genome dataset.

gnomad_v4_genome_fafmax_faf95_max_gen_anc	Genetic ancestry group with maximum filtering allele frequency (using Poisson 95% CI), from gnomAD genome dataset.
gnomad_v4_genome_grpmax	Genetic ancestry group with the maximum allele frequency, from gnomAD genome dataset.
gnomad_v4_genome_lcr	Flag represents whether the variant falls within a low complexity region or not.
gnomad_v4_genome_nhomalt	Count of homozygous individuals from gnomAD genome dataset.
gnomad_v4_genome_nhomalt_xx	Count of homozygous individuals in female samples, from gnomAD genome dataset.
gnomad_v4_genome_nhomalt_xy	Count of homozygous individuals in male samples, from gnomAD genome dataset.
gnomad_v4_genome_pangolin_largest_ds	Pangolin's largest delta score across 2 splicing consequences, which reflects the probability of the variant being splice-altering.
gnomad_v4_genome_spliceai_ds_max	Illumina's SpliceAI max delta score; interpreted as the probability of the variant being splice-altering.
ga4k_af	The Children's Mercy Research Institute frequency (~500 samples).
freq_max	Maximum of GnomAD exome/GnomAD genome filtering allele frequencies(faf) .
a1000g_freq_max	Maximum of 1000 Genomes allele frequencies.
dbSNP	exact match (position, allele) to dbSNP.
dbSNP_common	exact match (position, allele) to common dbSNP track USCS; this is not meant as a replacement of frequency-based filtering, and is not practically used for hard filters.
dbSNP_region	overlap-based match for dbSNP; this is useful to look up variants that are split into multiple dbSNP entries, or otherwise only partially match to dbSNP entries.
sift_score	SIFT score for predicted protein impact, values ≤ 0.05 correspond to damaging (can be interpreted as a p-value); note that SIFT is based on amino acid conservation/substitution rates inferred from protein sequence alignments.
polyphen_score	Polyphen2 scores for predicted protein impact, values ≥ 0.95 correspond to damaging; note that Polyphen2 is based on a set of sequence and structural attributes, combined in a machine classifier trained on a data-set of positive cases, and thus it partially is complementary and partially correlated to SIFT and MA.
ma_score	mutation assessor scores for predicted protein impact, values ≥ 2 correspond to damaging; note that MA is based on amino acid conservation / substitution rates inferred from protein sequence alignments, additionally modeling protein family groupings, thus can be regarded as an improved version of SIFT (improved performance has been shown particularly for somatic variants), although to this date SIFT remains more popular / broadly used.
mt_score	Mutationtaster prediction scores.

mpc_score	MPC (Missense badness, PolyPhen-2, and Constraint) score. MPC is a missense deleteriousness metric based on the analysis of genic regions depleted of missense variants in the Exome Aggregation Consortium (ExAC) data.
revel_score	Rare Exome Variant Ensemble Learner (REVEL) scores for predicting the deleteriousness of each nucleotide change. The REVEL score for an individual missensevariant can range from 0 to 1, with higher scores reflecting greater likelihood that the variant is disease-causing.
alpha_missense_score	A continuous score between 0 and 1 which can be interpreted as the predicted probability of the variant being pathogenic.
alpha_missense_aa	Alphamissense amino acid change.
alpha_missense_transcript	Refseq mane transcript id related to this alphamissense score.
cadd_raw	Raw score from CADD(Combined Annotation Dependent Depletion).
cadd_phred	PHRED-like c-score ranking from CADD.
phylop_mam_avg	Average value of PhyloP nucleotide-level conservation inferred from the Placental Mammal genome group. This field is suitable for hard filters, especially for missense variants and non-frameshift substitutions; note that PhyloP is based uniquely on nucleotide substitutions and does not factor structural variation.
phylop_vert100_avg	Average value of PhyloP nucleotide-level conservation inferred from the 100 Vertebrate genome group; this field is suitable to hard filter, especially for missense variants and non-frameshift substitutions; note that PhyloP is based uniquely on nucleotide substitutions and does not factor structural variation.
phast_cons_placental	PhastCons score for the Placental Mammal genome group, based on UCSC track (calculated by HMM scan of PhyloP values to infer conserved status); this is useful to assess conservation at the *regional level* (rather than PhyloP's *nucleotide level*); this field can be used as an evidence *suggestive* of functional relevance of a sequence (especially if non-coding), and thus could be used to hard filter non-coding variation.
pfam_annovar	overlap with coding sequence matching to a PFAM protein domain; can help further assess protein impact, but not meant for hard-filter.
per_cds_affected	percentage of coding exonic sequence affected, should be used to prioritize loss-of-function variation (stopgain, frameshift, splicing), but not meant for hard filters.
per_transcripts_affected	percentage of transcripts with variant overlapping them, should be used to prioritize variation, but not meant for hard filters.
dbsc_snv_ada_score	Splice site prediction scores from dbscSNV

	(https://www.ncbi.nlm.nih.gov/pubmed/26555599)
dbsc_snv_rf_score	Splice site prediction scores from dbscSNV (https://www.ncbi.nlm.nih.gov/pubmed/26555599)
splice_ai_symbol	Gene symbol from SpliceAI (https://pypi.org/project/spliceai/)
splice_ai_ds_ag	SpliceAI Delta score, the probability of the variant being splice-altering (acceptor gain). T o filter use 0.5 (recommended), for higher recall 0.2 can be used.
splice_ai_ds_al	SpliceAI Delta score, the probability of the variant being splice-altering Delta score (acceptor loss). T o filter use 0.5 (recommended), for higher recall 0.2 can be used.
splice_ai_ds_dg	SpliceAI Delta score, the probability of the variant being splice-altering Delta score (donor gain). T o filter use 0.5 (recommended), for higher recall 0.2 can be used.
splice_ai_ds_dl	SpliceAI Delta score, the probability of the variant being splice-altering Delta score (donor loss). T o filter use 0.5 (recommended), for higher recall 0.2 can be used.
splice_ai_dp_ag	SpliceAI Delta position (acceptor gain).
splice_ai_dp_al	SpliceAI Delta position (acceptor loss).
splice_ai_dp_dg	SpliceAI Delta position (donor gain).
splice_ai_dp_dl	SpliceAI Delta position (donor loss).
seg_dup	overlap with segmental duplications, often useful to explain why ploidy is not 2 outside male sex chromosomes.
repeat	Repeatmasker annotation from UCSC.
effect_impact_str	pre-computed effects and impacts (eg: Missense-Low).
effect_impact_str_select	pre-computed effects and impacts based refseq select gene definitions.

ACMG, American College of Medical Genetics and Genomics; CI, confidence interval; gnomAD, Genome Aggregation Database; GVCF, genomic variant call format; HMM, hidden Markov model; HPO, Human Phenotype Ontology; LOEUF, Loss-of-Function Observed/Expected Upper bound Fraction; MA, mutation assessor; MGI, Mouse Genomic Informatics; MPO, Mammalian Phenotype Ontology; MT, mutation taster; OMIM; Online Mendelian Inheritance in Man; PFAM: protein families database; pLOF, probability loss of function; RCV, Reference ClinVar record; SIFT, sorting intolerant from tolerant; SNV, single nucleotide variant; TCAG, The Centre for Applied Genomics; TSS, transcription start site; UCSC; University of California Santa Cruz; VCF, variant call format.

Table A2. VCF annotations and annotation descriptions for CNVs and SVs, as per SickKids TCAG. Annotations ordered from top to bottom in order provided by SickKids TCAG. The presence of an ‘X’ indicates that the annotation is provided for that variant class.

Annotation	TCAG Description	CNV	SV
chrAnn	Chromosome (chr1 to chr22, plus chrX, chrY, and chrM).	X	X
pos	CNV/SV start (padded for INS)	X	X
end	CNV/SV end (equals to the “end” column, or the END field in the INFO column, padded for INS).	X	X
svlen	Difference in length between REF and ALT alleles	X	X
svtype	Type of structural variant	X	X
ref	Reference allele		X
alt	Alternate allele		X
qual	Genotype Quality	X	X
filter	Sample filter, 'PASS' indicates that all filters have passed for this sample	X	X
cistart	Confidence interval around POS	X	X
ciend	Confidence interval around END	X	X
imprecise	Imprecise structural variation	X	X
variantTypeAnn	Variant type as interpreted by the annotation pipeline; for CNVs: deletions and duplications. For CNVs, complex is used when the calling software assigns a type which cannot be identified as either duplication or deletion (e.g. translocations, inversions, and other events); in this case the annotation pipeline ignores the type in the database when matching it with the variant. SV types recognized by the pipeline are: duplications, deletions, inversions, and insertions. Insertions of size 1 are padded (+/- 50bp) prior to frequency calculations.	X	X
cipos	Confidence interval around POS	X	X
mateid	ID of mate breakend		X
matedist	Distance, in base pairs, between the two breakends of a translocation when both occur on the same chromosome/contig.		X

gt	Genotype	X	X
ad	Allelic depth.		X
dp	Filtered basecall depth used for site genotyping.		X
sac	Stranded allele counts, representing allele-supporting reads separated by strand.		X
cn	Estimated copy number		X
chr_ann	Chromosome (chr1 to chr22, plus chrX, chrY, and chrM).	X	X
start_ann	CNV/SV start (padded for INS) ¹	X	
end_ann	CNV/SV end (equals to the “end” column, or the END field in the INFO column, padded for INS).	X	X
size_ann	CNV/SV size as determined by the pipeline (end – start +1).	X	X
gc_content_perc	GC content based on the UCSC genome reference, hg19 or hg38.	X	X
cytoband_ann	Cytoband.	X	X
number_of_gene_symbols	number of official gene symbol(s) for genes spanned by the CNV/SV based on the UCSC RefSeq gene definitions.	X	X
gene_symbol	official gene symbol(s) for genes spanned by the CNV/SV based on the UCSC RefSeq gene definitions.	X	X
gene_eg_id	entrez gene IDs.	X	X
gene_symbol_cnvstart	official gene symbol(s); overlap on CNV/SV start coordinate.	X	X
gene_symbol_cnvend	official gene symbol(s); overlap on CNV/SV end coordinate.	X	X
exon_symbol	official gene symbol(s); exons overlap only.	X	X
exon_eg_id	official gene entrez gene IDs, exons overlap only.	X	X
cds_symbol	official gene symbol(s); coding exons overlap.	X	X
cds_eg_id	official gene entrez IDs; coding exons overlap.	X	X
tx_symbol_lnc_rna	official lncRNA gene symbol (or reference to ID in lncRNAKB if novel gene) (only hg38).	X	X
ex_symbol_lnc_rna	official lncRNA gene symbol (or reference to ID in lncRNAKB if novel gene), exons overlap (only hg38).	X	X
dark_genes_symbol	official gene symbol of genes spanning the dark regions (hard-to-map).	X	X
dark_genes_pct_ovlp	% overlap with the dark regions.	X	X
clin_gen_haploinsufficient	(array of) entrez gene ID, gene symbol, and score from the dosage sensitivity map, haploinsufficient phenotype defined in ClinGen (ClinGen). ClinGen is a National Institutes of Health (NIH)-funded	X	X

	resource dedicated to building an authoritative central resource that defines the clinical relevance of genes and variants for use in precision medicine and research.		
clin_gen_triplosensitive	(array of) entrez gene ID, gene symbol, and score for dosage sensitivity map, triplosensitive phenotype defined in ClinGen (ClinGen).	X	X
gnom_ad_oe_lof	observed/expected ratio for loss of function from the genome aggregation database (gnomAD)	X	X
gnom_ad_oe_lof_upper	observed/expected upper bound loss of function from the genome aggregation database (gnomAD)	X	X
gnom_ad_oe_mis	observed/expected missense ratio from the genome aggregation database (gnomAD)	X	X
gnom_ad_oe_mis_upper	observed/expected upper bound missense from the genome aggregation database (gnomAD)	X	X
gnom_ad_p_li	probability that a gene falls into the class of intolerant of a single LoF gene (LoF-haploinsufficient intolerant genes), from the genome aggregation database (gnomAD)	X	X
gnom_ad_p_rec	probability that a gene falls into the class of intolerant of two LoF genes (recessive genes), from the genome aggregation database (gnomAD)	X	X
repeat_masker_pct_ovlp	percent overlap with repeat regions (RepeatMasker annotation from UCSC).	X	X
dirty_region_pct_ovlp	percent overlap with gaps (including centromeres and telomeres), and segmental duplications.	X	X
chr_region	telomere/centromere tag.	X	X
mpo_nervous_system	(array of) entrez gene ID, gene symbol, and inheritance for genes in the MPO terms part of nervous system phenotype (for MPO description, see MPO_Other).	X	X
mpo_growth	(array of) entrez gene ID, gene symbol, and inheritance for genes in the MPO terms part of growth phenotype (for MPO description, see MPO_Other).	X	X
mpo_other	(array of) entrez gene ID, gene symbol, MPO term, and inheritance for genes in the MPO (Mammalian Phenotype Ontology) top level phenotype(s) excluding terms in nervous system or growth phenotypes, imported from MGI and mapped from an orthologous mouse gene; the	X	X

genotype- phenotype association is typically supported by a heterozygous/homozygous knock-out or other transgenic experiment, sometimes involving more than one gene: these details are exported by TCAG from MGI, but not included in this annotation field, so they should be confirmed on the MGI website.

hpo_nervous_system	(array of) entrez gene ID, gene symbol and inheritance for genes in the HPO terms part of nervous system phenotype (for HPO description, see HPO_Other).	X	X
hpo_growth	(array of) entrez gene ID, gene symbol and inheritance for genes in the HPO terms part of growth system phenotype (for HPO description, see HPO_Other).	X	X
hpo_other	(array of) entrez gene ID, gene symbol, MPO term, and inheritance for genes in the HPO (Human Phenotype Ontology) top level phenotype(s), imported from HPO, excluding nervous system and growth phenotypes. The genotype-phenotype association is typically supported by an OMIM entry.	X	X
cgd	(array of) entrez gene ID, gene symbol, disease name(s), and inheritance found in the Clinical Genomics Database; it is compiled by curators and maintained by the NHGRI (National Human Genome Research Institute); for every gene in the database, the CGD provides a list of one or more genetic disorders and a mode of inheritance (AD, AR, AD/AR, XL, more complex modes); since the CGD mode of inheritance is directly added by a curator and is tied to specific genetic disorder(s), it could be considered more accurate than the mode of inheritance for top-level HPO phenotypes.	X	X
omim_pheno	(array of) entrez gene ID, gene symbol, and disorder/disease name(s), with their respective inheritance modes coded as “i:”, found in OMIM.	X	X
omim_inh	list of all the inheritance modes for OMIM phenotypes	X	X
as_dgenes	ASD associated genes. See appendix for the source.	X	X
acm_ggenes	disease associated genes listed by ACMG.	X	X
clin_gen_region	Genomic disease region from ClinGen.	X	X
cnv_clin_gen_pct_ovlp	% length of CNV/SV overlapped by ClinGen region(s).	X	X
clin_gen_cnv_pct_ovlp	(array of) % length of ClinGen region(s) overlapped by CNV/SV.	X	X
clin_gen_cnv_pct_ovlp_max	highest % length of ClinGen region overlapped by CNV/SV.	X	X

clin_gen_match_cn_vmax_pct_ovlp	% length of CNV/SV overlapped by largest ClinGen region.	X	X
exon_symbol_clin_gen	gene symbol(s) in the region covered by ClinGen.	X	X
decipher_region	Genomic disease region from Decipher; the Database of genomic Variation and Phenotype in Humans using Ensembl Resources is an interactive web-based database which incorporates a suite of tools designed to aid the interpretation of genomic variants. For more information, see: https://decipher.sanger.ac.uk/ .	X	X
cnv_decipher_pct_ovlp	% length of CNV/SV overlapped by Decipher region(s).	X	X
decipher_cnv_pct_ovlp	% length of Decipher region(s) overlapped by CNV	X	X
decipher_cnv_pct_ovlp_max	highest % length of Decipher region(s) overlapped by CNV.	X	X
decipher_match_cn_vmax_ptc_ovlp	% length of CNV/SV overlapped by largest Decipher region.	X	X
exon_symbols_decipher	gene symbol(s) in the region covered by Decipher.	X	X
clin_var_symbol	genes present in the ClinVar database, if listed.	X	X
clin_var_clin_sign	clinical significance of the genes present in the ClinVar database, by relevance.	X	X
clin_var_var_id	ClinVar identifier	X	X
ntsd186_pct_ovlp	% overlap with variants in the NTSD186 dataset from NCBI.	X	X
nstd_type_list	list of variants from the NTSD186 dataset from NCBI, which overlap the CNV/SV.		
gnom_ad_af_max90pct_rec_ovlp	Maximum overall gnomAD allele frequency among matching SVs with at least 90% reciprocal overlap.		X
gnom_ad_af_max_max90pct_rec_ovlp	Maximum value of the gnomAD AF_MAX field among matching SVs with at least 90% reciprocal overlap.		X
gnom_ad_af_popmax_max90pct_rec_ovlp	Maximum population-specific allele frequency reported by gnomAD among matching SVs with at least 90% reciprocal overlap.		X
gnom_ad_freq_homref_max90pct_rec_ovlp	Maximum frequency of homozygous-reference genotypes among matching gnomAD SVs with at least 90% reciprocal overlap.		X
gnom_ad_freq_het_max90pct_rec_ovlp	Maximum frequency of heterozygous genotypes among matching gnomAD SVs with at least 90% reciprocal overlap.		X
gnom_ad_freq_homalt_max90pct_rec_ovlp	Maximum frequency of homozygous-alternate genotypes among matching gnomAD SVs with at least 90% reciprocal overlap.		X

gnom_ad_lr_artifact_max90pct_rec_ovlp	Long-read artefact flag for matching gnomAD SVs with at least 90% reciprocal overlap.		X
dgv_nstudies_50pct_rec_ovlp	Number of studies in DGV where at least one subject in the study has a variant overlapping the CNV/SV, restricted to 50% reciprocal overlap.	X	X
dgv_nsubjs_50pct_rec_ovlp	Number of subjects in DGV where the variant overlaps the CNV/SV, restricted to 50% reciprocal overlap.	X	X
dgv_pct_ovlp_any	% length of CNV/SV overlapped by DGV region(s) (no cutoff used). The Database of Genomic Variants provides a comprehensive summary of structural variation in the human genome. For more information: http://dgv.tcag.ca/dgv/app/about?ref=GRCh37/hg19 . The DVG was lifted over to obtain the corresponding intervals in the GRCh38 reference genome.	X	X
dgv_50pct_rec_ovlp	% length covered by merged variants in DGV, restricted to those with at least 50% reciprocal overlap.	X	X
dgv_nsubjs_study_50pct_rec_ovlp	number of subjects in the study that detected the variant	X	X
dgv_common_pct	CNV/SV % overlap with DGV with a frequency higher than 1%.	X	X
otg_manta_pct_freq_90pct_rec_ovlp	Percentage frequency of matching SVs called by Manta in the 1000 Genomes dataset, requiring at least 90% reciprocal overlap.		X
otg_delly_pct_freq_90pct_rec_ovlp	Percentage frequency of matching SVs called by Delly in the 1000 Genomes dataset, requiring at least 90% reciprocal overlap.		X
eur_f_dgn_pct_freq_90pct_rec_ovlp	frequency based on internal database – unrelated female samples with European ancestry sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 90% reciprocal overlap, matched by variant type; the type is ignored when complex.		X
all_f_dgn_pct_freq_90pct_rec_ovlp	frequency based on internal database – unrelated female samples sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 90% reciprocal overlap, matched by variant type; the type is ignored when complex.		X
eur_m_dgn_pct_freq_90pct_rec_ovlp	frequency based on internal database – unrelated male samples with European ancestry sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 90% reciprocal overlap, matched by variant type; the type is ignored when complex.		X

all_m_dgn_pct_freq_90pct_rec_ovlp	frequency based on internal database – unrelated male samples sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 90% reciprocal overlap, matched by variant type; the type is ignored when complex.	X	
eur_dgn_pct_freq_90pct_rec_ovlp	frequency based on internal database – unrelated samples with European ancestry sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 90% reciprocal overlap, matched by variant type; the type is ignored when complex.	X	
all_dgn_pct_freq_90pct_rec_ovlp	frequency based on internal database – unrelated samples sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 90% reciprocal overlap, matched by variant type; the type is ignored when complex.	X	
pac_bio_pct_freq_90pct_rec_ovlp	frequency based on SVs called on 502 unrelated samples sequenced by Children’s Mercy Hospital, with at least 50% reciprocal overlap, matched by variant type; the type is ignored when complex. This is a placeholder until a suitable PacBio dataset is found.	X	
nearest_left_exon_boundary	gene with exon/intron junction closest to the left CNV/SV boundary.	X	X
nearest_left_exon_distance	distance of closest exon/intron junction to the left boundary; negative values indicate that the CNV/SV is upstream of the junction.	X	X
nearest_right_exon_boundary	gene with exon/intron junction closest to the right CNV/SV boundary.	X	X
nearest_right_exon_distance	distance of the closest exon/intron junction to the right boundary; negative values indicate that the CNV/SV is upstream of the junction.	X	X
eur_f_dgn_pct_freq_50pct_rec_ovlp	frequency based on internal database – unrelated female samples with European ancestry sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 50% reciprocal overlap, matched by variant type; the type is ignored when complex.	X	
all_f_dgn_pct_freq_50pct_rec_ovlp	frequency based on internal database – unrelated female samples sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 50% reciprocal overlap, matched by variant type; the type is ignored when complex.	X	
eur_m_dgn_pct_freq_50pct_rec_ovlp	frequency based on internal database – unrelated male samples with European ancestry sequenced by Illumina NovaSeq 6000 with average	X	

	coverage of 35x, called by Dragen 3.8.4, with at least 50% reciprocal overlap, matched by variant type; the type is ignored when complex.	
all_m_dgn_pct_freq_50pct_rec_ovlp	frequency based on internal database – unrelated male samples sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 50% reciprocal overlap, matched by variant type; the type is ignored when complex.	X
eur_dgn_pct_freq_50pct_rec_ovlp	frequency based on internal database – unrelated samples with European ancestry sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 50% reciprocal overlap, matched by variant type; the type is ignored when complex.	X
all_dgn_pct_freq_50pct_rec_ovlp	frequency based on internal database – unrelated samples sequenced by Illumina NovaSeq 6000 with average coverage of 35x, called by Dragen 3.8.4, with at least 50% reciprocal overlap, matched by variant type; the type is ignored when complex.	X
pac_bio_pct_freq_50pct_rec_ovlp	frequency based on SVs called on 502 unrelated samples sequenced by Children's Mercy Hospital, with at least 50% reciprocal overlap, matched by variant type; the type is ignored when complex. This is a placeholder until a suitable PacBio dataset is found.	X

AD, autosomal dominant; ALT, alternate; AR, autosomal recessive; ASD, autism spectrum disorder; chr, chromosome; CNV, copy number variant; DGV, Database of Genomic Variants; gnomAD, Genome Aggregation Database; GRCh37/38, Genome Reference Consortium human build 37/38; HPO, Human Phenotype Ontology; ID, identity; INS, insertion; lncRNA, long non-coding RNA; lncRNAKB, long non-coding RNA Knowledgebase; LoF, loss-of-function; MGI, Mouse Genomic Informatics; MPO, Mammalian Phenotype Ontology; OMIM, Online Mendelian Inheritance in Man; PacBio, Pacific Biosciences; POS, position; TCAG, The Centre for Applied Genetics; REF, reference; SV, structural variant; UCSC, University of California Santa Cruz; VCF, variant call file; XL, X-linked.

Appendix C

STR loci derived from STRchive to perform TRGT repeat analysis with LR-WGS data.

Table A3. List of 34 STR loci assessed using TRGT repeat analysis with LR-WGS data.

Chr	Start	End	Disease	Associated gene	Motif(s)	Repeat structure
chr1	1435798	1435818	HMNR7_VWA1	<i>VWA1</i>	GGCGCGGAGC	(GGCGCGGAGC) _n
chr1	57367044	57367125	SCA37_DAB1	<i>DAB1</i>	AAAAT, GAAAT	(AAAAT) _n (GAAAT) _n (AAAAT) _n
chr1	149390803	149390842	NIID_NOTCH2NLC	<i>NOTCH2NLC</i>	GGC	(GGC) _n
chr2	96197067	96197121	FAME2_STARD7	<i>STARD7</i>	AAATG, AAAAT	(AAATG) _n (AAAAT) _n
chr3	63912685	63912728	SCA7_ATXN7	<i>ATXN7</i>	CAG, CCG CAGG, CAGA, CA	(CAG) _n (CCG) _n (CAGG) _n (CAGA) _n (CA) _n
chr3	129172577	129172737	DM2_CNBP	<i>CNBP</i>	CA	(CAGG) _n (CAGA) _n (CA) _n
chr4	39348425	39348483	CANVAS_RFC1	<i>RFC1</i>	AAGGG, ACAGG.	NA
chr5	10356339	10356411	FAME3_MARCHF6	<i>MARCHF6</i>	TTTTA, TTTCA	(TTTTA) _n (TTTCA) _n
chr5	146878728	146878759	SCA12_PPP2R2B	<i>PPP2R2B</i>	GCT	(GCT) _n
chr6	16327634	16327724	SCA1_ATXN1	<i>ATXN1</i>	CTG	(CTG) _n
chr6	170561907	170562017	SCA17_TBP	<i>TBP</i>	GCA	(GCA) _n
chr8	104588972	104588999	OPDM1_LRP12	<i>LRP12</i>	CGC	(CGC) _n
chr8	118366813	118366918	FAME1_SAMD12	<i>SAMD12</i>	TAAAA, TGAAA	(TAAAA) _n (TGAAA) _n (TAAAA) _n
chr9	27573484	27573546	FTDALS1_C9orf72	<i>C9orf72</i>	GGCCCC	(GGCCCC) _n
chr9	69037270	69037304	FRDA_FXN	<i>FXN</i>	A, GAA	(A) _n (GAA) _n
chr9	130681606	130681639	HSAN-VIII_PRDM12	<i>PRDM12</i>	GCC	(GCC) _n
chr12	111598950	111599019	SCA2_ATXN2	<i>ATXN2</i>	CTG	(CTG) _n
chr12	123533720	123533755	OPDM4_RILPL1	<i>RILPL1</i>	GGC	(GGC) _n
chr13	70139353	70139429	SCA8_ATXN8OS	<i>ATXN8OS</i>	CTA, CTG	(CTA) _n (CTG) _n
chr13	102161577	102161726	SCA27B_FGF14	<i>FGF14</i>	GAA	(GAA) _n

chr14	23321472	23321502	OPMD_PABPN1	<i>PABPN1</i>	GCN	(GCN) _n
chr14	92071012	92071053	SCA3_ATXN3	<i>ATXN3</i>	CTG	(CTG) _n
chr15	22786677	22786701	ALS1_NIPA1	<i>NIPA1</i>	GCG	(GCG) _n
chr15	89333589	89333629	CPEO_POLG	<i>POLG</i>	CTG	(CTG) _n TTG(CTG) _n
chr16	66490397	66490466	SCA31_BEAN1	<i>BEAN1</i>	TGGAA, TAGAA	(TGGAA) _n (TAGAA) _n
chr19	13207859	13207898	SCA6_CACNA1A	<i>CACNA1A</i>	CTG	(CTG) _n
chr19	14496041	14496074	OPDM2_GIPC1	<i>GIPC1</i>	CCG	(CCG) _n
chr19	45770204	45770266	DM1_DMPK	<i>DMPK</i>	CAG	(CAG) _n
					GGCCTG,	
chr20	2652733	2652775	SCA36_NOP56	<i>NOP56</i>	CGCCTG	(GGCCTG) _n (CGCCTG) _n
chr22	45795355	45795424	SCA10_ATXN10	<i>ATXN10</i>	ATTCT	(ATTCT) _n
chrX	25013530	25013565	PRTS_ARX	<i>ARX</i>	NGC	(NGC) _n
chrX	25013649	25013697	EIEE1_ARX	<i>ARX</i>	NGC	(NGC) _n
chrX	31284557	31284605	DMD_DMD	<i>DMD</i>	TTC	(TTC) _n
chrX	67545317	67545419	SBMA_AR	<i>AR</i>	GCA	(GCA) _n

LR-WGS, long-read whole genome sequencing; STR, short tandem repeat; TRGT, Tandem Repeat Genotyping Tool.

Table A4. Summary of participant phenotypes, and LR-WGS and LR-RNA-seq outcomes in a cohort of 22 individuals with NMDs.

NMID	Onset	Phenotype	LR-WGS candidate variants	LR-RNA- seq outlier NMD/HPO genes	Ultimate finding	Case resolution:
NM01 (Son to NM02)	C	Static diffuse weakness	4	6/5	<i>ATXN8OS</i> repeat expansion of uncertain pathogenicity	Uncertain candidate variant
NM02 (Father to NM01)	C	Static diffuse weakness, Progression in adulthood Mixed myopathy and neuropathy on EMG	5	Not performed ¹	<i>ATXN8OS</i> repeat expansion of uncertain pathogenicity	Uncertain candidate variant
NM03 (brother to NM06)	P	Periodic pain and weakness with myoglobinuria	4	5/11	Pathogenic <i>DMD</i> CNV (Section 4.1.1)	Solved
NM04	A	Progressive myalgias and stiffness	4	2/9	RNA expression profile supportive of previously identified mosaic D4Z4 deletion (Section 4.3.2)	Solved
NM05	C	Complex multisystemic syndrome with severe weakness	2	3/10	Biallelic, likely pathogenic <i>NUP88</i> variants (Section 4.1.5)	Solved
NM06 (Brother to NM03)	P	Periodic pain and weakness with myoglobinuria	4	Not performed ¹	Pathogenic <i>DMD</i> CNV (Section 4.1.1)	Solved

NM07	C	Static diffuse weakness	5	4/11	No variants of interest	Not solved
NM09	A	Progressive lower limb weakness	4	4/18	Pathogenic reclassification of previously identified <i>FLNC</i> missense variant (Section 4.1.3)	Solved
NM10	A	Progressive proximal and axial weakness Nocturnal hypoventilation	3	4/12	No variants of interest	Not solved
NM12	C	Complex multisystemic syndrome with static weakness and ptosis	11	7/11	LR-RNA-seq evidence of pathogenic impact of previously identified homozygous <i>TTN</i> splice variant (Section 4.1.4)	Solved
NM15	C	Static diffuse muscle weakness	5	3/9	No variants of interest	Not solved
NM16	C	Static diffuse weakness, chronic myalgia	4	5/12	No variants of interest	Not solved
NM17	A	Progressive proximal weakness, ptosis, ophthalmoplegia	7	6/9	RNA expression profile supportive of mitochondrial myopathy (Section 4.3.1)	Solved
NM18	A	Progressive proximal weakness	5	9/13	No variants of interest	Not solved
NM21	A	Progressive proximal weakness	9	4/15	<i>ORAI1</i> novel missense VUS with predicted pathogenicity (Section 4.1.2)	Solved

NM22	A	Progressive proximal and distal weakness	6	Not performed	No variants of interest	Not solved
NM23	C	Static diffuse weakness	3	6/14	No variants of interest	Not solved
NM24	A	Progressive proximal weakness	3	14/8	No variants of interest	Not solved
NM25	C	Complex multisystemic syndrome with static weakness and ptosis	8	6/7	<i>PRX</i> in-frame SV (deletion) of uncertain significance	Uncertain candidate variant
NM26	A	Progressive proximal weakness	5	5/11	No variants of interest	Not solved
NM27	A	Progressive myalgia and cramping	7	3/13	No variants of interest	Not solved
NM28	A	Progressive proximal muscle weakness	5	7/10	No variants of interest	Not solved

¹LR-RNA-seq not performed as skeletal muscle biopsy not performed for individual. ²LR-RNA-seq not performed as skeletal muscle biopsy was previously performed, but no sample was available for the present study. CNV, copy number variant; LR-RNA-seq, long-read RNA sequencing; LR-WGS, long-read whole genome sequencing; NMD, neuromuscular disorder; SV; structural variant; VUS, variant of uncertain significance.

