CHRD 2024: Abstract Submission Form

Presenter Name
Alexandria Martin

Presenter Status
Undergraduate Students

Role in the project
Perform Experiments
Analyze Data
Write Abstract

Research Category
Clinical

Title
Investigating key features of neuromuscular disorder genes using machine learning

Background
Neuromuscular disorders (NMDs) encompass a broad, heterogeneous group of pediatric- and adult-onset disorders that primarily affect the peripheral nervous system. They are often caused by genetic alterations that impair skeletal muscle function and result in debilitating symptoms. Obtaining an accurate molecular diagnosis remains a challenge for almost half of NMD cases, with a significant portion remaining undiagnosed due to unidentified causal genes.

Objective
Utilize computational methods to explore the underlying genetic architecture of NMDs, focusing on identifying key features that distinguish NMD-associated genes from other genes in the broader genome.

Methods
Genes implicated in NMDs (n=613) were obtained from the curated GeneTable of NMDs and merged with a comprehensive set of genic features. Machine learning-based feature selection using Boruta in R was then employed to identify the most important discriminative features related to gene complexity (n=90), genetic variation (n=10), expression patterns (n=200), and other general traits (n=9).

Results
As expected, NMD-associated genes exhibit enriched expression in disease-relevant tissues, such as the skeletal muscle and heart. Additionally, these genes displayed broader expression profiles within these tissues, increased complexity (i.e., greater transcript lengths, exon counts), conservation across species, and metrics associated with short tandem repeats.

Conclusion
This study identified several key genic features that may aid in distinguishing NMD-associated genes from the rest of the genome. This may contribute to enhanced identification of causal genes and facilitate earlier interventions. We plan to use these data to assist in NMD gene identification. In this regard, we have already sequenced 22 DNA samples from genetically undiagnosed individuals using long-read whole genome sequencing on the PacBio Revio. We will incorporate the novel data from this study into our bioinformatics pipeline aimed at identifying disease-causing variants in these individuals.

Do you have a table/figure to upload?
No

Authors

Name	Email	Role	Profession
Alexandria Martin	marti29@myumanitoba.ca	Presenting Author	Student
Jessica Hartley	Jessica.Hartley@umanitoba.ca	Co Author	Assistant Professor
Patrick Frosk	pfrosk@hsc.mb.ca	Co Author	Assistant Professor
Britt Drögemöller	Britt.Drogemoller@umanitoba.ca	Co Author	Assistant Professor
Galen Wright	Galen.Wright@umanitoba.ca	Co Author	Assistant Professor