Identifying the link between non-coding regulatory RNAs and phenotypic severity in a zebrafish model of *gmppb* dystroglycanopathy

Grace Smith[1,2], Erin Bailey[3], Michelle Goody[4], Clarissa Henry[3,4], Benjamin King[1,2,3]

[1]The Honors College, [2]Department of Molecular and Biomedical Sciences, [3]Graduate School of Biomedical Science and Engineering, [4]School of Biology and Ecology, University of Maine, Orono, ME, grace.smith@maine.edu

Abstract

Muscular Dystrophy (MD) is characterized by varying severity and time-of-onset by individuals afflicted with the same forms of MD, a phenomenon that is not well understood. Mutations in *gmppb,* an enzyme that glycosylated dystroglycan, cause dystroglycanopathic MD[1]. Like human patients, *gmppb* mutant zebrafish present both mild and severe phenotypes. In order to understand the molecular mechanisms involved, we performed high-throughput RNA Sequencing (RNA-Seq) and small RNA Sequencing at 4 and 7 days-post-fertilization (dpf) in mild and severe *gmppb* mutants and controls. We hypothesize that variable phenotypes in *gmppb* mutants are due to differences in gene regulation; therefore, we identified differentially expressed (DE) long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) - both potent genetic regulators. In the 4dpf severe mutants, we identified DE "MD-relevant" Ensembl-annotated genes that were predicted targets of DE miRNAs – identifying 55 of these interactions. We utilized a novel method of visualizing gene expression networks by generating co-expression miRNA networks and subsequently removing miRNA nodes to identify miRNAs that maintain network stability. We identified 95 potential lncRNAs for further analysis. By integrating analyses of both coding *and* non-coding genes, we hope to better understand the molecular mechanisms of dystroglycanopathy, highlighting potential phenotypic modulators.

Gmppb zebrafish mutants, like human patients, exhibit varying phenotypes and age of onset

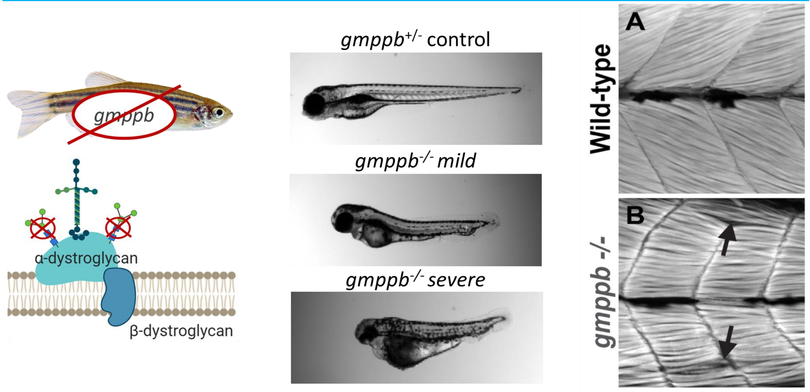
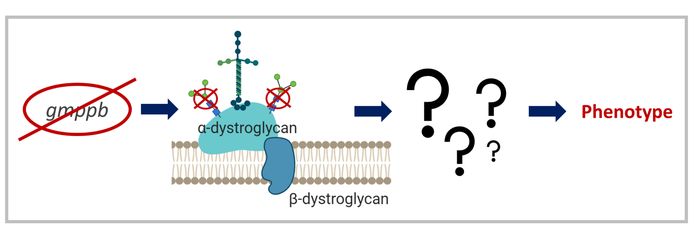


Fig 1. Mutations in *gmppb* are expected to decrease alpha-dystroglycan glycosylation. In 2dpf *gmppb* homozygous mutants, fish could be categorized as either mild or severe based on birefringence. The mutants exhibited lower skeletal muscle density and improperly formed Myotendinous Junctions.

Research Goals

A. Determine whether phenotypic differences in the mild and severe mutants are supported by RNA sequencing data
B. Identify DE protein-coding genes, miRNAs, and lncRNAs
C. Identify and predict interactions between miRNAs and protein-coding genes
D. Use co-expression networks to visualize miRNA expression



We propose that genetic regulation plays a role in modulating phenotype

RNA Sequencing supports genetic differences in mild vs. severe *gmppb* mutants.

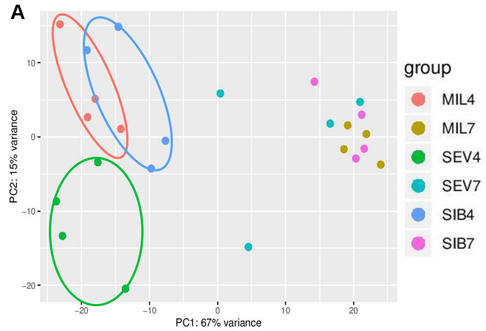


Fig 2. A) Principal component analysis of mutants shows genetic differences between control, mild mutant, and severe mutant zebrafish samples at 4dpf using RNA Sequencing and quadruplicate sample. B) Number of differentially expressed protein-coding genes according to adjusted and unadjusted p values <0.05 . C) Number of differentially expressed miRNAs according to adjusted and unadjusted p values <0.05.

Identification of DE miRNAs and predicted DE ensembl annotated gene targets

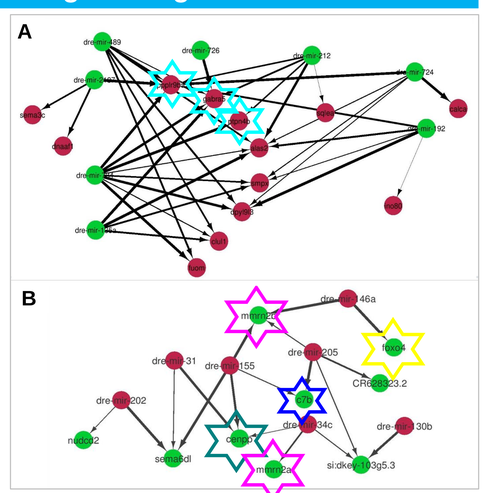


Fig 3. Nodes represent either DE miRNAs (p<0.05) or DE ensembl annotated genes (adj p<0.05). MiRNA targets predicted using TargetScan-Fish and are represented with arrows. A) Green nodes are upregulated miRNAs; red nodes are down-regulated ensembl genes. B) Red nodes are downregulated miRNAs; Green nodes are upregulated ensembl genes.

A total of 55 predicted miRNA/mRNA interactions were identified. Genes were categorized to identify "MD-relevant" interactions.

Functions:

Cell growth | Immune system | Angiogenesis | Neuronal function | Skeletal muscle atrophy

Co-expression miRNA networks & sequential hub removal

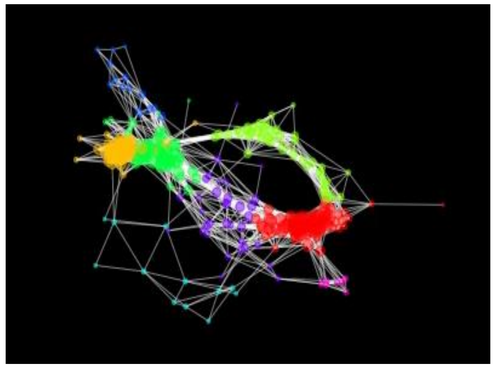
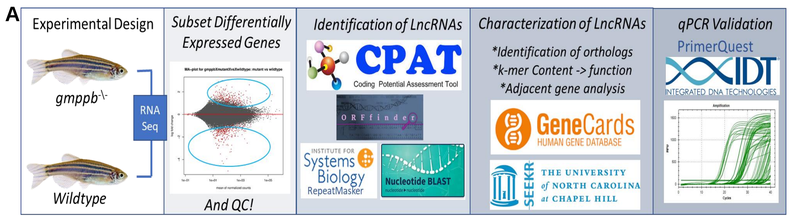


Fig 4. A) An example network. Each node represents a DE miRNA that has a similar pattern of expression to other miRNAs in the sample if connected by an edge (co-expression). By sequential "hub" removal, miRNAs with the highest number of edges, one can potentially identify miRNAs that are integral to the biological systems stability. Hub removal that causes large changes in characteristic path length, are candidates for further research. B) A video illustrating this process using 4/7dpf severe DE miRNAs.

Table 1. Network attack was performed in each of the samples, and miRNAs that when removed caused drastic changes in characteristic path length are indicated. Dre-Mir_103 was present in both the mild and severe networks.

4dpf 7 dpf Sibling	4dpf 7dpf Mild	4dpf 7dpf Severe
Dre-Mir-204-P2a-5p	Dre-Mir_103-P3b-3p	Dre-Mir-17-P2a1-5p
Dre-Mir-15-P2a2-5p	Dre-Mir-126-P1-3p	Dre-Mir-132-P1a-3p
Dre-Mir-132-p2a-5p		Dre-Mir-130-P3b1-3p
		Dre-Mir-103-P3a-3p
		Dre-Let-7-P1b-5

Identification of potential novel long non-coding RNAs



B Comparison of Ensembl Annotated and Novel Genes in 4dpf severe mutants

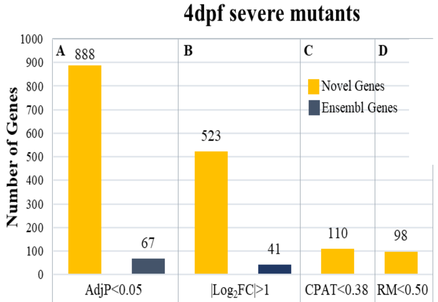


Fig 5. A) Workflow (2) used to identify lncRNAs based on low alignment to annotated genes, low-coding potential, low percentage of repetitive bases, lack of open reading frames, etc. B) Filtering of unannotated DE transcripts in 4dpf severe (adj p <0.05) based on $|Log_2FC|>1$, coding potential <0.38, and less than 50% repetitive bases. A total of 98 transcripts in the 4dpf severe mutants were identified for further analysis via the aforementioned workflow to indicate the likelihood they are lncRNAs.

Conclusions and Future Directions

- Phenotypic differences in mild and severe *gmppb* mutant zebrafish were supported by RNA Sequencing data.
- 55 interactions between DE miRNA and DE predicated ensembl annotated gene targets were identified with opposite expression patterns.
  - These targets included proteins that function in "MD-relevant" processes including cell growth, angiogenesis, the immune response, neuronal function, and skeletal muscle atrophy.
- A novel method of visualizing miRNA networks was used to identify miRNAs for further research.
- 98 potential lncRNAs were identified for further analysis.
- Compare expression of miRNA 5' and 3' arms in different samples to look for miRNA class switching .
- Design and validate lncRNA, miRNA, and protein-coding gene expression using qPCR

References and Acknowledgements

1 https://rarediseases.org/rare-diseases/duchenne-muscular-dystrophy
2. King, B. L. *et al. npj Regen.* 2018.

Maine INBRE NIH P20 GM103423NIH R01 AR075836; University of Maine System Research Reinvestment Fund; University of Maine Medicine Seed Grant