Improving genetic diagnosis in LGMDs using a new method of DNA sequencing

Christine C. Bruels[1], Hannah R. Littel[1], Carrie Walls[2], Seth A. Stafki[1], Elicia A. Estrella[3], Lauren Brady[4], Mark Tarnopolsky[4], James J. Dowling[5], Carla D. Zingariello[6], Basil T. Darras[3], Partha S. Ghosh[3], Peter I. Karachunski[1], Georgios E. Manousakis[1], Randal C. Richardson[7], Louis M. Kunkel[8], Christina A. Pacak[1], Christopher Faulk[2], Peter B. Kang[1,9]

[1]Greg Marzolf Jr. Muscular Dystrophy Center and Department of Neurology, University of Minnesota Medical School, Minneapolis, MN; [2]Department of Animal Science, College of Food, Agricultural and Natural Resource Sciences, University of Minnesota, St. Paul, MN; [3]Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA; [4]Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada; [5]Division of Neurology, Hospital for Sick Children, Toronto, Ontario, Canada; [6]Department of Pediatrics, University of Florida College of Medicine, Gainesville, FL; [7]Gillette Children's Hospital, St. Paul, MN; [8]Division of Genetics and Genomics, Boston Children's Hospital and Harvard Medical School, Boston, MA; [9]Institute for Translational Neuroscience, University of Minnesota, Minneapolis, MN

Greg Marzolf Jr. Muscular Dystrophy Center

UNIVERSITY OF MINNESOTA
Driven to Discover®

Summary

Study goal: Use a new method of DNA sequencing to identify causative variants in individuals who did not get a genetic diagnosis after clinical genetic testing (Figure 1).

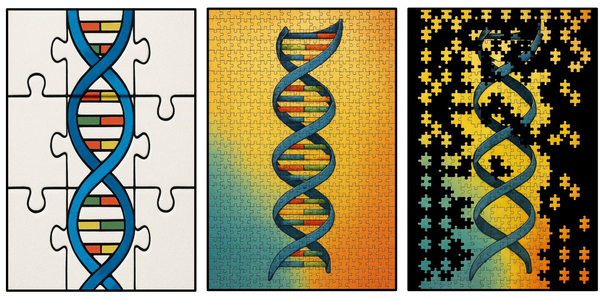
Short term goal: Enroll additional individuals with LGMDs in this research study.

Conclusions: We have used nanopore long read whole genome sequencing to sequence DNA from 68 individuals with muscular dystrophy. We identified and/or clarified pathogenic and likely pathogenic variants (mutations) in 43% of the patient samples we sequenced. This increase in diagnostic capability will enable additional individuals with LGMDs to participate in clinical trials and to receive targeted care including better informed genetic counseling and family planning guidance.

Background

Study rationale: Genetic diagnostic techniques for limb girdle muscular dystrophies (LGMDs) have advanced dramatically in recent decades. And yet, many affected individuals still struggle to obtain definitive genetic diagnoses. Clinical genetic testing, which typically includes sequencing and copy number variant analysis, yields a clear diagnosis in less than half of patients with LGMDs, and standard research-based whole exome and whole genome sequencing identifies causative variants in up to half of the remainder. This leaves a significant number of LGMD patients without a genetic diagnosis. Currently, clinical diagnostic and research based genetic testing relies primarily on short read DNA sequencing methods that generate sequence data in short reads of ~200 nucleotides.

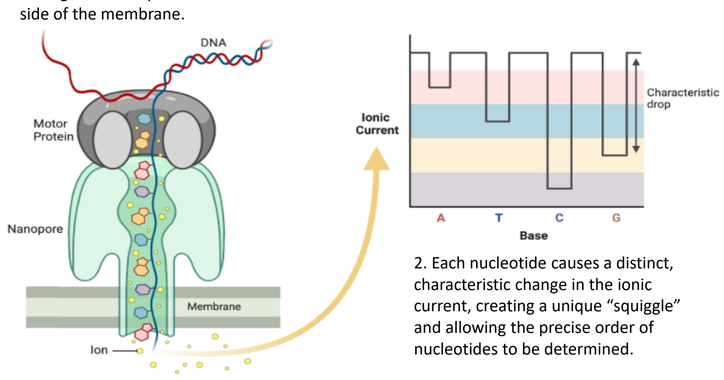
What does our study do differently? We are using nanopore long read whole genome sequencing (LR-WGS) which can sequence the entire length of a DNA molecule. These long reads make it easy to span repetitive regions and structural variants, and to look for variants that are outside the regions typically sequenced in whole exome data.



Assembling the nanopore long reads is like putting together a genetic puzzle that only has 9 large pieces, compared to short read sequencing which is like putting that same puzzle together when it has 900 small pieces. Using whole genome sequencing of the entire genome means that we can see all of the genetic information, while whole exome sequencing (including only the protein coding regions of the genome) is like figuring out that 900 piece puzzle when it doesn't have all of the pieces.

How does nanopore sequencing work? DNA and RNA are made up of 4 nucleotides. Sequencing DNA or RNA reveals the precise order of the nucleotides. In nanopore sequencing, a flow cell is used which contains an array of thousands of biological nanopores. These nanopores are embedded in a membrane that has an ionic current. A motor protein is attached to one end of the DNA molecule and feeds the DNA molecule through the nanopore one nucleotide at a time, disrupting the current. Each of the 4 nucleotides causes a distinct disruption in the current, creating a unique "squiggle".

1. DNA is unwound by the motor protein. One strand is passed through the nanopore to the other side of the membrane.



2. Each nucleotide causes a distinct, characteristic change in the ionic current, creating a unique "squiggle" and allowing the precise order of nucleotides to be determined.

These squiggles can be analyzed to determine the precise order of nucleotides in the molecule. And the entire molecule is sequenced – which can be more than 4 million nucleotides long!

Figures

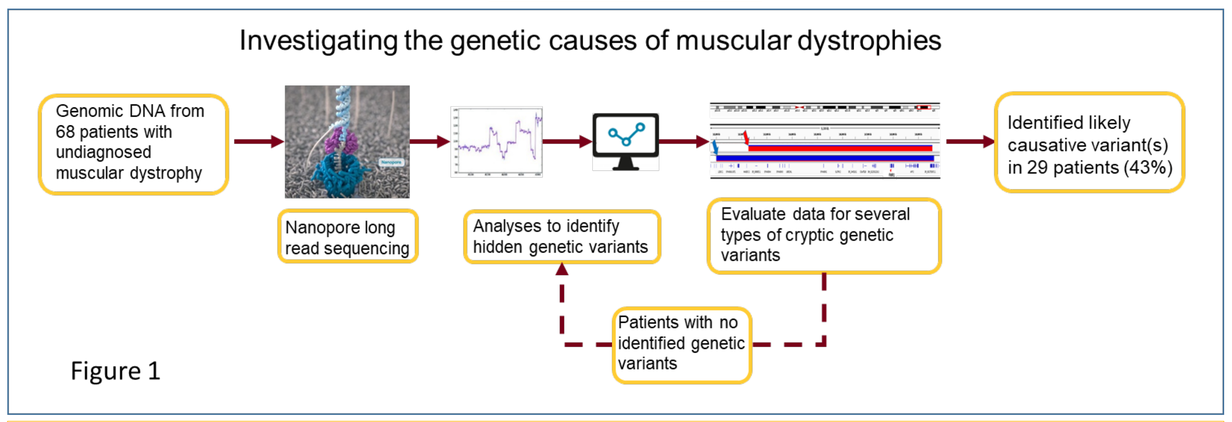


Figure 1

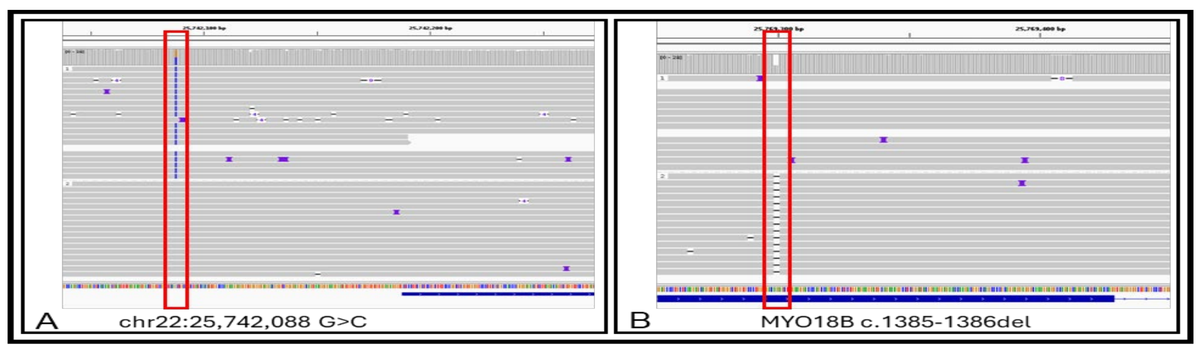


Figure 2. Discovery and phasing of compound heterozygous single nucleotide variants (SNVs) affecting *MYO18B* in patient 56-1-M. A. Detailed view of the SNV chr22:25,742,088 G>C (hg38), which is located within promoter EH38E2154405. **B.** Detailed view of the SNV chr22:25,769,299-25,769,300del (hg38) *MYO18B* c.1385_1386del (p.Glu462fs).

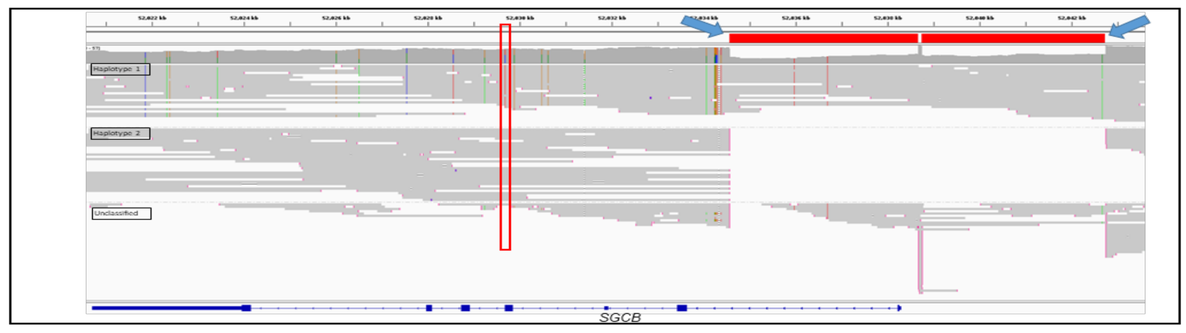


Figure 3. Discovery and phasing of deletions affecting *SGCB* in patient 1479-1. The horizontal red boxes highlighted by blue arrows represent 2 heterozygous deletions with an intervening 86 bp region that is not deleted. Sniffles (v2.0.7) called the deletions from chr4:52,034,569-52,038,672 and chr4:52,038,758-52,042,752, with 86bp between two deletions. The deleted portion spans exon 1 of *SGCB*, as well as multiple regulatory regions for *SGCB*. The initial SNV in *SGCB* (highlighted by a red rectangle), chr4:52029691C>T (hg38), *SGCB* c.416G>A, p.Gly139Asp, was identified as a VUS on clinical genetic testing.

Results & Conclusions

Notable findings include the identification of second heterozygous pathogenic or likely pathogenic variants in the CAPN3 gene in six individuals with LGMDs who each had only one heterozygous pathogenic variant on clinical genetic testing. We also identified or clarified pathogenic or likely pathogenic variants in the DMD gene in six patients with suspected Duchenne Muscular Dystrophy who did not have genetic diagnoses after clinical genetic testing. LRS identified 29 cryptic variants in 68 patients (Table 1). Representative data are presented in Figures 2 & 3. Nanopore data is visualized in Integrative Genome Viewer (IGV) (v2.16.1) in the figures.

Table 1: LRS identified 29 cryptic variants in 68 patients

Gene	# patients
CAPN3	6
DMD	6
ANO5	2
LAMA2	2
DTNA	1
FKTN	1
GNE	1
MYO18B	1
SGCB	1
SGCG	1
Other genes	7

This study demonstrates the ability of nanopore LR-WGS to improve genetic diagnostic yields by identifying single nucleotide variants, identifying large structural variants, and phasing complex variants from a single DNA sample in order to determine their parental origin. This increase in diagnostic capability will enable more individuals with LGMDs to receive targeted care and participate in clinical trials. Increasing our scientific understanding of the genetic mechanisms of these diseases will lead to improved therapeutic development.

Acknowledgments

We thank the patients and their families who contributed the samples used in this study and helped to make this research possible; the University of Minnesota Medical School Department of Neurology; the Minnesota Supercomputing Institute (MSI) at the University of Minnesota; and Research Computing, University of Florida. The long read nanopore sequencing was performed in the Kang laboratory (MinION and PromethION P2 solo) and at the University of Minnesota Genomics Center (GridION). The authors thank the Speak Foundation and the Coalition to Cure Calpain 3 for their support with case ascertainment. This study was funded in part by Sarepta Therapeutics (PBK) and by a Bernard F. and Alva B. Gimbel Foundation grant to Boston Children's Hospital and Harvard Medical School (LMK).