Arts & Sciences Electronic Theses and Dissertations

Arts & Sciences

Winter 12-15-2022

Generation, Characterization, and Treatment of Functional Zebrafish Models for Musculoskeletal Disorders

Julia Nicole Whittle
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Biology Commons

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Developmental, Regenerative, and Stem Cell Biology

Dissertation Examination Committee:
Christina Gurnett, Chair, Mentor
Kory Lavine, Co-Chair
Gabriel Haller
Aaron Johnson
Mayssa Mokalled
Chris Weihl

Generation, Characterization, and Treatment of Functional Zebrafish Models for
Musculoskeletal Disorders
by
Julia Whittle

A dissertation presented to

Washington University in St. Louis

in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

December, 2022
St Louis, MO

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIS ....Adolescent Idiopathic Scoliosis

BMD ....Bone Mineral Density

CADD ....Combined Annotation Dependent Depletion

CRISPR.. Clustered Regularly Interspaced Short Palindromic Repeats

DA...Distal Arthrogryposis

DA2A... Distal Arthrogryposis 2A (Freeman-Sheldon Syndrome)

DA2B ....Distal Arthrogryposis 2B (Sheldon-Hall Syndrome)

del..deletion

dpf ....days post fertilization

ECM...Extra Cellular Matrix

EHR...Electronic Health Record

Fn .... n^{th} Filial Generation

FNDC1 .... Fibronectin domain containing 1

GFP ....green fluorescent protein

gnomAD... The Genome Aggregation Database

gRNA .... guide RNA

GWAS.. Genome Wide Association Study

HA...calcium-hydroxyapatite

HDR ....Homology-Directed Repair

hpf .... hours post fertilization

ICD...International Classification of Diseases

kDa..Kilodaltons

MAC ....Minor Allele Count

Mhc .... Myosin Heavy Chain (*Drosophila*)

MicroCT.. Micro Computed Tomography

mpf .... months post fertilization

MYH3 ....Myosin Heavy Chain 3 (Myosin, Heavy Chain 3, Skeletal Muscle, Embryonic)

MyHC .... Myosin Heavy Chain

NHEJ... Non-Homologous End Joining

OR .... Odds Ratio

pMOI...Polar Moment of Inertia

QC .... Quality Control

smyhc1..*slow myosin heavy chain 1*

SNP .... Single Nucleotide Polymorphism

TALEN .... Transcription activator-like effector nuclease

UTR...Untranslated region

VoI .... Variant of Interest

VUS.. Variant of Unknown Significance

WES ....Whole Exome Sequencing

<u>Acknowledgements</u>

After so much time at Wash U, it's hard to believe I'm almost finished. Science is never really "finished," and there's so much to do here and on these projects- in some ways it's difficult to let go and pass the baton to other lab members. But it's always a collaborative endeavor, and I'm happy to leave the work in the hands of my wonderful colleagues at Wash U, and whoever else might join these labs in the future.

On the subject of collaborative endeavors, I want to thank everyone at wash U who has been so instrumental is making my projects a success. First, I want to give so many thanks to my wonderful mentor, Dr. Christina Gurnett. Chris has always been endlessly supportive of my research- by the nature of this work, many things go wrong, many experiments don't work, and a lot of troubleshooting must be done. You quickly realize how much heartbreak goes into a research project. But Chris has always been so supportive and encouraging- her faith in my ability to do the job has lifted me up and made it so much easier to persevere and reach my goals.

I want to thank the rest of the Gurnett lab as well- although we all have wildly different areas of research, I've always received great feedback on my practice talks and presentations. I want to thank Lili for setting the groundwork for these projects, allowing me to pick up and excel in my research. I want to especially thank Kevin for helping me get my experiments to work, and allowing me to bounce ideas off of him- always offering a great new perspective on an idea (as well as many corny dad jokes).

Many thanks to my committee members; Kory Lavine, Aaron Johnson, Mayssa Mokalled, Chris Weihl, and Gabe Haller. Your suggestions and perspective have been incredibly helpful in guiding my projects (and finally deciding when to graduate).

A special thank you to Lila Solnica-Krezel's lab for a great help during my start here at Wash U, and continued support throughout my grad school career- I especially want to thank Diane for being a wonderful friend to talk to about science, politics, or raspberry gardening.

Thank you to all the wonderful friends I've made in and out of lab- the commiserating has always been excellent stress relief. I don't know what I'd do without other grad students working so hard as well.

Thank you all the friends I've made outside of school (especially Will) for taking my mind off of science for a bit with Magic or book club or anything else.

Lastly, I want to thank for family for their unwavering support. From art to science, anything I've ever wanted to do has seemed possible.

Julia Whittle

Washington University in St Louis

December 2022

ABSTRACT OF THE DISSERTATION

Generation, Characterization, and Treatment of Functional Zebrafish Models for
Musculoskeletal Disorders

by

Julia Whittle

Doctor of Philosophy in Biology and Biomedical Sciences
Developmental, Regenerative, and Stem Cell Biology

Washington University in St. Louis, 2022

Professor Christina Gurnett, Chair & Mentor
Professor Kory Lavine, Co-Chair

Musculoskeletal disorders in humans present with a broad range of phenotypes that vary in severity, onset, and genetic cause. The musculoskeletal system is comprised of complex and diverse tissues that make up the majority of total body mass. Thus, mutations in many different genes with differing structure and function can produce phenotypically similar developmental and morphological defects in the muscle and skeleton. In addition, the muscular and skeletal systems develop non-autonomously, wherein development and function of each will affect development of the other. Thus, mutations in genes responsible for development of this complex system have the potential to influence morphogenesis outside of their tissue of expression. Many different gene defects have the potential to cause musculoskeletal disorders and to properly treat afflicted patients it is necessary to understand the underlying causes of pathogenesis.

One such example of pathogenic morphogenesis is Distal Arthrogryposis (DA), a group of ten congenital disorders characterized by contractures of the joints of the distal limbs that presents at birth. The most severe form of DA, Freeman-Sheldon syndrome (DA2A), is most commonly associated with autosomal dominant missense mutations in the motor domain of MYH3, a sarcomeric embryonic myosin heavy chain gene expressed in developing skeletal muscle. While this gene is expressed only transiently in skeletal muscle tissue, symptoms present in the structure of skeletal joints and persist beyond birth.

To examine the disease mechanism responsible for developing skeletal muscle and joint contracture and to explore treatment options for afflicted patients, we created the most commonly observed human DA2A pathogenic variant ($MYH3^{R672H}$) in the analogous gene in zebrafish ($smyhc1^{R673H}$). The heterozygous zebrafish had contractures of the spine that modeled the phenotype of DA2A patient's muscle contractures. We observed shortened muscle fibers and sarcomeres, disorganized skeletal muscle tissue, and disrupted motor activity. The phenotype was rescued with an actin-myosin inhibitor, para-aminoblebbistatin, suggesting that contractures are caused by constitutive muscle contraction.

While we were able to characterize the phenotype of this pathogenic variant, many variants of unknown significance have been recorded in the motor domain of *MYH3*. Given the labor and time required to generate missense mutants using traditional methods, a higher throughput means of mutant generation is necessary to examine the function of dozens of reported variants. To develop a higher throughput method of analyzing single variants, we created zebrafish lines that would be expected to enable high-throughput screening. A GFP reporter tag at the end of *smyhc1* is used as a marker for successful homology directed repair. A deletion is induced where reported variants cluster in the human analog that also serves to put GFP out of frame. In-frame repair and

integration of a variant of interest via a repair template returns GFP into frame, allowing for efficient screening of F0 embryos.

Finally, we analyzed a fibronectin domain containing 1 (*fndc1)* zebrafish knockout to better understand its association with human scoliosis. *fndc1* zebrafish mutants displayed a higher bone mineral density and slightly larger vertebral centra, though scoliosis was not present. Using zebrafish to model musculoskeletal disease, we were able to elucidate disease mechanisms of distal arthrogryposis and scoliosis, as well as develop more efficient methods for variant analysis in zebrafish.

Chapter 1: Introduction and Perspective

1.1 Preface

Parts of this chapter were adapted from the following manuscript:

Whittle, Julia, Aaron Johnson, Matthew B. Dobbs, and Christina A. Gurnett. "Models of distal arthrogryposis and lethal congenital contracture syndrome." *Genes* 12, no. 6 (2021): 943.

Author Contributions:

J.W. and C.A.G. wrote the first draft of the manuscript. J.W., A.J., M.B.D. and C.A.G. wrote or critically reviewed the manuscript, and reviewed and approved the final version. All authors have read and agreed to the published version of the manuscript.

1.2 Arthrogryposis

Arthrogryposis (arth = joint; grp = curved; osis = pathological state) describes a broad range of disorders consisting of multiple congenital joint contractures presenting at birth (Hall, 2014). About 1 in 3000 live births presents with some form of arthrogryposis, many of which are nonprogressive and improve with physiotherapy (Hall, 1997; Hall, 2014; Ravenscroft et al, 2020).

The core cause of arthrogryposis is fetal akinesia, or lack of fetal movement, that results in contractures forming in the joints (Hall, 1997; Hall, 2014; Ravenscroft et al, 2020). Movement is required for normal joint development; it influences the structure of the joints and promotes cellular signaling to guide normal tissue development. Mechanical forces also influence bone morphology, which affects the organization of chondrocytes, bone elongation, and differential growth, to impact the shape of bones as they develop. Fetal akinesia impairs joint formation by inhibiting both normal development of joints and bones, which may lead to joint and bony fusions. Furthermore, tension is required for normal tendon development, which forms a connection

between bone and muscle (Felsenthal & Zelzer, 2017). Arrested movement during development has significant impact on the formation of the skeleton, joints, muscle, and connective tissues.

Akinesia can be caused by both genetic and environmental factors. Over 400 genes are associated with arthrogryposis broadly, encapsulating a wide diversity of genes affecting different pathways, including axon structure, circulatory development, or synaptic transmission (Kiefer & Hall, 2019). Extrinsically, maternal disease or exposures, uterine space limitation, and decreased blood supply are also root causes for contraction defects that may lead to arthrogryposis (Hall, 1997; Hall, 2014).

1.2.1 Distal Arthrogryposis

A subset of arthrogryposis is described as Distal Arthrogryposis (DA), a group of congenital contractures that predominantly affect the joints of the distal limbs, including the hands, wrists, ankles, and feet. Clinically, the lower extremity manifestations commonly include clubfoot and vertical talus. There are currently 10 classifications of DA, including Sheldon-Hall syndrome (DA2B) and Freeman-Sheldon syndrome (DA2A) (Bamshad et al, 2009; Beck et al, 2014; Scala et al, 2018; Toydemir et al, 2006). Freeman-Sheldon syndrome is considered the most severe form of DA, also presenting with facial contractures (Toydemir et al, 2006). This range of disorders can negatively impair patient quality of life through muscle weakness, generalized pain, and diminished motor function (Kimber et al, 2012).

1.2.2 Current treatment methods

Currently, DA patients are offered supportive care to improve quality of life, including occupational therapy, physical therapy, and surgery (Desai et al, 2020). While these treatments improve outcome for patients, they often fall short of complete restoration of range of motion in the joints and functionality. This strategy also fails to address underlying causes for DA, such as

muscle weakness and impaired neurotransmission. Therefore, further investigation is necessary to understand the impact of disease variants which will allow us to determine the most effective treatment options for patients.

1.2.3   Genetics of Distal Arthrogryposis

While over 400 genes are associated with arthrogryposis, including a wide diversity of genes affecting different pathways, mutations associated with DA are observed at a high rate in skeletal muscle and neural genes (Bamshad et al, 2009; Desai et al, 2020). Specifically, the most common genetic causes of severe DA are autosomal dominant missense mutations in *MYH3*, a myosin heavy chain (MyHC) gene expressed early in human development (Alvarado et al, 2011; Beck et al, 2013; Beck et al, 2014; Scala et al, 2018; Tajsharghi et al, 2008; Toydemir et al, 2006; Zhao et al, 2022).

1.3 Embryonic Myosin Heavy Chain

MyHCs are large, dimerizing, ATP-dependent motor proteins that bundle together in thick filaments to drive muscle contraction. They consist of a large ATPase domain and long tail domain linked by a coiled coil. These proteins dimerize and bundle together in thick filaments that drive muscle contraction by binding to actin in the sarcomere (Carraro & Catani, 1983; Schiaffino et al, 2015; Zhao et al, 2022) (Figure 1.1). MyHCs are critical for both structural organization of the sarcomere and driving contraction.

It is common for embryonic isoforms of proteins to be expressed transiently in development before adult isoforms populate the tissue (Schiaffino et al, 2015; Tajsharghi et al, 2008; Zhao et al, 2022). Developmental MyHC genes are conserved in vertebrates, with these isoforms identified in birds and fish as well as mice and humans. In developing skeletal muscle, embryonic and neonatal forms of MyHCs are expressed. They are downregulated significantly as

adult MyHCs begin expression (Schiaffino et al, 2015; Smerdu, 2002). However, there is evidence that some adult muscle tissues continue to express developmental forms of MyHC in adulthood. Muscle such as the extraocular muscle and jaw-closing muscles appear to continue to express embryonic MyHC in adulthood. Typically, these genes are co-expressed with other MyHC isoforms (Schiaffino et al, 2015; Wieczorek et al, 1985). There is also evidence that these genes resume expression during muscle regeneration (Schiaffino et al, 2015; Tajsharghi et al, 2008; Zhao et al, 2022).

The evolutionary advantage of distinct developmental MyHC proteins is not well known. It is possible that these proteins have structural properties that allow them to organize the sarcomere more effectively. There is also functional evidence that these proteins consume ATPase at a lower rate than adult isoforms (Sieck et al, 2003). This suggests that embryonic MyHCs may have a lower contractile capacity (Schiaffino et al, 2015). Indeed, human muscle fibers containing embryonic MyHC isoforms have been found to generate less than $1/10^{th}$ the force of adult muscle fibers (Racca et al, 2013). While the physical properties of these proteins have been studied, the evolutionary advantage of dedicated embryonic MyHC isoforms is not well known. It is possible that the lower force production is well-suited to delicate embryonic tissue. This may explain why these proteins have been observed to be expressed in delicate muscles such as the extraocular muscle as well (Wieczorek et al, 1985).

1.3.1   *MYH3*

MYH3 is the earliest expressed embryonic MyHC gene in humans where it is predominantly expressed in myotubes destined to become fast-twitch myofibers (Toydemir et al, 2006). *MYH3* expression peaks during fetal development, and is significantly downregulated before birth (Karsch-Mizrachi et al, 1989; Smerdu, 2002). Missense mutations in *MYH3* are

strongly associated with DA clinically, and contribute to multiple subtypes including DA1, DA2A, and DA2B which have varying degrees of severity (Alvarado et al, 2011; Beck et al, 2013; Beck et al, 2014; Scala et al, 2018; Toydemir et al, 2006; Zhao et al, 2022). Notably, phenotypes caused by these mutations persist into adulthood without surgical intervention, long after *MYH3* expression has ceased expression.

1.3.2 *smyhc1*

slow myosin heavy chain 1 (*smyhc1*) is an embryonic isoform of MyHC expressed in developing zebrafish slow muscle. *In situ* data has detected *smyhc1* expression as early as 12 hours post fertilization (hpf) in the somites (Rauch et al, 2003). Specifically, the adaxial cells of the somites, adjacent to the notochord, express *smyhc1*. These cells migrate away from the notochord toward the lateral sides of the trunk, and differentiate into slow muscle tissue (Devoto et al, 1996).

1.4 Embryonic MyHC models for DA

The most common genetic causes of DA are autosomal dominant missense mutations in *MYH3* (Scala et al, 2018; Schiaffino et al, 2015; Toydemir et al, 2006; Zhao et al, 2022). *MYH3* mutations have been identified in multiple DA syndromes, including distal arthrogryposis, type 1 (DA1), DA2A, and distal arthrogryposis type 2B (DA2B) (also called Sheldon-Hall syndrome) (Alvarado et al, 2011; Beck et al, 2013; Toydemir et al, 2006). MYH3 mutations have additionally been identified in patients with multiple pterygium syndrome, and spondylocarpotarsal synostosis syndrome, which is associated with vertebral, carpal, and tarsal fusions in addition to joint contractures (Cameron-Christie et al, 2018; Carapito et al, 2016; Chong et al, 2015; Zieba et al, 2017).

Various disease models have been developed to examine the mechanisms behind DA pathogenesis and to test therapeutic interventions. Molecular and single-cell studies are useful for

precisely examining the effects of DA-causing variants on the affected proteins and tissues (Racca et al, 2015; Walklate et al, 2016). Access to human tissues, particularly from muscle biopsies, has facilitated molecular analysis for research, yet is clinically useful only in select cases as patient variability is so high (Kimber et al, 2012). Protein modeling can help predict the impact of various amino acid substitutions on molecular interactions and protein activity (Del Angel et al, 2020; Yip et al, 2004). However, animal models are necessary to analyze the effect of single gene variants on organisms on scales larger than single cells. The interaction between cells and tissues in animal models allows us to study non-cell autonomous effects of genes and genetic variants. The effect of zygosity and gene dosage may also be better studied in animal models to assess interactions between normal and abnormal gene products. Animal models are also useful for studying experimental interventions that may improve patient quality of life and outcome, acting as stand-ins for potential human patients.

Models of human disease are rapidly becoming more sophisticated with the ability to knock-in single nucleotide variants and create conditional (tissue specific or time-dependent) knockouts (Ablain et al, 2015; Zhang et al, 2012a; Zhuo et al, 2021). Loss-of-function alleles, which are often easier to generate, provide critical information about gene function, but may not fully explain autosomal dominant phenotypes, in which gain-of-function or dominant negative alleles can cause markedly different phenotypes. Therefore, in many cases loss-of-function alleles do not act as analogs for pathogenic alleles. Conditional knockouts, while very helpful in defining gene function, rarely replicate the human phenotype in its entirety, but may be required when early lethality limits further study. These methods allow researchers to design models that more accurately represent these human conditions and replicate pathogenic effects broadly or in specific tissues.

1.4.1 Single cell and single molecule studies

Single molecule and single cell studies are useful to examine the precise impact of a variant on protein function. The effects of amino acid substitutions are difficult to predict without mechanistic examination. These substitutions can affect protein folding, ligand interaction, protein stability, and more (Stefl et al, 2013; Yip et al, 2004; Zhang et al, 2012b). Single-molecule studies facilitate understanding the effect of a missense variant on protein function and can later be translated into an understanding of how small mechanical differences affect tissues and whole body systems (Walklate et al, 2016). Missense variants can be directly studied in human skeletal muscle biopsies as well (Kimber et al, 2012). However, these are not routinely performed for DA diagnosis, which makes these studies challenging. To study the mechanistic link between DA phenotype and gene variant, Racca et al. performed contractility studies on isolated muscle cells and myofibrils derived from biopsied muscle tissue from DA2A patients. They found that a DA2A-associated *MYH3* variant inhibited cross-bridge detachment, thereby slowing muscle relaxation and lowering force production (Racca et al, 2015). A later study replicated these results while examining multiple *MYH3* variants associated with DA2. In addition to slower actin-myosin detachment, ATP binding and ATPase activity were lower in variant *MYH3* molecules (Walklate et al, 2016).

Development of single cell models, in which *MYH3* variants are exogenously expressed or overexpressed from a plasmid or virus, have been limited by the difficulties of expressing large genes, like *MYH3*, *in vitro*. For similar reasons, RNA-induced expression induction in multicellular organisms is also limited. Likewise, because many features of DA are due to complex relationships between different cell types, co-cultures of muscle cells with tenocytes and bone may

be required to recapitulate the human condition. Thus, many investigators have preferred to study DA genes in whole organisms.

1.4.2 *Drosophila* models

Drosophila melanogaster (fruit flies) are useful tools for studying muscle function and myofibril assembly, particularly as introduction of single variants are traditionally simpler in this system compared to other models. *MYH3* and the Drosophila myosin heavy chain gene, Mhc, are highly conserved. Drosophila have the advantage of having only a single myosin heavy chain, which eliminates the possible obfuscation of effects due to compensation by other myosin heavy chain analogs. Therefore, the effect of variants on protein function can be examined in a setting without other myosin heavy chain isoforms.

Drosophila transgenic models have been generated by overexpressing Mhc constructs containing DA variants (Das et al, 2019). Guo et al. predicted that a DA1 variant would perturb a hydrophobic interaction, while a DA2B mutation would introduce a hydrogen bond that was not present in the wild type (Guo et al, 2020). The effect of these predicted interactions was tested mechanistically in Drosophila models. Muscle fibers containing the DA alleles were extracted and found to have lower actin affinity, reduced power output, and increased stiffness, which may explain the motor deficits (Guo et al, 2020).

Morphological studies of Drosophila skeletal muscle expressing three DA2A Mhc transgenes (R672H, R672C, and T178I) showed branching and splitting defects, which were most severe in the R672C variant, which caused Z-discs to be split and malformed. In addition, Z-disc distance was shorter in the transgenic flies indicating an overall shortening of the sarcomeres, perhaps due to an enhanced contractile state of the myofibers. Presumably, the shortened sarcomeres observed in Drosophila contribute to the formation of contractures in human patients.

Indeed, ATPase activity was reduced in these transgenic flies, leading to functional defects in muscle activity (Das et al, 2019).

In studying the intact adult Drosophila, Guo et al. showed that the MhcF437I mutants had a much longer lifespan than MhcA234T mutants, consistent with the less severe phenotype of DA1 patients compared to DA2B patients (Guo et al, 2020). In addition, the researchers found that both mutants displayed aberrant myofibril assembly, as well as misaligned sarcomere structure including distorted M and Z lines. Again, this phenotype was more severe in A234T mutants than in the F437I mutants. Mhc^{F437I} mutants displayed essentially normal myofibers and sarcomeres, while Mhc^{A234T} mutants had small myofibers with disrupted morphology, as well as abnormal sarcomeres (Guo et al, 2020). Das et al. also found that decreased climbing capability of adult flies also correlated with the phenotypic severity in humans (Das et al, 2019).

1.4.3 *C. elegans* models

Like Drosophila, *Caenorhabditis elegans* (*C. elegans*) has also been used to study myosin heavy chain genes (Gil-Gálvez et al, 2020). There are many advantages of *C. elegans* and *Drosophila* for disease modeling, including large numbers of progeny, knowledge of ontogeny of individual cells, and ease of functional studies for drug screening. However, as described in Gil-Galvez et al., the evolutionary distance, and differences in number of myosin heavy chain genes between invertebrates and humans makes it difficult to determine whether the gene being studied has the same function, particularly in terms of spatial and temporal expression, as its human counterpart. Furthermore, Gil-Galvarez et al. caution against overexpression studies in C. elegans broadly, citing interference with muscle cell function overall (Gil-Gálvez et al, 2020). The major drawback of invertebrate models is, quite obviously, the lack of skeletal structures which limits their use in understanding the complex relationship between muscle, nerve, and bone.

1.4.4   Limitations of current models

The musculoskeletal system is complex, with an entwined relationship between bone, muscle, nerve, and joints. Because the previously described models are invertebrate or *in vitro*, this system cannot be holistically examined analogously to human patients. While these models are helpful in analyzing kinetic properties and molecular effects of these mutations, it is difficult to examine the functional disease mechanism by which MyHC mutations lead to skeletal and joint deformities. Therefore, there is a need for vertebrate models for DA to more accurately depict the disorder and assess treatment options.

1.5 Scoliosis

Scoliosis is defined as a lateral curvature of the spine. Severity is calculated by the Cobb angle. Typically, a deviation of 10 degrees or greater is defined as scoliosis (Carman et al, 1990; Cobb, 1948) (Figure 1.2). In severe cases, patients may be treated with surgery or a back brace to correct the curvature. Scoliosis can be congenital and comorbid with other musculoskeletal conditions (Janicki & Alman, 2007). However, it commonly arises in patients with no apparent cause, which is known as idiopathic.

1.5.1   Adolescent Idiopathic Scoliosis

Scoliosis is the most common spinal deformity in children, characterize by a lateral spinal curve with a Cobb angle > 10° (Konieczny et al, 2013). In more than 80% of patients, scoliosis is idiopathic, meaning the cause in that specific patient is unclear, and uncoupled with apparent spinal deformities such as bony fusions. It also most commonly arises during adolescence (Lonstein, 1994). Adolescent idiopathic scoliosis (AIS) is a common disorder, with a prevalence reported up to 5.2% in current population studies (Konieczny et al, 2013). Both prevalence and severity are more common in females than males (Lonstein, 1994; Tajsharghi et al, 2008). In severe cases,

surgery and bracing are required. This condition typically runs in families, implying underlying genetic causes for its onset and progression (Wynne-Davies, 1968). Although the genetic factors that contribute to severe AIS are largely unknown, it is suspected that the disorder is a complex trait caused by multiple genetic risk factors contributing to an additive effect on morphology.

1.5.2   Genetics of AIS

While AIS has been observed to run in families, the specific genetic factors that contribute to onset and progression are largely unknown. It follows that the mechanistic effects of these mutations that lead to scoliosis are unclear. While some common genetic loci such as LBX1 (ladybird homeobox 1), GPR126 (G-protein coupled receptor 126)7, BNC2 (basonuclin 2), PAX1 (paired box 1), and SLC39A8 (Solute Carrier Family 39 Member 8) have been associated with AIS, reported loci account for less than 5% of AIS heritability (Haller et al, 2018; Kou et al, 2019; Kou et al, 2013; Ogura et al, 2015; Sharma et al, 2015; Takahashi et al, 2011). To further complicate AIS genomic study, these genes encode a wide variety of protein function, from cell migration regulatory agents to solute transporters (Gross et al, 2000; Park et al, 2015). The pathways attributable to AIS are therefore largely unknown, and the mechanistic basis for AIS onset and progression are unclear.

To further complicate study, AIS is suspected to be a complex trait, where multiple common and rare variants culminate in a deleterious effect that would not be detectable with a single mutant. AIS is not a Mendelian disorder and the phenotype can only rarely be linked to a specific genetic locus that segregates cleanly. Therefore, modern methods of genome analysis are necessary to find the complex gene associations that additively culminate in AIS (Risch, 2000).

Currently, genome-wide association studies (GWAS) of common variants and rare variant gene burden analysis are strong tools for determining genetic risk factors for AIS. GWAS studies

aim to connect complex traits to the genetic loci associated with them. With a large enough sample size, statistical analysis of an association of common variants to a disease phenotype can be detected when compared to a control set. They are useful in detecting single nucleotide polymorphisms (SNPs) in association with diseases (such as AIS) on a population-wide scale (Cantor et al, 2010; Visscher et al, 2012). On the other hand, rare variant gene burden analyses seeks to identify genes in which rare variants are found more commonly in a disease cohort versus a control cohort (Guo et al, 2018).

Conducting these studies in families to find genes and variants associated with AIS gives us a broader understanding of alleles that carry risk for AIS development, and how they may combine to manifest in differing scoliosis severity. In addition, documentation of mutations in different genes that contribute to AIS will allow us to develop models, test and determine the mechanisms leading to the deformity in specific cases, and how best to approach treatment precisely for each patient.

1.6 Dissertation objectives

Disorders such as DA and AIS are caused by a broad array of gene variants, so that it is difficult to study without analyzing a multitude of variants and genes. Furthermore, patients' cases may require specialized treatment. To approach this, a precision-based approach to research will allow us to better understand the multitude of mechanisms driving these disorders and tailor treatment accordingly. To do this, we used both a knock-in DA zebrafish model, and a GWAS exome analysis to identify genes associated with severe AIS.

In Chapter 2, we created and characterized a DA zebrafish model by knocking in the most common variant associated with DA2A, (the most severe form of DA), $MYH3^{R672H}$. In zebrafish, the analogous variant is $smyhc1^{R673H}$, and these mutants phenocopied the human disorder. In

human DA patients, the *MYH3* missense mutations are always heterozygous and present with contractures of the hands and feet. In our heterozygous zebrafish, we observed notochord contractures in early larval stages that persisted to adulthood. We also functionally assessed muscle structure and function in these fish. This phenotype was corrected by treating the fish with para-aminoblebbistatin, a MyHC ATPase inhibitor that prevents MyHC and actin from interacting. This suggests that this allele causes a DA phenotype by causing constitutive contraction in the developing muscle.

In Chapter 3, we continue this investigation by developing a higher-throughput method of examining variants of interest in a zebrafish model. By inserting a GFP tag at the end of *smyhc1*, we are able to screen for successful variant integration more easily. A large region of interest excised from the gene puts the GFP out of frame, and a successful repair with a variant of unknown significance (VUS) will put the GFP back in frame, allowing for efficient screening of F0 embryos. Because the tissue *smyhc1* is expressed in does not fuse, future studies may be able to examine and characterize the sarcomeric structure and muscle phenotype of GFP positive cells in mosaic F0 zebrafish.

Finally, in Chapter 4, we perform a gene burden analysis on whole exome data in patients affected by severe AIS compared to a control cohort. We were able to identify an enrichment of core matrisome genes, with the most significant enrichment of variants in *FNDC1* (fibronectin domain containing 1). To gain a better understanding of the function of this gene and how it may contribute to scoliosis development, we examined *fndc1* mutant zebrafish. In these mutants, we found an increase in bone mineral density in the vertebrae, a phenotype that also scaled with mutant allele frequency.

This work uses zebrafish models to better understand the musculoskeletal disorders of DA and AIS. These conditions are currently treated symptomatically, but a greater understanding of the underlying genetic causes will allow us to intervene to prevent disease or progression. An understanding of the genes at play also gives us a broader perspective of their role in development.
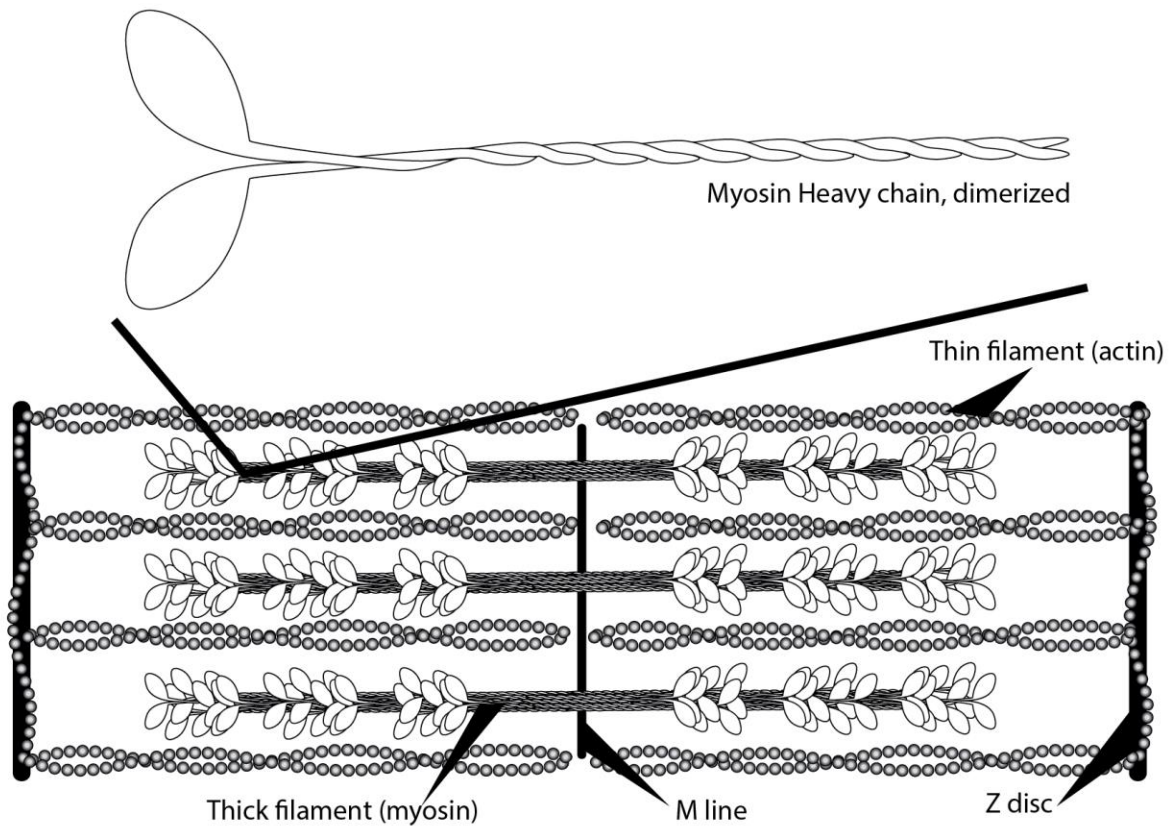
Figure 1.1 Diagram of dimerized myosin heavy chain and sarcomere. Myosin heavy chain dimerizes and bundles into thick filaments that drive contraction of the sarcomere.
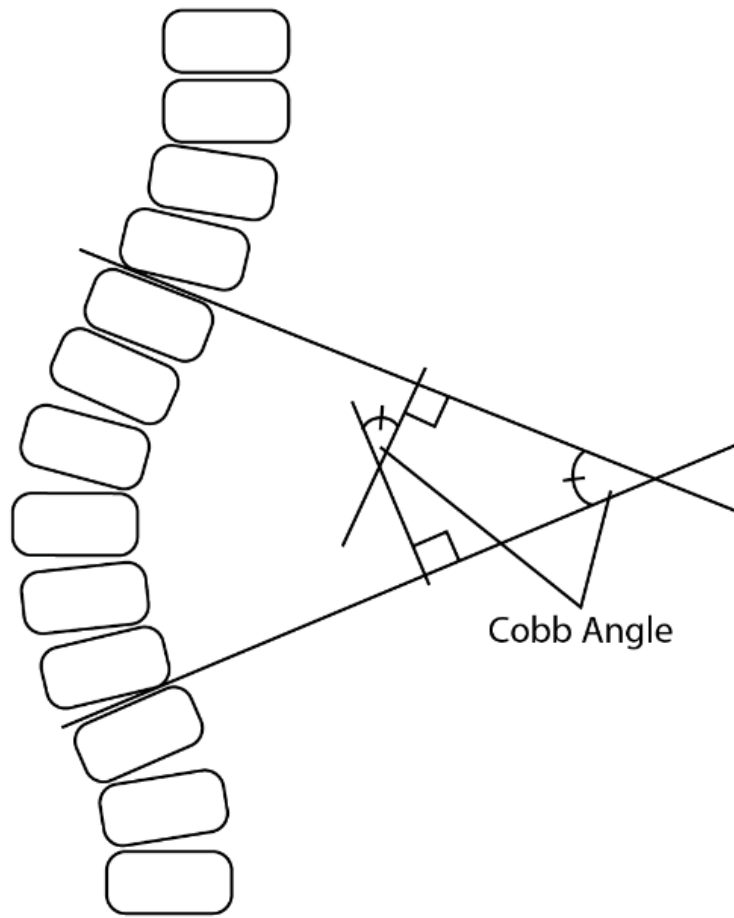
Figure 1.2 Diagram of Cobb angle calculation.

Chapter 2: *MYH3*-associated distal arthrogryposis zebrafish model is normalized with para-aminoblebbistatin

Authors

Julia Whittle[1], Lilian Antunes[1], Mya Harris[2], Zachary Upshaw[2], Diane S. Sepich[3], Aaron N. Johnson[3], Mayssa Mokalled[3], Lilianna Solnica-Krezel[3], Matthew B. Dobbs[2], Christina A. Gurnett[1,2,4]

Affiliations

Departments of [1]Neurology, [2]Orthopedic Surgery, [3]Developmental Biology, and [4]Pediatrics; Washington University in St. Louis, St. Louis, MO, USA.

Whittle, Julia, Lilian Antunes, Mya Harris, Zachary Upshaw, Diane S. Sepich, Aaron N. Johnson, Mayssa Mokalled, Lilianna Solnica-Krezel, Matthew B. Dobbs, and Christina A. Gurnett. "MYH 3-associated distal arthrogryposis zebrafish model is normalized with para-aminoblebbistatin." *EMBO molecular medicine* 12, no. 11 (2020): e12356.

2.1 Abstract

Distal arthrogryposis (DA) is group of syndromes characterized by congenital joint contractures. Treatment development is hindered by the lack of vertebrate models. Here, we describe a zebrafish model in which a common *MYH3* missense mutation (R672H) was introduced into the orthologous zebrafish gene *smyhc1* (*slow myosin heavy chain 1*) (R673H). We simultaneously created a *smyhc1* null allele (*smyhc1⁻*), which allowed us to compare the effects of both mutant alleles on muscle and bone development, and model the closely related disorder, spondylocarpotarsal synostosis syndrome. Heterozygous *smyhc1*$^{R673H/+}$ embryos developed notochord kinks that progressed to scoliosis with vertebral fusions; motor deficits accompanied the disorganized and shortened slow-twitch skeletal muscle myofibers. Increased dosage of the mutant allele in both homozygous *smyhc1*$^{R673H/R673H}$ and transheterozygous *smyhc1*$^{R673H/-}$ embryos exacerbated the notochord and muscle abnormalities, causing early lethality. Treatment of *smyhc1*$^{R673H/R673H}$ embryos with the myosin ATPase inhibitor, para-aminoblebbistatin, which decreases actin–myosin affinity, normalized the notochord phenotype. Our zebrafish model of *MYH3*-associated DA2A provides insight into pathogenic mechanisms and suggests a beneficial therapeutic role for myosin inhibitors in treating disabling contractures.

2.2 Introduction

Distal arthrogryposis (DA) describes a group of congenital musculoskeletal syndromes characterized by contractures in the joints of the hands and feet. Classification systems currently describe ten closely related DA subtypes, the most severe of which is Freeman-Sheldon Syndrome (also called distal arthrogryposis, type 2A [DA2A]). Children born with DA2A present with characteristic contractures of the hands, clubfeet, and facial contractures. They often develop scoliosis (Beck et al, 2013; Toydemir et al, 2006). Complications of DA, particularly DA2A,

include difficulty eating, respiratory issues, and impaired speech and mobility. Currently, supportive treatments including bracing, physical therapy, and surgical interventions, are the only therapeutic options for DA patients; however these therapies are often suboptimal (Boehm et al, 2008; Stevenson et al, 2006).

The most common genetic causes of DA are autosomal dominant missense mutations in the *MYH3* gene, encoding the embryonic myosin heavy chain (MyHC) that is expressed first during slow skeletal muscle development. *MYH3* expression peaks during fetal development, and is significantly downregulated after birth (Cameron-Christie et al, 2018; Chong et al, 2015; Schiaffino et al, 2015). *MYH3* mutations have been identified in multiple DA syndromes, including distal arthrogryposis, type 1 (Alvarado et al, 2011), DA2A, and distal arthrogryposis, type 2B (DA2B)(also called Sheldon-Hall syndrome) (Beck et al, 2013; Toydemir et al, 2006). *MYH3* mutations have additionally been identified in patients with multiple pterygium syndrome (Chong et al, 2015), and spondylocarpotarsal synostosis syndrome, which is associated with vertebral, carpal and tarsal fusions in addition to joint contractures (Cameron-Christie et al, 2018; Carapito et al, 2016; Zieba et al, 2017).

MyHCs are large, dimerizing, ATP dependent motor proteins that bundle together in thick filaments to drive muscle contraction. The most common DA-causing *MYH3* mutations cluster in the ATPase region of the MyHC motor domain (Fitts, 2008; Toydemir et al, 2006). The dominant inheritance pattern of DA mutations implies a hypermorphic, neomorphic, or antimorphic nature. The functional effects of DA-associated *MYH3* missense mutations include slowing the muscle relaxation time and prolonging the muscle fiber contracted state (Racca et al, 2015; Walklate et al, 2016). While the vast majority of described DA mutations are dominant missense variants, the recent identification of autosomal recessive *MYH3* variants in spondylocarpotarsal synostosis

syndrome suggests that some mutations may also contribute to disease pathogenesis through a loss of function, or hypomorphic mechanism (Cameron-Christie et al, 2018).

Understanding the mechanism by which dominant or recessive *MYH3* mutations cause contractures or bony fusions has been limited by poor access to human tissue, particularly during early development when the gene is most highly expressed. Analysis of muscle biopsies from adults with the *MYH3* R672C mutation, which is the most common recurrent variant causing DA2A and DA2B (Toydemir et al, 2006), revealed increased relaxation time and impaired sarcomeric cycling (Racca et al, 2015). Recombinantly expressed *MYH3* R672C, R672H, and T178I mutations in cultured cells also caused marked abnormalities in molecular kinetic properties including slower cycling time (Walklate et al, 2016). While the molecular effects of pathogenic *MYH3* mutations have been modeled in *Drosophila* (Rao et al, 2019), mutations have yet to be studied in vertebrate models where their impact on muscle and skeletal development can be concurrently assessed.

To address the need for vertebrate models of DA, we created a zebrafish line in which the most common human DA2A *MYH3* mutation, R672H, was precisely edited into the corresponding amino acid of the *slow myosin heavy chain 1* gene (*smyhc1^{R673H}* [stl583]) (Roy et al, 2001). This gene is expressed in the somites as early as 5–9 somite stages, and in the lateral slow muscle of the trunk, as evidenced by previously conducted *in situ* hybridization (Bessarab et al, 2008; Li et al, 2020; Rauch et al, 2003). We simultaneously created a *smyhc1* null allele (*smyhc1$^-$* [stl582]), which allowed us to compare the effects of both mutant alleles on slow skeletal muscle and bone development.

Remarkably, the *smyhc1$^-$* allele in the homozygous state caused transient muscle abnormalities in larvae and late onset scoliosis, whereas the *smyhc1^{R673H}* allele resulted in

notochord kinks, abnormally shortened muscle, and vertebral fusions with the severity of the phenotype correlating with allele dosage. The early notochord phenotype was suppressed or normalized with para-aminoblebbistatin, a drug that reduces myosin-actin affinity by inhibiting myosin ATPase activity, which directly inhibits muscle contraction (Várkuti et al, 2016). Because removing muscle contraction suppressed the abnormal notochord phenotype, we interpret this result as supporting our hypothesis that the $smyhc1^{R673H}$ allele causes hypermorphic muscle contraction that creates excessive tension on the developing zebrafish notochord. This secondarily results in kinking of the notochord that contributes to later vertebral fusions. Furthermore, by showing that the $smyhc1^{R673H}$ zebrafish model replicates key aspects of the congenital disorder that can be normalized with myosin inhibitors, we have demonstrated its value for evaluating therapeutics for use in human DA patients.

2.3 Results

2.3.1 $smyhc1^{R673H}$ is an autosomal dominant mutation that acts in a dose-dependent manner

To study the mechanism by which *MYH3* mutations cause DA, we genetically engineered a mutant zebrafish line in which a single nucleotide substitution was introduced via homologous recombination into exon 16 using a donor oligonucleotide and TALENs (Figure 2.1A-C). The resultant zebrafish $smyhc1^{R673H}$ missense substitution is orthologous to the most common *MYH3* mutation causing DA2A in humans, R672H (Figure 2.1A, B). We concurrently generated a zebrafish $smyhc1^{-/-}$ line containing a frame-shifting seven base pair deletion that is predicted to result in a premature truncation in exon 16 and transcript loss by nonsense mediated decay. Indeed, Smyhc1 protein expression was not detected in $smyhc1^{-/-}$ embryos at 24 hours post fertilization (hpf) when Smyhc1 is normally most highly expressed in wild-type ($smyhc1^{+/+}$) embryos (Figure 2.1C-E). In contrast, $smyhc1^{R673H}$ heterozygotes and homozygotes produce detectable Smyhc1

protein, confirmed via immunohistochemistry staining (Fig 2.1E). The brief period of Smyhc1 protein expression is demonstrated by its absence by 48 hpf in $smyhc1^{+/+}$ larvae on Western blot (Fig 2.1D) and in all genotypes by 3 dpf (days post fertilization) as shown by immunohistochemistry (Supplemental Figure 2.1).

Consistent with prior studies that used morpholinos to transiently reduce *smyhc1* expression (Codina et al, 2010)*, smyhc1$^{-/-}$* embryos displayed no gross phenotype in larval stages despite being paralyzed until 48 hpf (described below) (Figure 2.2A, 2.5A). In contrast, we noted that nearly half of the embryos heterozygous for the *smyhc1^{R673H}* allele displayed mild notochord kinks or slight curves in the body axis beginning at 48 hpf, while fewer showed more severe kinks, often in multiple locations along the body axis. Furthermore, all homozygous *smyhc1$^{R673H/R673H}$* zebrafish manifested severely abnormal morphology and developed multiple notochord bends and kinks that markedly compressed and disfigured the body axis and were apparent by ~30 hpf (Figure 2.2A, B; Supplemental Figure 2.3A). These early notochord kinks resemble those seen in the *accordion* zebrafish mutant, which has a muscle relaxation defect due to a mutation in the sarcoplasmic reticulum Ca^{2+} - ATPase pump (*atp2a1*) (Hirata et al, 2004).

To determine whether the presence of a wild-type copy of *smyhc1* influences the phenotype of fish harboring a single *smyhc1^{R673H}* allele, we crossed *smyhc1$^{-/-}$* and *smyhc1$^{R673H/+}$* mutant lines. The absence of a wild-type copy of *smyhc1* resulted in more severe morphological abnormalities. In fact, the phenotype of s*myhc1^{R673H}* transheterozygous embryos (*smyhc1$^{R673H/-}$*) was identical to that of s*myhc1$^{R673H/R673H}$* embryos, in which similarly severe notochord kinks and bends were observed (Figure 2.2A). The observed partial rescue of the *smyhc1*^{R673H} phenotype via dilution of the allele with the wild-type allele, and the failure of *smyhc1*^{R673H} homozygotes, heterozygotes, or

transheterozygotes to phenocopy $smyhc1^{-/-}$ fish support the hypothesis that the $smyhc1^{R673H}$ substitution is hypermorphic or neomorphic.

2.3.2 $smych1^{R673H}$ mutation is homozygous lethal and reduces survival of heterozygotes

Because of the disruptive embryonic phenotypes of $smyhc1^{R673H}$ heterozygotes and homozygotes, we systematically determined the impact of genotype on survival. Because no $smyhc1^{R673H/R673H}$ individuals survive to adulthood for breeding, we instead incrossed $smyhc1^{R673H/+}$ fish, and genotyped the resulting progeny of various ages to determine the ratio of surviving $smyhc1^{+/+}$:$smyhc1^{R673H/+}$:$smyhc1^{R673H/R673H}$ fish (Figure 2.3A). In contrast to the expected Mendelian ratio of 1:2:1, we discovered that $smyhc1^{R673H/R673H}$ embryos did not survive past larval stages (Figure 2.3B). We determined that the ratio of $smyhc1^{R673H/R673H}$ embryos sharply declined between 1 and 2 dpf, with the remaining homozygotes perishing before juvenile stages. Many $smyhc1^{R673H/+}$ fish survive into adulthood and are fertile, however survival is reduced compared to wild type, with a decline continuing throughout adulthood. The $smyhc1^{R673H}$ allele increases mortality of the fish in a dose-dependent manner, consistent with the severity of the gross phenotype.

2.3.3 $smyhc1^{-/-}$ adults develop late-onset spinal curvature

We noticed an increase in spinal curvature of $smyhc1^{-/-}$ fish as they reached adulthood, and decided to examine the phenotype further. Despite the apparent short and early temporal expression of $smyhc1$, which is strongly expressed at 24 hpf but markedly reduced by 48 hpf (Figure 2.1D), its disruption has lasting effects into adulthood. Although $smyhc1^{-/-}$ larvae appeared morphologically normal (Figure 2.2A), nearly all developed spinal curvatures in adulthood (Figure 2.4A, Supplemental Figure 2.2). To determine whether the observed curvature is associated with vertebral fusions, alizarin red/alcian blue staining of bone and cartilage, respectively, was

performed on adult zebrafish between 1 and 2 years old. Alizarin red staining of adult $smyhc1^{-/-}$ fish revealed variable spinal curvatures along the length of the spine in both lateral and sagittal planes (Figure 2.4B). Most $smyhc1^{-/-}$ fish were afflicted with mild-to-moderate curvature (Figures 2.4C; Supplemental Figure 2.2,3B). Despite the apparent curvature of $smyhc1^{-/-}$ fish, vertebral fusions were not observed. This is consistent with the lack of notochord abnormalities displayed in larval stages in $smyhc1^{-/-}$ fish, a phenotype associated with later vertebral fusions in other mutant fish lines (reviewed in (Ellis et al, 2013a; Ellis et al, 2013b; Gray et al, 2014).

2.3.4 $smyhc1^{R673H/+}$ adults have both spinal curvature and vertebral fusions

Similar to the $smyhc1^{-/-}$ fish, the $smyhc1^{R673H/+}$ fish also displayed spinal curvatures in adulthood, but the fish were often shortened in length (Figure 2.4A; Supplemental Figure 2.2). In addition, the spinal curves developed much earlier; notochord kinks and bends that were seen as early as 48 hpf often preceded the later development of structural vertebral abnormalities (Figure 2.2A; Figure 2.4B, much as those previously reported in *leviathan*/*col8a1a* zebrafish mutants (Gray et al, 2014).). Upon observation of alizarin red stained bone, most $smyhc1^{R673H/+}$ adults were found to have moderate-to-severe vertebral abnormalities, with complex and variable vertebral fusions often involving multiple vertebrae and their jaws often had a closed mouth appearance (Figures 2.4B, C; Supplemental Figure 2.2,2.3B). Similar to leviathan, the vertebral fusions in $smyhc1^{R673H/+}$ were present at regions of severe spinal curve and were not detected in regions where the spine was straight. Spinal curves were predominantly present in the lateral plane and were nearly all restricted to the distal tail. However, we noted that some $smyhc1^{R673H/+}$ larvae with proximal notochord kinks did not survive to adulthood, which suggests that the propensity toward distal spinal curves may be due to a selection bias rather than a developmental predisposition to

distal scoliosis. These data suggest that larval notochord abnormalities predispose *smyhc1*$^{R673H/+}$ fish to vertebral fusions in adulthood.

2.3.5 *smyhc1* **mutants have motor deficits**

Because *smyhc1* is a myosin heavy chain gene critical for motor function, and because of the observed effects of *smyhc1* mutations on gross anatomy described above, we assessed the effects of *smyhc1* mutations on muscle function by quantifying the movement of fish at various times during development. The *smyhc1* gene is expressed embryonically, as early as 10-13 somites as evidenced by *in situ* hybridization (Bessarab et al, 2008; Li et al, 2020; Rauch et al, 2003); therefore, we first assessed embryonic movements by counting the number of light-induced twitches beginning at 24 hpf when embryos first respond to light stimulation (Kokel & Peterson, 2011; Saint-Amant & Drapeau, 1998) (Figure 2.5A). The *smyhc1*$^{-/-}$ embryos were completely paralyzed up to 48 hpf, after which they exhibited normal movements in response to light stimulation. There was not a statistically significant difference between *smyhc1*$^{R673H/+}$ and their wild-type siblings until 34 hpf, when mutants displayed a slight decrease in the number of twitches (Figure 2.5A). At early stages when muscle function was grossly measured as the ability to twitch in response to light, the *smyhc1*^{R673H} allele had only minimally apparent effects while lack of Smyhc1 abolished muscle function in early embryonic stages.

Motor function was also assessed at 6 dpf to determine later effects of *smyhc1* mutations (Figure 2.5B). Automated motion tracking software was used to quantitatively measure spontaneous swimming distance, which was not possible at earlier stages due to their small size and transparency. The total distance traveled over 5 minutes was less in *smyhc1*$^{R673H/+}$ larvae compared to their wild-type siblings. The *smyhc1*$^{-/-}$ larvae swam a significantly shorter distance

than both of the other genotypes (Figure 2.5B). At larval stages, it appears that the effects of *smyhc1* mutations, even after apparent downregulation of the gene, have an enduring effect on muscle function and locomotion.

Muscle function was measured in adults at 6 months post fertilization (mpf) by assessing endurance upon exposure to a gradually escalating flow of water in a swim tunnel (Figure 5C). The *smyhc1*$^{R673H/+}$ fish fatigued in the swim tunnel sooner than *smyhc1*$^{+/+}$ and *smyhc1*$^{-/-}$ fish. No significant difference was noted in time spent swimming between the *smyhc1*$^{+/+}$ and *smyhc1*$^{-/-}$ fish (Figue 2.5D). While the *smyhc1*$^{-}$ allele had a more severe effect on muscle function in larval stages compared to the *smyhc1*^{R673H} allele, this result provides evidence that *smyhc1*$^{-/-}$ mutants recover a normal degree of motor function in adulthood.

2.3.6 *smyhc1* R673H mutants have disorganized muscle and shortened myoseptal intervals

Because *smyhc1* is the first and earliest myosin heavy chain expressed in slow skeletal muscle (Devoto et al, 1996), we examined the effect of mutant *smyhc1* alleles on morphological muscle development. Phalloidin-rhodamine was used to stain filamentous actin in *smyhc1*$^{+/+}$, *smyhc1*$^{-/-}$, s*myhc1*$^{R673H/+}$, and s*myhc1*$^{R673H/R673H}$ larvae. At 24 hpf, the *smyhc1*$^{-/-}$ muscle was highly disorganized and had no distinguishable sarcomeres or myofibers (Figure 2.6A, C). Wavy strands of filamentous actin and numerous puncta of actin bundles were present (Figure 2.6A). In contrast, *smyhc1*^{R673H} heterozygotes and homozygotes formed sarcomeres, but the filamentous actin was unevenly distributed across the myofiber, which occasionally had a flared appearance near the myosepta (Figure 2.6A). At 3 dpf, discrete sarcomeres became evident in s*myhc1*$^{-/-}$ myofibers; however, they were poorly aligned with each other and with the myosepta. s*myhc1*$^{-/-}$ larvae were also found to have centralized nuclei (Figure

2.6B,D). Organized sarcomeres were also present in *smyhc1*^{R673H/+} muscle, but myofibers were thin, becoming bulbous at their termini, leading to poorly defined myosepta. The *smyhc1*^{R673H/R673H} myofibers were also had poorly defined myoseptal boundaries; bundles of actin were commonly observed (Figure 2.6B,D).

To better define the role of *smyhc1* in sarcomere development, Z-disks were labeled with anti-α-actinin antibodies. At 24 hpf, Z-disks were assembled in *smyhc1*^{+/+}, *smyhc1*^{R673H/+}, and s*myhc1*^{R673H/R673H} larvae, while s*myhc1*^{−/−} larvae failed to develop any Z-disk structure (Figure 2.6C). However, by 3 dpf, s*myhc1*^{−/−} larvae formed Z-disk structures, with a disorganized appearance consistent with the thin filament morphology (Figure 2.6D). We then quantified the distance between Z-disks to determine whether sarcomere length was affected by the R673H mutation or lack of Smyhc1. s*myhc1*^{R673H/R673H} and *smyhc1*^{R673H/+} genotypes displayed significantly shorter sarcomere lengths compared to wild type at both 24 hpf and 3 dpf, with a large range of sarcomere length variability in *smyhc1*^{R673H/R673H} larvae (Figure 2.6G,H). At 3 dpf, *smyhc1*^{−/−} fish displayed a significantly shorter Z-disk interval than any other genotype.

To determine whether the *smyhc1* alleles affected the overall somite length as was previously observed in the hypercontractile *accordion* mutant (Hirata et al, 2004), we measured the distance between myosepta in slow skeletal muscle. Myoseptal intervals were significantly reduced in s*myhc1*^{R673H/R673H} and *smyhc1*^{R673H/+} muscle at 24 hpf and 3 dpf. *smyhc1*^{−/−} larvae did not display a shortened myoseptal distance (Figure 6E,F). Because both the Z-disk interval and myoseptal interval are reduced in *smyhc1*^{R673H} mutants, we hypothesize that the shortened myoseptal interval results from increased tension shortening the length of the sarcomere, and subsequently the myofiber, as was proposed for the *accordion* mutant (Hirata et al, 2004).

Conversely, the shortened Z-disk interval observed in $smyhc1^{-/-}$ fish is not accompanied by a shortened myoseptal distance. These data, in addition to the observed disorganization of sarcomere alignment in $smyhc1^{-/-}$ fish, suggest a different mechanism, perhaps in which the structural effects caused by the lack of Smyhc1 result in both a shortened Z-disk interval and sarcomere misalignment.

2.3.7 Cell death in $smyhc1^{R673H/R673H}$ zebrafish

To explore the effects of the $smyhc1$ alleles on cell viability, larvae at 24 hpf and 3 dpf were stained with Caspase-3 antibodies(Dalgin & Prince, 2015; Dalgin et al, 2011). Increased signal of the cleaved Caspase-3 protein indicates an activation of the apoptosis pathway (Sorrells et al, 2013). While there was no difference in cleaved Caspase-3 at 24hpf, $smyhc1^{R673H/R673H}$ larvae had increased cleaved Caspase-3 in the muscle tissue at 3dpf (Supplemental Figure 2.4). This suggests that increased tissue stress and muscle cell morphological decline due to the $smyhc1^{R673H}$ allele may lead to an increase in cell death in the homozygous state.

2.3.8 Para-aminoblebbistatin rescues $smyhc1^{R673H}$ mutant phenotypes

To understand the proposed mechanism and explore new therapeutic approaches to prevent contractures and disability in DA patients, we evaluated the ability of select pharmacological agents to normalize the phenotype of $smyhc1^{R673H}$ heterozygous and homozygous embryos. Tricaine methanesulfonate, an anesthetic that acts as a neuromuscular blocking agent by directly inhibiting neuronal sodium channels (Attili & Hughes, 2014), was applied to dechorionated embryos from 12 hpf to 72 hpf. Although the drug completely suppressed embryonic movement as expected, it failed to improve the ventral body curvature or reduce the development of notochord kinks in $smyhc1^{R673H}$ heterozygotes and homozygotes (Supplemental Figure 2.5). This suggests

that the contractures observed in *smyhc1*[R673H] mutants arise independently of stimulated muscle contraction, but may alternatively be intrinsic to the muscle and present even without neural stimulation.

Myosin ATPase inhibitors are being developed for the treatment of cardiomyopathy caused by similar mutations of cardiac myosin heavy chain genes (Green et al, 2016), therefore we sought to determine whether drugs that directly inhibit function of myosin (Kovács et al, 2004) could reduce the adverse phenotypic effects of the s*myhc1*[R673H] allele. Blebbistatin, a myosin ATPase inhibitor, which reduces actin-myosin affinity, was applied to dechorionated embryos at a concentration of 25 µM from 24 hpf to 48 hpf (Wang et al, 2015). Blebbistatin prevented the development of notochord kinks in both *smyhc1*[R673H/+] and s*myhc1*[R673H/R673H] embryos (Rauscher et al, 2018) (Supplemental Figure 2.5B). However, numerous developmental defects, including pericardial edema, dorsal tail curvature, and small eyes were observed in all fish exposed to this drug, including wild-type larvae, consistent with the known cytotoxic effects of blebbistatin that are independent of its myosin inhibition (Várkuti et al, 2016).

A blebbistatin derivative, para-aminoblebbistatin, which is more photostable and less cytotoxic (Várkuti et al, 2016), was therefore applied to dechorionated embryos at 25 µM for 24 or 48 hours, starting at 24 hpf. While para-aminoblebbistatin does not share the cytotoxic effects of blebbistatin, its effect on cardiac myosin appears to cause lethal pericardial edema in both mutant and wild-type embryos (Várkuti et al, 2016). However, similarly to blebbistatin, its effect on skeletal myosin completely abolished notochord kinks and bends in both *smyhc1*[R673H/+] and *smyhc1*[R673H/R673H] embryos at 2 dpf (Figure 7A). This effect persisted to 3 dpf (Figur 7B). The molecular inhibition of actin–myosin interaction removes the ability of the muscle to contract. This effect suppresses the *smyhc1*[R673H] phenotype, suggesting that the allele

causes notochord kinking as a secondary effect of constitutive myofiber contraction, independent of neural stimulation.

A MYH3
Chromosome 17

Transcription →

10628K 10630K 10632K 10634K 10636K 10638K 10640K 10642K 10644K 10646K 10648K 10650K 10652K 10654K 10656K

10640K 10641K 10642K

R672H

B smyhc1
Chromosome 24

Transcription →

40656K 40658K 40660K 40662K 40664K 40666K 40668K

40662K 40663K

R672H
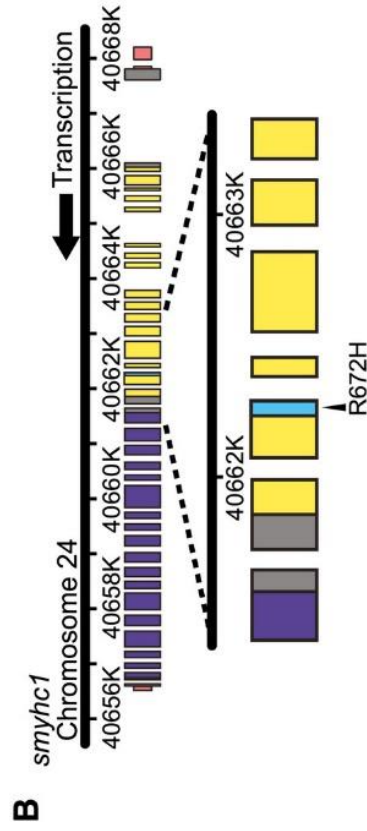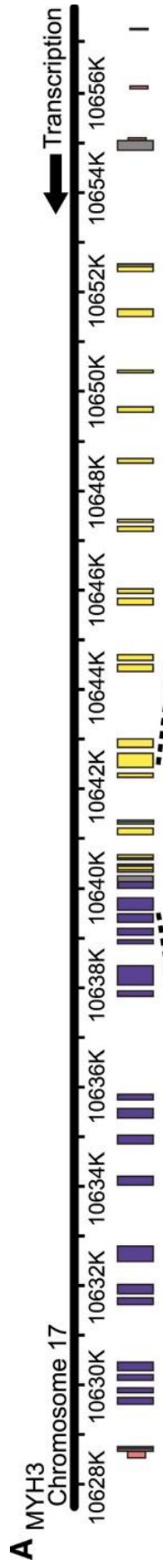
C DNA Sequence

MYH3	A	T	A	T	G	T	G	T	G	T	T	T	A	C
smyhc1+	G	T	C	C	G	C	G	C	G	T	T	T	C	A	C
smyhc1−	G	T	C	C	G	T	*	*	*	T	T	T	A	C
smyhc1 R673H	G	T	C	C	G	C	A	C	G	T	T	T	C	A	C

Polypeptide Sequence

MYH3	I	C	R	V	F	H
smyhc1+	L	C	R	V	F	H
smyhc1−	L	C	STOP	A	F	H
smyhc1 R673H	L	C	H	V	F	H

D

hpf   +/+   −/−
      24 48   24 48

F59 260kDa

GAPDH 37kDa

E   WT   KO   R673H/+   R673H/R673H
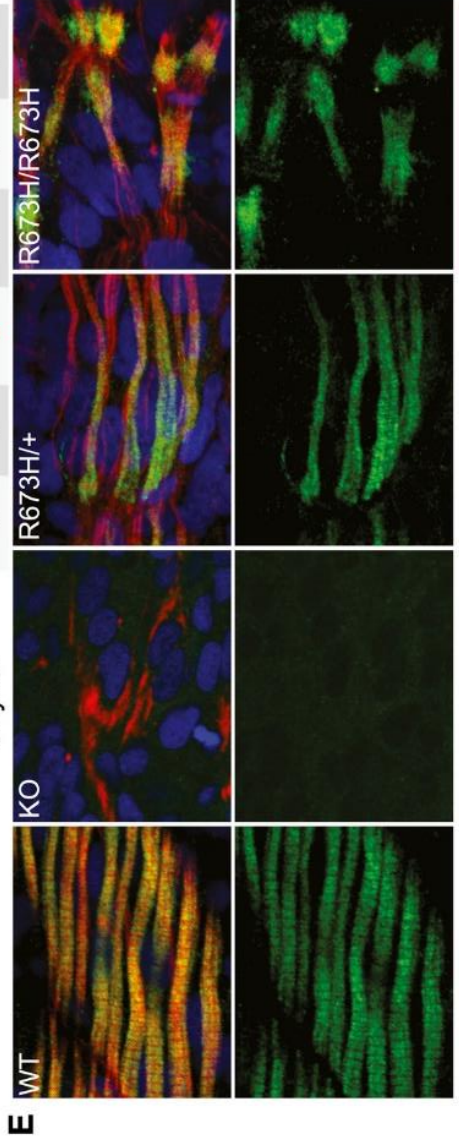
Figure 2.1 Generation of *smyhc1* mutant lines and embryonic muscle development. (A) Schematic to scale of human MYH3 gene on chromosome 17. Noncoding regions are displayed in pink. Coiled coil domain (840–1,933 bp) displayed in purple. Motor domain (86–779 bp) displayed in yellow. Actin binding site (656–678/758–772 bp) shown in cyan. Location of R672H mutation is enlarged and labeled. (B) Schematic to scale of zebrafish *smyhc1* gene on chromosome 24. Noncoding regions are displayed in pink. Coiled coil domain (842–1,929 bp) displayed in purple. Motor domain (85-778 bp) displayed in yellow. Actin binding site (655–677 bp) displayed in cyan. Location of R672H mutation is enlarged and labeled. (C) Aligned DNA and amino acid sequences of *MYH3* and *smyhc1* alleles surrounding the *smyhc1*[R673H] and *MYH3* R672H substitutions. The *smyhc1*[−] allele has a 7 base pair deletion that results in a frameshift in which one errant amino acid precedes a premature stop codon. The *smyhc1*[R673H] allele results from a G>A transition single point mutation. (D) Western blot of whole zebrafish larvae, Smyhc1 stained with the F59 antibody. *smyhc1*[+/+] larvae express Smyhc1 at 24 hpf, but not at 48 hpf. *smyhc1*[−/−] larvae do not express Smyhc1. (E) Smyhc1 immunohistochemistry stain of 24 hpf (hours post fertilization) zebrafish larvae. Filamentous actin is stained with phalloidin (red). Nuclei are stained with DAPI (blue). Smyhc1 is stained with the F59 antibody (green) (Elworthy et al, 2008). *smyhc1*[+/+], *smyhc1*[R673H/+], and *smyhc1*[R673H/R673H] larvae display Smyhc1 in the muscle fibers at 24 hpf. *smyhc1*[−/−] larvae at 24 do not stain positively for Smyhc1. Scale bar represents 50 μm.
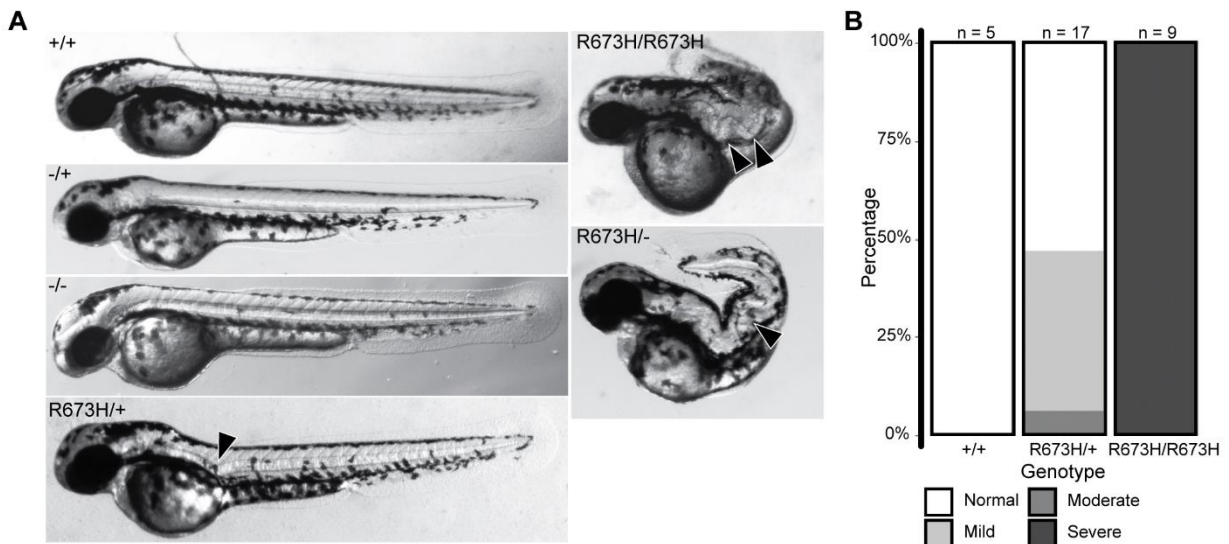
Figure 2.2 Embryonic phenotypes of *smyhc1* mutants. (A) Morphologies of *smyhc1*$^{+/+}$, *smyhc1*$^{-/+}$ and *smyhc1*$^{-/-}$ embryos are grossly normal at 2 days post fertilization (dpf), despite complete paralysis of smyhc1$^{-/-}$ embryos. In contrast, the *smyhc1*$^{R673H/+}$ embryos have notochord kinks in larval stages (arrow head), while *smyhc1*$^{R673H/R673H}$ embryos have much more severe notochord kinks that compress and distort the body axis (arrow heads). The *smyhc1*$^{R673H/-}$ embryos completely phenocopy *smyhc1*$^{R673H/R673H}$ embryos. (B) Quantification of *smyhc1*^{R673H} embryonic phenotypes at 2 dpf. Representative examples of fish in each of the four phenotypic groups (normal, mild, moderate, and severe) are shown in Supplemental Figure 2.4.
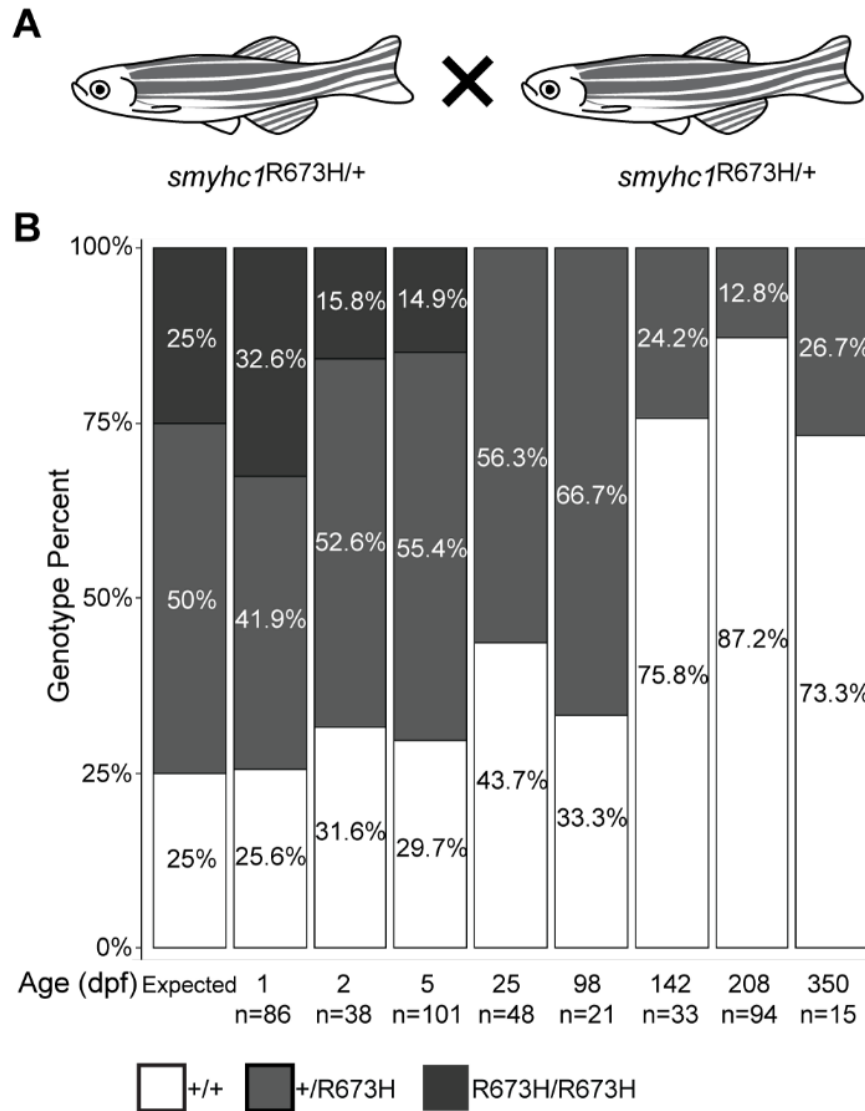
Figure 2.3 The *smyhc1^R673H^* allele reduces survival in a dose-dependent manner. (A) Schema of in-crosses used to generate clutches for survival assay. (B) The predicted mendelian (1:2:1) and experimental ratio of genotypes of progeny from *smyhc1^R673H/+^* in-crosses demonstrates survival at the indicated ages. This ratio quickly skewed toward an overrepresentation of *smyhc1^+/+^* fish with no representation of *smyhc1^R673H/R673H^* after 5 dpf. There is also dropout of *smyhc1^+/R673H^* fish into adulthood. Number of fish counted per group is indicated.
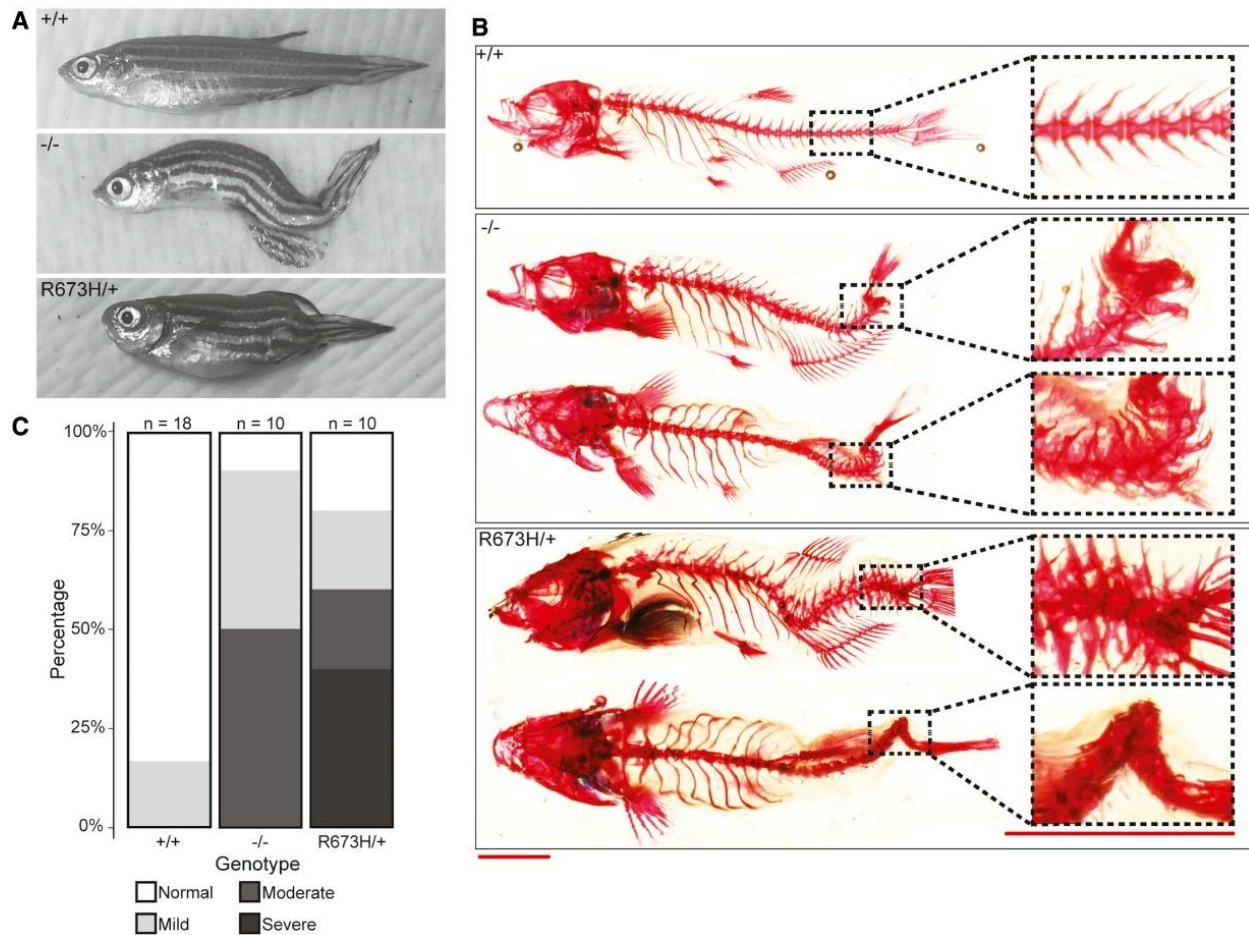
Figure 2.4 *smyhc1* mutants exhibit skeletal abnormalities in adulthood. (A) Gross morphology of *smyhc1* mutant adults. Most *smyhc1*$^{-/-}$ adult fish display dorsal tail curvature, while *smyhc1*$^{R673H/+}$ adults have a shortened body axis and variable spinal curves. (B) Alizarin red staining of bone shows spinal curvature but no bony fusions in the skeleton of *smyhc1*$^{-/-}$ adults, in contrast to the compression and fusion of vertebrae seen in *smyhc1*$^{R673H/+}$ adults. Distal tail regions are highlighted and enlarged. (C) Quantification of skeletal phenotype severity in adults. Representative examples of fish in each of the four phenotypic groups (normal, mild, moderate, and severe) are shown in Fig EV3. Scale bars represent a length of 5 mm.
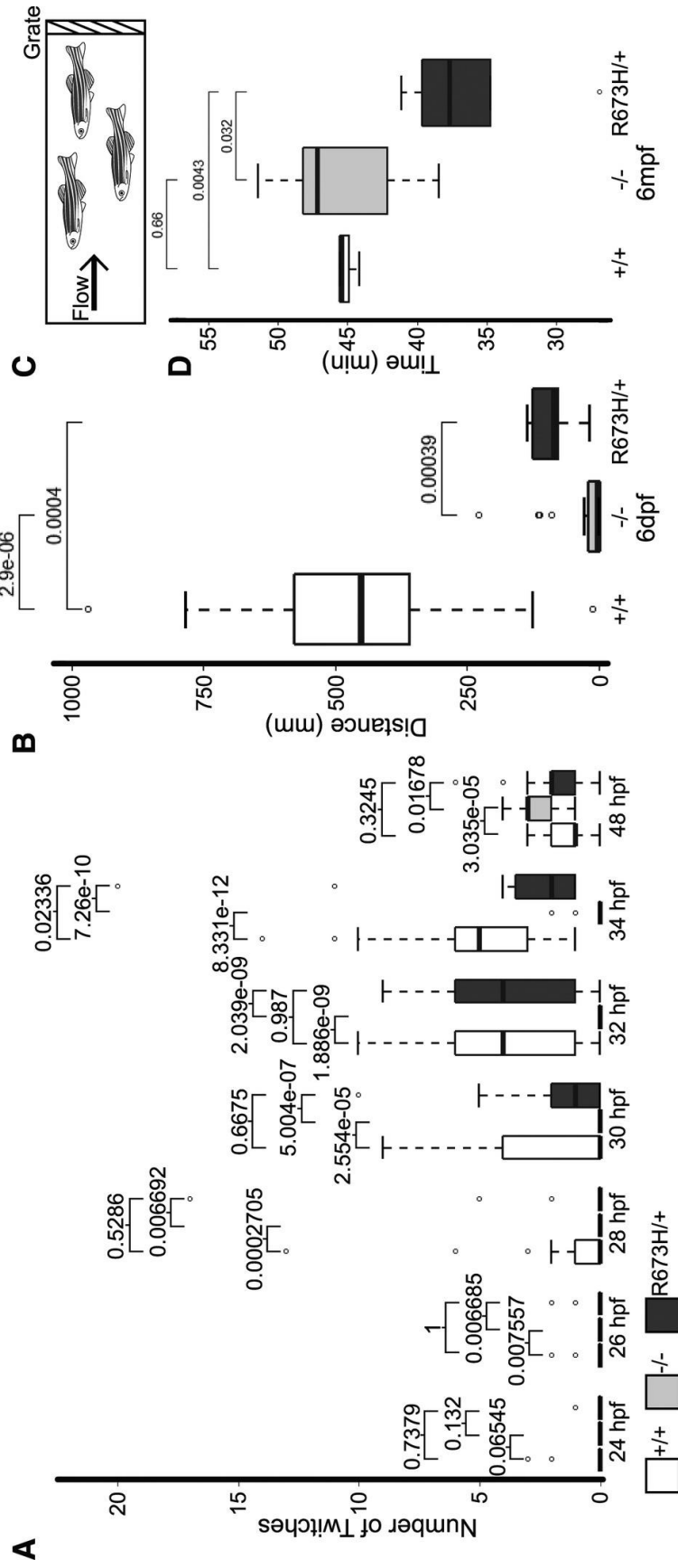
Figure 2.5 Motor deficits in *smyhc1* mutants are variable during development.

(A) Time course of light-triggered larval movements was manually counted at indicated times between 24 and 48 hpf. The *smyhc1*$^{-/-}$ larvae (n = 36) were paralyzed up to 48 hpf. *smyhc1*$^{R673H/+}$ mutants (n = 15) twitched slightly less than their *smyhc1*$^{+/+}$ siblings (n = 21) at 34 hpf. The central bands of the boxplots represent the median, the boxes of the boxplots represent the interquartile range (between the first and third quartile), and the whiskers represent the minimum and maximum values, up to 1.5 times the interquartile range. Outliers displayed are outside of this range. (B) Distance traveled during 5 min of spontaneous swimming at 6 dpf quantified by motion tracking software (Noldus Ethovision). At 6 dpf, both *smyhc1*$^{-/-}$ (n = 24) and *smyhc1*$^{R673H/+}$ (n = 11) larvae traveled significantly less than *smyhc1*$^{+/+}$ larvae (n = 13). The central bands of the boxplots represent the median, the boxes of the boxplots represent the interquartile range (between the first and third quartile), and the whiskers represent the minimum and maximum values, up to 1.5 times the interquartile range. Outliers displayed are outside of this range. (C) Diagram of swim tunnel used to quantify adult swimming behavior. Water flow was gradually increased until the fish fatigued and were collected at the grate. (D) Time spent swimming in swim tunnel before fatigue in adults at 6 months post fertilization (mpf). *smyhc1*$^{+/+}$ (n = 6), *smyhc1*$^{-/-}$ (n = 5), *smyhc1*$^{R673H/+}$ (n = 5).

Figure 2.5 Cont. Data Information: Wilcoxon rank-sum test used to calculate significance. p-values comparing each group displayed above data. The central bands of the boxplots represent the median, the boxes of the boxplots represent the interquartile range (between the first and third quartile), and the whiskers represent the minimum and maximum values, up to 1.5 times the interquartile range. Outliers displayed are outside of this range.
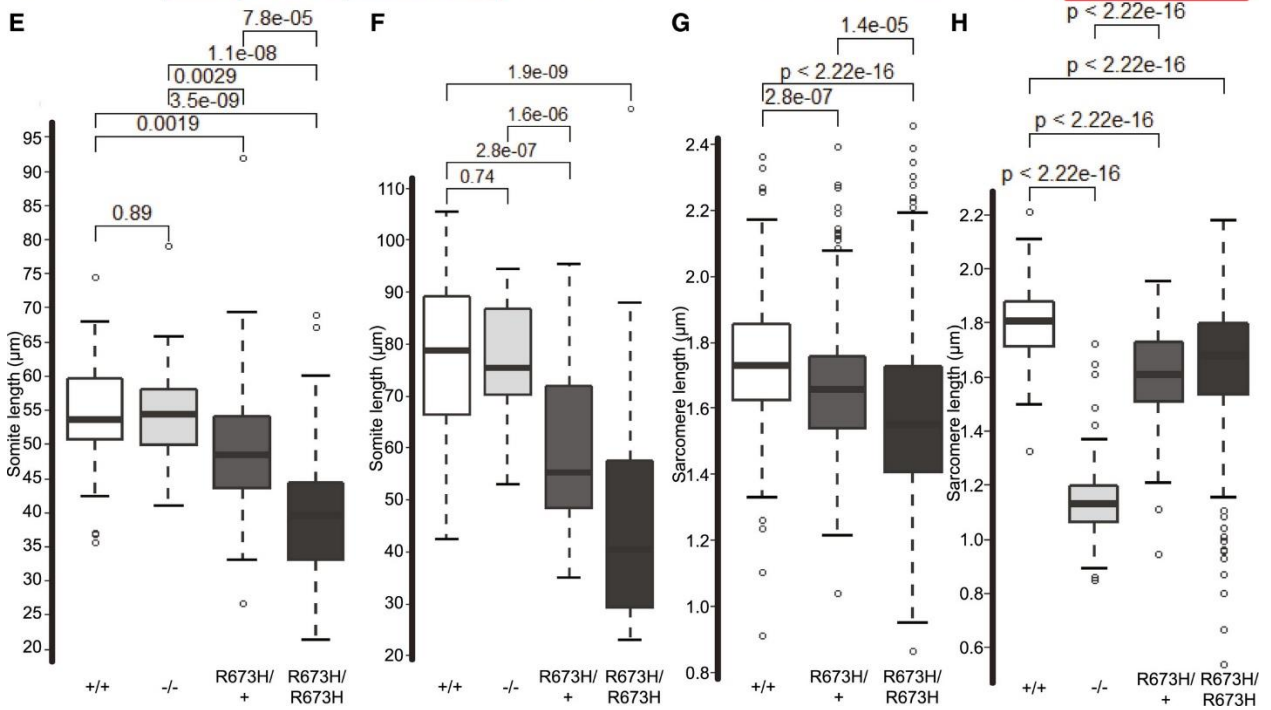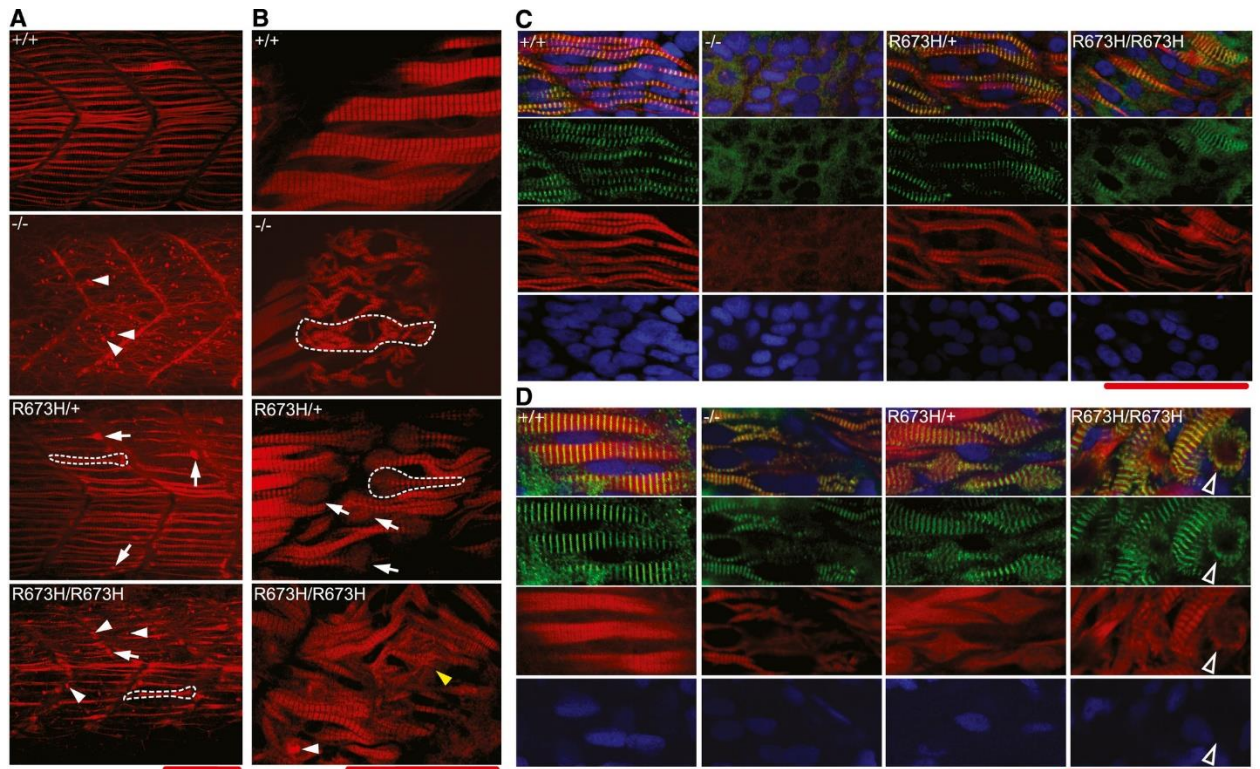
Figure 2.6 Slow skeletal muscle disorganization of *smyhc1* mutants. (A) Confocal fluorescence images of slow skeletal muscle of 1 dpf larvae, stained with phalloidin-rhodamine to detect actin. Puncta of filamentous actin are indicated with white arrowheads, and misshapen myofibers are indicated with white arrows and outlined. (B) Confocal microscope fluorescence images of slow skeletal muscle of 3 dpf larvae, stained with phalloidin-rhodamine. Puncta of filamentous actin are indicated with white arrowheads, distorted myofibers are indicated with arrows and outlined, and frayed myofibers are indicated with yellow arrowhead. (C)Confocal fluorescence images of slow skeletal muscle of 1 dpf larvae, stained with phalloidin-rhodamine (red), anti-α-actinin antibodies (green), and DAPI (blue). (D) Confocal fluorescence images of slow skeletal muscle of 3 dpf larvae, stained with phalloidin-rhodamine (red), anti-α-actinin antibodies (green), and DAPI (blue). Bundle of actin ringed with α-actinin indicated with white-bordered black arrowhead. (E) Myoseptal intervals of slow skeletal muscle at 1 dpf in *smyhc1*$^{+/+}$ ($n = 54$), *smyhc1*$^{-/-}$ ($n = 31$), *smyhc1*$^{R673H/+}$ ($n = 45$), and *smyhc1*$^{R673H/R673H}$ ($n = 49$). Distance between myosepta was measured perpendicular to rostral-caudal body axis at defined mid-body regions in phalloidin stained slow skeletal muscle fluorescence images. The central bands of the boxplots represent the median, the boxes of the boxplots represent the interquartile range (between the first and third quartile), and the whiskers represent the minimum and maximum values, up to 1.5 times the interquartile range. Outliers displayed are outside of this range.

Figure 2.6 Cont. (F) Myoseptal intervals of slow skeletal muscle at 3 dpf in *smyhc1*$^{+/+}$ ($n = 51$), *smyhc1*$^{-/-}$ ($n = 32$), *smyhc1*$^{R673H/+}$ ($n = 64$), and *smyhc1*$^{R673H/R673H}$ ($n = 38$). The central bands of the boxplots represent the median, the boxes of the boxplots represent the interquartile range (between the first and third quartile), and the whiskers represent the minimum and maximum values, up to 1.5 times the interquartile range. Outliers displayed are outside of this range. (G) Z-disk intervals (sarcomere length) of slow skeletal muscle at 1 dpf in *smyhc1*$^{+/+}$ ($n = 380$), *smyhc1*$^{R673H/+}$ ($n = 293$), and *smyhc1*$^{R673H/R673H}$ ($n = 268$). Distance between z-disks was measured in anti-α-actinin stained slow skeletal muscle fluorescence images. The central bands of the boxplots represent the median, the boxes of the boxplots represent the interquartile range (between the first and third quartile), and the whiskers represent the minimum and maximum values, up to 1.5 times the interquartile range. Outliers displayed are outside of this range. (H) Z-disk intervals (sarcomere length) of slow skeletal muscle at 3 dpf in *smyhc1*$^{+/+}$ ($n = 249$), *smyhc1*$^{-/-}$ ($n = 278$), *smyhc1*$^{R673H/+}$ ($n = 246$), and *smyhc1*$^{R673H/R673H}$ ($n = 252$). The central bands of the boxplots represent the median, the boxes of the boxplots represent the interquartile range (between the first and third quartile), and the whiskers represent the minimum and maximum values, up to 1.5 times the interquartile range. Outliers displayed are outside of this range.

Data Information: Wilcoxon rank-sum test used to calculate significance. *P*-values comparing each group displayed above data. Scale bars represent a length of 50 μm.
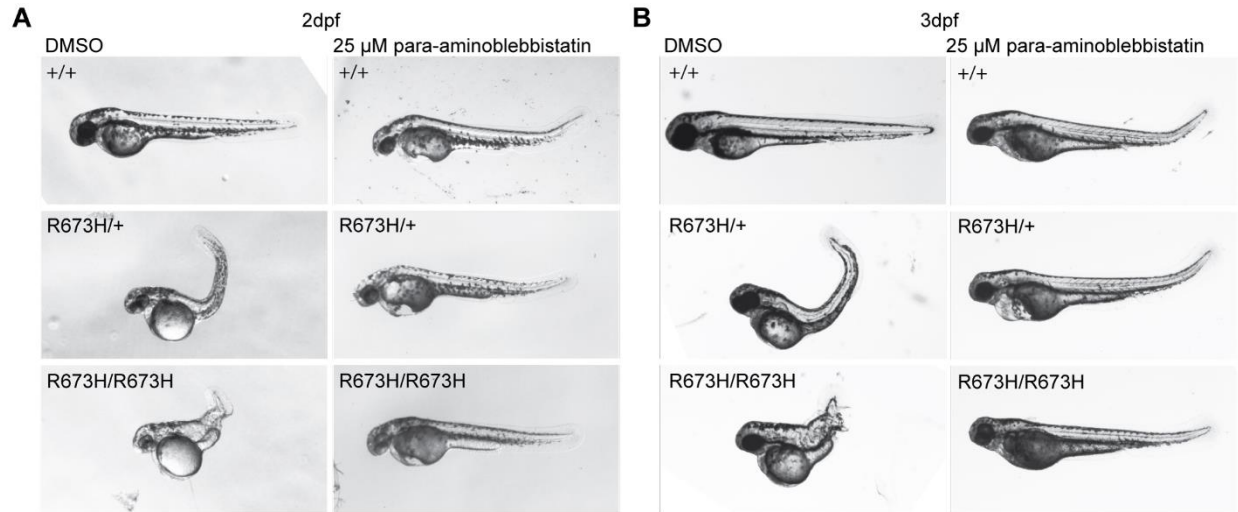
Figure 2.7 Para-aminoblebbistatin ameliorates the *smyhc1*[R673H] larval phenotype. (A) Representative morphologies of *smyhc1*[+/+], *smyhc1*[R673H/+], and *smyhc1*[R673H/R673H] larvae at 2 dpf after 24 hour treatment with 0.25% DMSO (control) or 25 µM para-aminoblebbistatin. All larvae treated with para-aminoblebbistatin develop a slight dorsal tail curve and pericardial edema regardless of genotype. All para-aminoblebbistatin-treated *smyhc1*[R673H/+] (n=9) and *smyhc1*[R673H/R673H] (n=3) larvae are indistinguishable from *smyhc1*[+/+] unlike the untreated larvae which have markedly curved or distorted body axis. (B) Representative images of *smyhc1*[+/+], *smyhc1*[R673H/+], and *smyhc1*[R673H/R673H] larvae at 3dpf after 48 hour treatment with 0.25% DMSO (control) or 25 µM para-aminoblebbistatin. All larvae treated with para-aminoblebbistatin develop a slight dorsal tail curve and pericardial edema regardless of genotype. Para-aminoblebbistatin-treated *smyhc1*[R673H/+] (n=9) and *smyhc1*[R673H/R673H] (n=3) larvae are indistinguishable from *smyhc1*[+/+] unlike the untreated larvae which have markedly curved or distorted body axis.

2.4 Discussion

Understanding the functional consequences of *MYH3* variants associated with human disease has been limited by the lack of access to human skeletal muscle tissue during embryonic development. The recent description of multiple human skeletal muscle and bone phenotypes associated with *MYH3* variants, including autosomal dominant and recessive spondylocarpotarsal synostosis syndromes, in addition to DA, provides a strong motivation to develop animal models to study and understand these variable phenotypes. By generating two viable zebrafish *smyhc1* alleles, including a *smyhc1* null allele and a *smyhc1^{R673H}* missense variant, we now have tools to assess inheritance models and effects of gene dosage on phenotypic expression.

Our results not only confirm that the *smyhc1^{R673H}* allele is more severe than a *smyhc1* null allele (which was also recently described in (Li et al, 2020)), but also demonstrate the dosage sensitivity of the missense allele to the wild-type allele, which modulates the effects of the disease-causing R673H variant in our model system. Expression of the *smyhc1^{R673H}* allele in the absence of a wild-type allele resulted in embryos with a lethal phenotype identical to *smyhc1$^{R673H/R673H}$* embryos. Myosin is comprised of a hexamer containing a pair of heavy chains, a pair of regulatory light chains, and a pair of essential light chains (Schiaffino & Reggiani, 1996). Because of the dimerization of the myosin heavy chains, and subsequent assembly of the hexamers into the thick filament, it is reasonable to hypothesize that compound heterozygosity of a missense mutation in trans with a hypomorphic or null allele will increase the phenotypic severity of afflicted individuals compared to a missense mutation in trans with a wild-type allele. In fact, a Dutch study recently identified a high impact *MYH3* 5′ UTR variant with a gnomAD minor allele frequency of 0.008 that diminishes *MYH3* translational efficiency (Cameron-Christie et al, 2018). This variant was reported in patients with presumed autosomal recessive spondylocarpotarsal synostosis syndrome,

demonstrating a biallelic phenotype consisting of a haploinsufficient allele and a hypomorphic allele. This genotype exacerbated the severity of the disorder, and our results in zebrafish similarly demonstrate that the phenotypic severity of an autosomal dominant allele is greater when in trans with a hypomorphic than a wild-type allele. Because intrafamilial phenotypic variability has previously been reported in DA1 families with *MYH3* mutations (Alvarado et al, 2011), we now hypothesize that expression differences of the wild-type allele may be responsible.

Zebrafish offer many practical advantages for research, including accessibility at early time points in development due to their optical clarity, speed of development, large clutch size, and convenient external development. *smyhc1* is the only myosin gene expressed during early zebrafish embryogenesis; therefore, we hypothesized that its manipulation would accurately model the human condition (Devoto et al, 1996; Elworthy et al, 2008; Schiaffino et al, 2015). Zebrafish also offer a unique opportunity to study slow skeletal muscle development in isolation due to the spatial segregation of slow and fast skeletal muscle in zebrafish, with the former developing as a thin border along the lateral aspect and the latter making up the bulk of the trunk (Elworthy et al, 2008). In contrast, human skeletal muscle is composed of interspersed slow and fast twitch muscle fibers. Muscle hypercontractility in $smyhc1^{R673H}$ heterozygous and homozygous fish was evidenced by the shortened slow muscle sarcomere length, and shortened myoseptal interval that phenocopies the *accordion* hypercontractile zebrafish mutant which has a muscle relaxation defect due to a mutation in *atp2a1* (sarcoplasmic reticulum Ca^{2+}-ATPase 1)(Hirata et al, 2004). In our model, some myofibers even appeared to physically dissociate from the myosepta by flaring bulbously at the end of the myofiber, which we hypothesize was a consequence of excessive biomechanical force and may have contributed to increased apoptosis in the homozygous $smyhc1^{R673H}$ larvae. This resembles the hypercontractile phenotype seen

in *Drosophila* upon expression of the R672C mutation, in which normal myofibril assembly is followed by degeneration that ultimately results in a complete impairment of flight (Rao et al, 2019).

One of the earliest morphological abnormalities that we observed in *smyhc1^{R673H}* heterozygous and homozygous mutant embryos was prominent notochord kinks and bends. Results from our study suggest that notochord abnormalities arise as a consequence of excessive tension on the notochord due to hypercontraction and/or uncoordinated contraction of the surrounding muscle. Data to support this mechanism include the much more severe phenotype seen in homozygous *smyhc1^{R673H}* mutants, and the complete rescue of this phenotype by treatment with the myosin ATPase inhibitor, para-amino blebbistatin. While the notochord abnormalities in the *accordion* mutant were suppressed with drugs inhibiting neural transmission (Hirata et al, 2004), the *smyhc1^{R673H}* mutant notochord defects could not be similarly suppressed, but improved only when muscle contraction itself was directly impaired via myosin-inhibiting drugs. Our *in vivo* results are consistent with prior *in vitro* studies of both patient muscle tissue and C2C12 cells in which the R673H variant was observed to negatively affect relaxation kinetics (Racca et al, 2015; Walklate et al, 2016). As predicted by their work, gross phenotypic abnormalities in our *smyhc1*^{R673H} mutants, including severe notochord kinks and bends, were completely abolished after treatment with myosin inhibitors.

The phenotypic spectrum of *MYH3* associated disease was recently broadened to include vertebral fusions, which are a defining clinical feature of spondylocarpotarsal synostosis syndrome. Spondylocarpotarsal synostosis syndrome is associated with both autosomal dominant and recessive inheritance; both missense and nonsense *MYH3* variants have been reported in recessive cases (Cameron-Christie et al, 2018; Carapito et al, 2016; Zieba et al, 2017). Congenital

spinal fusions were present in our *smyhc1^{R673H}* mutants but not in *smyhc1* null mutant fish, suggesting that bony fusions in our missense model result secondarily from increased tension generated by the mutant muscle protein, putting excessive tension and stress on the notochord, causing it to kink. Gray et al. previously observed a direct spatial relationship between the areas where notochord kinks were most severe and the location of later developing vertebral fusions in *leviathan*/*col8a1a* null zebrafish (Gray et al, 2014). Thus, our current model supports a hypermorphic or neomorphic effect of the *smyhc1^{R673H}* allele that causes increased muscle tension, which in turn puts excessive stress on the notochord resulting in kinks that progress into vertebral fusions in adult fish (Figure 2.8). Others have proposed alternative mechanisms for the development of skeletal abnormalities in *MYH3* associated disorders, including a direct effect of its postnatal expression in vertebral bone (Zieba et al, 2017). Furthermore, the postnatal persistence of *MYH3* expression in the small multifidus muscles connecting to the neural arches of the spine in mice was also proposed as a possible mechanism. It is possible that abnormal mechanical forces imposed upon the intervertebral disc by pathogenic *MYH3* mutations functioning postnatally in these small muscle increase the vulnerability of the spine to scoliosis and spinal fusions over time (Zieba et al, 2017). Furthermore, in cultured cells, *MYH3* expression also activates TGF-β signaling pathways that are associated with bone fusions in other disease models (Zieba et al, 2017). Interestingly, we also note that *smyhc1* is expressed in zebrafish somites at the 5-9 somite stage even prior to muscle development (Bessarab et al, 2008; Li et al, 2020; Rauch et al, 2003); its role at this early time period is unknown but may indicate a function independent from its role in muscle. More work is needed to understand the how *MYH3* mediates the complex developmental relationship between muscle and bone.

Animal models are necessary to advance the development of novel gene-directed therapies for patients with DA. Pharmacological approaches for DA should be informed by research in hypertrophic cardiomyopathy on the closely related cardiac myosin heavy chain gene, *MYH7* (Green et al, 2016). The R453C variant in *MYH7* is analogous to the most common MYH3 mutation in DA2A that we modeled here in the zebrafish gene, *smyhc1*. Prior work has modeled the *MYH7* R453C variant in mice, demonstrating that treatment with cardiac specific myosin inhibitors, including the small molecule MYK-461, reduces ventricular hypertrophy, cardiomyocyte disarray, and profibrotic gene expression that are hallmarks of human hypertrophic cardiomyopathy (Green et al, 2016). While inhibitors specific for embryonic skeletal muscle myosin have not yet been reported, they would be ideal drugs for the treatment and prevention of diseases associated with *MYH3* hypermorphic mutations. If *MYH3*-associated muscle contractures or bone fusions result entirely from early embryonic gene expression, then short drug treatments may be sufficient to prevent progression of pathogenic phenotypes. Although our data show that zebrafish with complete loss of *smyhc1* expression have a milder phenotype than those with *smyhc1*^{R673H} mutations, nonspecific inhibition of *smyhc1* or *MYH3* with antisense oligonucleotides risks worsening the phenotype of patients with missense mutations, as seen by the severe phenotype of *smyhc1*$^{R673H/-}$ mutants. Therefore, any proposed antisense oligonucleotide therapy would need to be designed specifically for each *MYH3* pathogenic variant. Our work supports the potential use of myosin ATPase inhibitors for the treatment of *MYH3*-associated DA, but considerable challenges remain, including elucidation of the optimal treatment period and reduction of toxicity through the development of embryonic myosin specific inhibitors.

2.5 Synopsis

2.5.1 Problem

The debilitating joint contractures that are characteristic of distal arthrogryposis (DA) syndromes are currently suboptimally treated with supportive care. Understanding how dominant or recessive *MYH3* mutations cause contractures or bony fusions has been limited by poor access to human tissue, particularly during early development when *MYH3* is most highly expressed. Previous single-cell and small molecule studies suggest that DA mutations cause muscle hypercontraction, but vertebrate models are required to study the complex interactions between bone and muscle and to develop novel targeted therapeutics.

2.5.2 Results

Zebrafish carrying a single copy of the most common DA-associated *MYH3* substitution (R672H) displayed notochord bends that developed into scoliosis and vertebral fusions in adulthood, as well as shortened muscle fibers and impaired swimming capacity. The direct chemical inhibition of muscle contraction with the myosin ATPase inhibitor para-aminoblebbistatin prevented the notochord bends from developing in both heterozygous and homozygous fish, suggesting that the mutant allele causes notochord and vertebral abnormalities through a mechanical increase in muscle tension.

2.5.3 Impact

We developed a viable zebrafish model of DA that is dually useful for both mechanistic studies and therapeutic drug development. Our demonstration that the muscle hypercontractility mediated by the MYH3 mutation secondarily leads to vertebral fusions highlights the interconnectedness of the muscular and skeletal systems during early development. Furthermore,

we have shown the beneficial effects of myosin ATPase inhibitors for the treatment of DA (Figure 2.8).

smyhc1^{+/+}

smyhc1^{R673H/+}

Z disc
Actin Thin Filament
Myosin Thick Filament
Muscle
Muscle
Notochord
Vertebrae
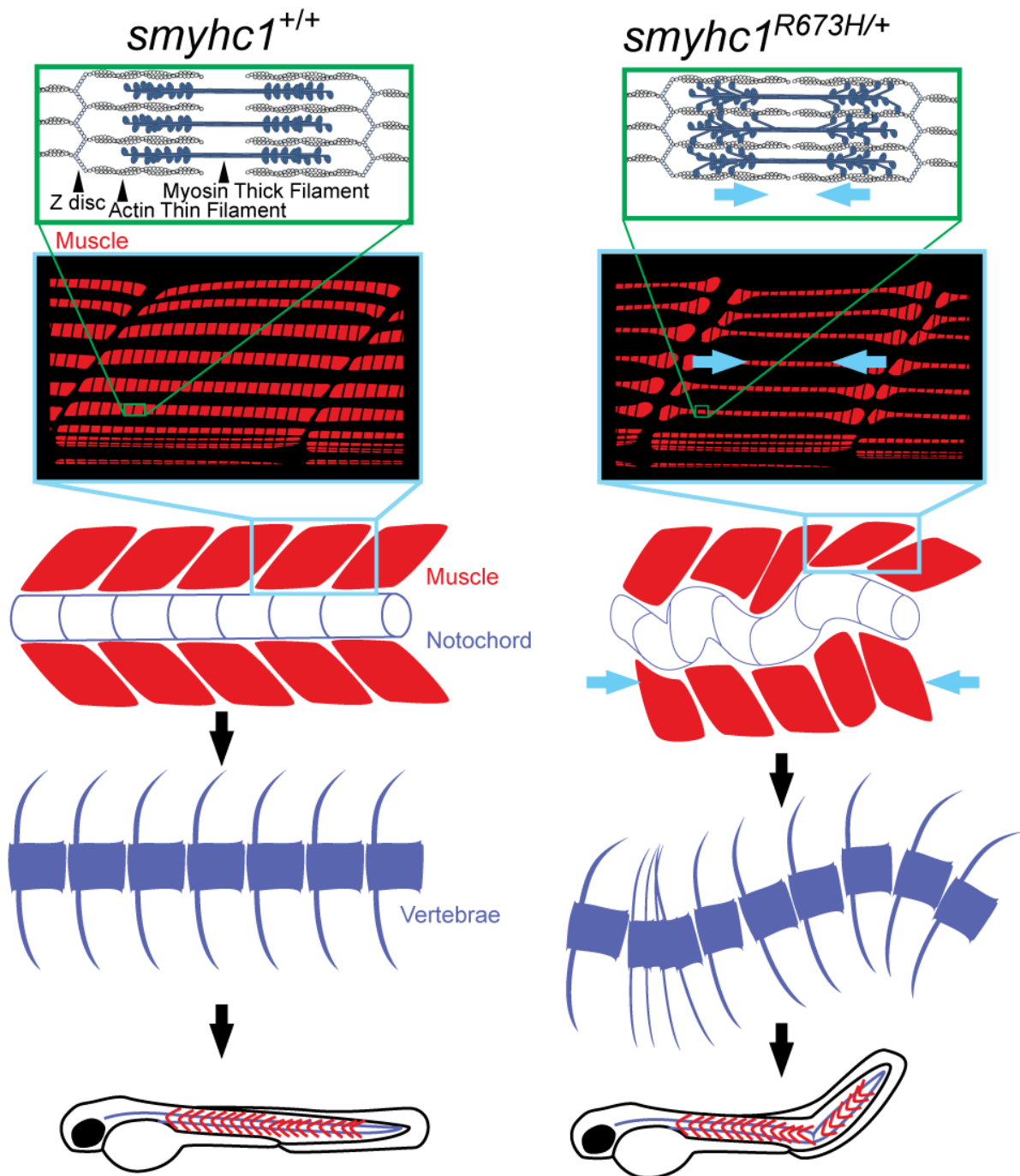
Figure 2.8 Synopsis Schematic. Introduction of the most common distal arthrogryposis (DA) associated pathogenic missense variant (*MYH3^{R672H}*) into the analogous zebrafish gene (*smyhc1^{R673H}*) largely reproduced the human phenotype. This model of provides key mechanistic insight into the pathogenesis of developmental disorders of muscle and bone as well as exploration of novel therapeutics. Here, we show Smyhc1^{R673H} causes muscle hypercontraction, which deforms muscle (shortening muscle fiber length). Tension caused by the hypercontracted muscle bends the notochord early in development, which develop into vertebral fusions and scoliosis. By reducing the hypercontracted state, myosin ATPase inhibitors normalize the notochord phenotype and show promise for human therapeutics.

2.6 Supplemental figures
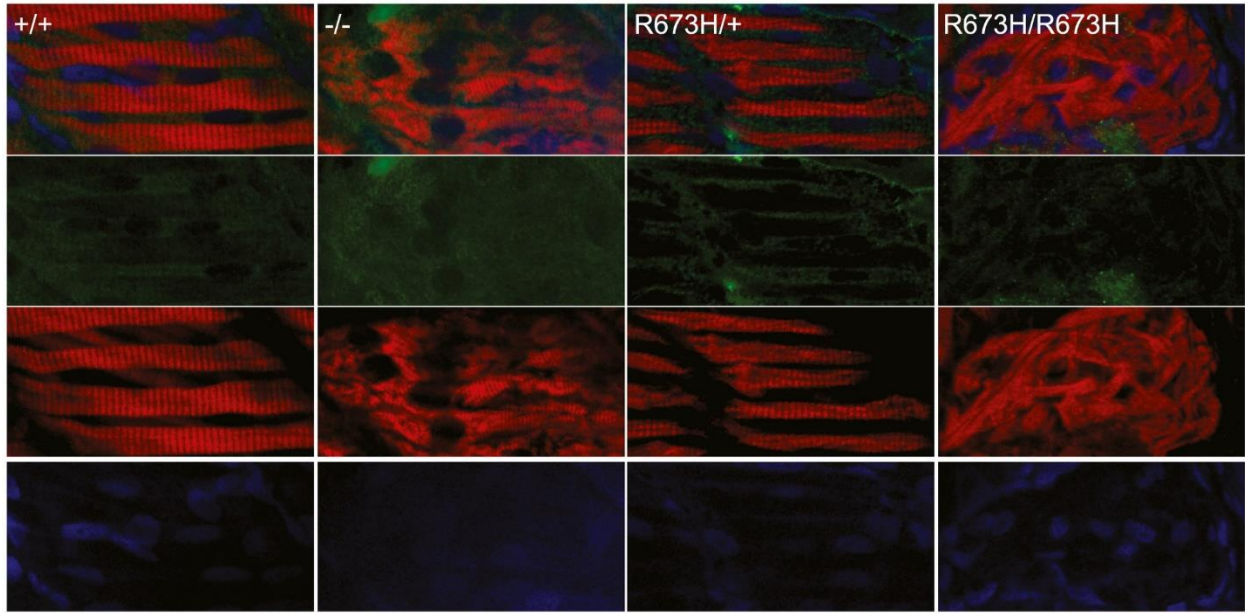
Supplemental Figure 2.1. Transient expression of *smyhc1.* Smyhc1
immunohistochemistry stain of 3 dpf zebrafish larvae. Filamentous actin is
stained with phalloidin (red). Nuclei are stained with DAPI (blue). Smyhc1 is
stained with the F59 antibody (green)
(Elworthy et al, 2008). *smyhc1*$^{+/+}$, *smyhc1*$^{R673H/+}$, and *smyhc1*$^{R673H/R673H}$ larvae
do not display Smyhc1 in the muscle fibers at 3dpf. Data Information: Scale
bar represents a length of 50 µm.

Supplemental Figure 2.2. Examples of adult phenotypes resulting from various *smyhc1* **mutant genotypes.**
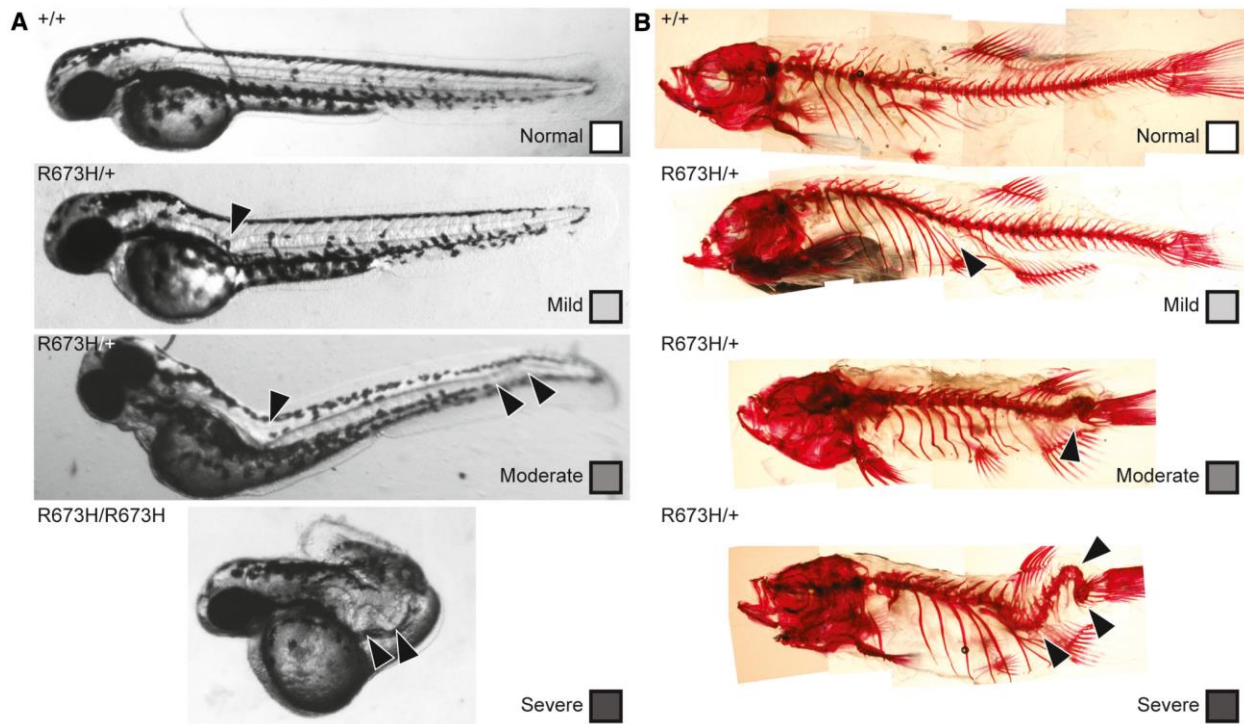
Supplemental Figure 2.3 Representative examples of each embryonic and adult graded phenotype. (A) Grading scale for 2 dpf embryos. Normal is classified as 0 visible abnormalities, mild is classified as 1 notochord curve or kink, moderate is classified as multiple notochord curves or kinks, and severe is classified as multiple notochord kinks that severely distort the body. Arrowheads indicate notochord abnormalities. (B) Grading scale for adult skeletal abnormalities. Normal is classified as 0 visible abnormalities, mild is classified as 1 spinal curve or kink, moderate is classified as multiple spinal curves or kinks, or apparent vertebral fusion, and severe is classified as multiple spinal kinks or vertebral fusions that severely distort the body. Arrowheads indicate spinal abnormalities.
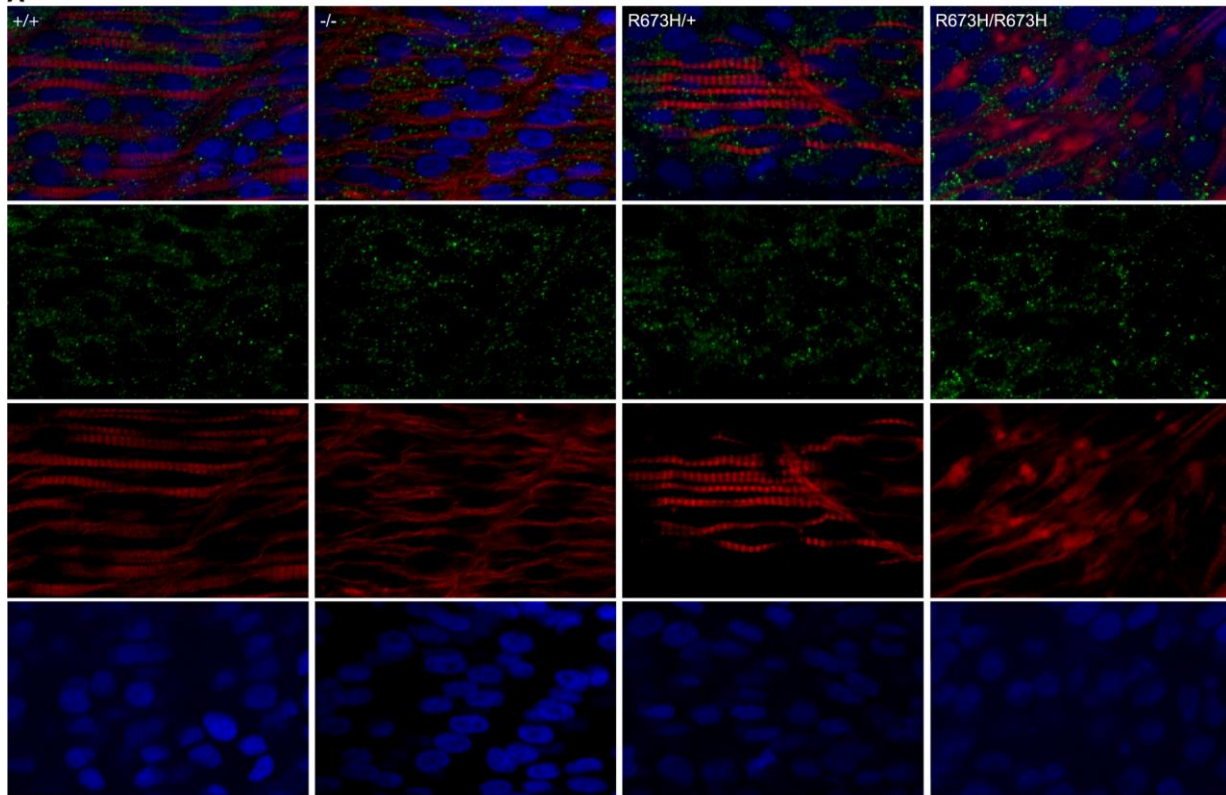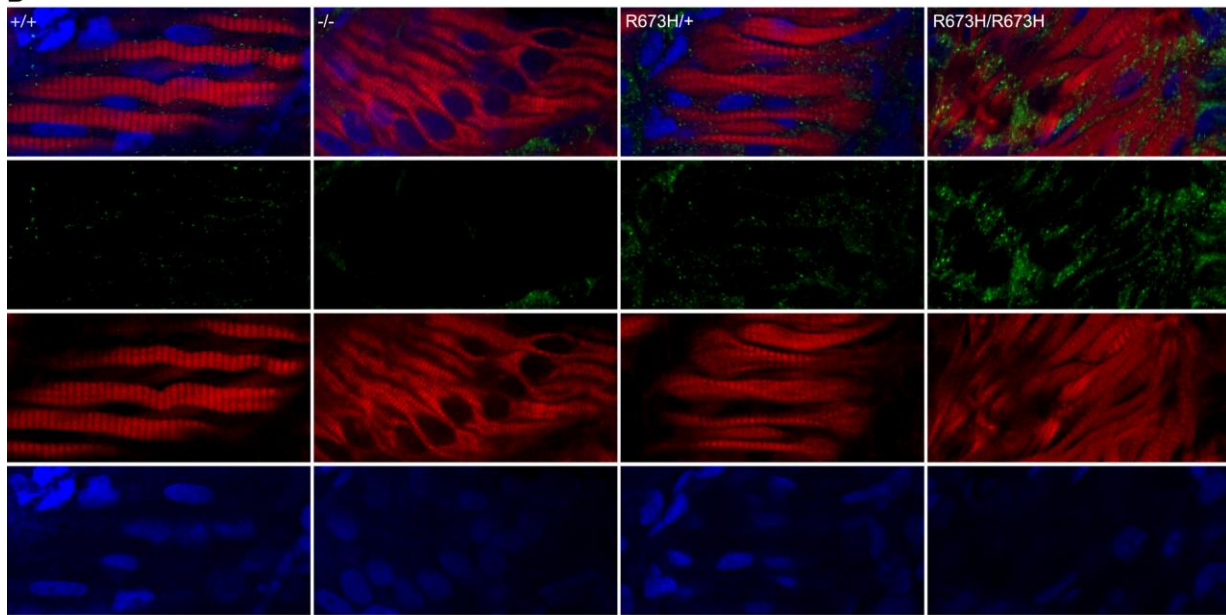
Supplemental Figure 2.4 Activated Caspase-3 presence in *smyhc1* mutant tissue. (A) Confocal fluorescence images of slow skeletal muscle of 1 dpf larvae, stained with phalloidin-rhodamine (red), Caspase-3 antibodies (green), and DAPI (blue). (B) Confocal fluorescence images of slow skeletal muscle of 3 dpf larvae, stained with phalloidin-rhodamine (red), Caspase-3 antibodies (green), and DAPI (blue).

A

Control

+/+
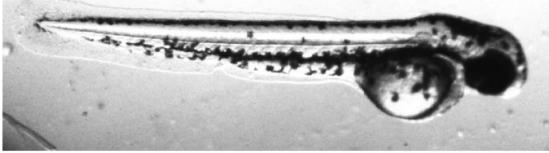
R673H/+

R673H/R673H

40 µg/mL Tricaine

+/+
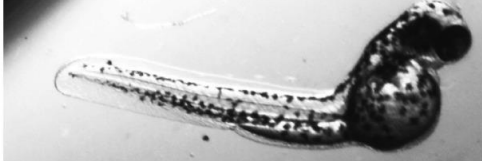
R673H/+

R673H/R673H

B

DMSO

+/+
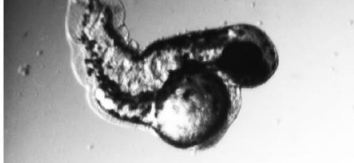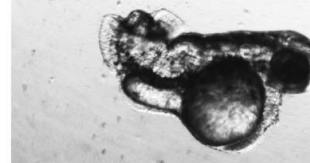
R673H/+

R673H/R673H

25 µM Blebbistatin

+/+

R673H/+
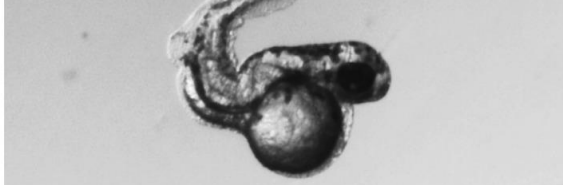
R673H/R673H

Supplemental Figure 2.5 Tricaine fails to ameliorate the *smyhc1*^{R673H} larval phenotype, and blebbistatin improves the *smyhc1*^{R673H} larval phenotype. (A) Zebrafish embryos (2 dpf) treated with 40 mg/ml tricaine methanylsulfonate or control egg medium for 24 h starting at 1 dpf. Tricaine suppressed movement but failed to normalize notochord kinks and bends. Tricaine appears to have no major effect on anatomy. (B) Zebrafish embryos (2 dpf) were treated with 25 µM blebbistatin or DMSO control for 24 h starting at 1 dpf. Blebbistatin appears to partially rescue *smyhc1*^{R673H} phenotype, but causes known off target effects. Note the severe dorsal tail curve and pericardial edema present in all larvae treated with blebbistatin regardless of genotype. Representative images of embryos at 48 hpf.

2.7 Materials and methods

2.7.1 Ethical statement

Husbandry and protocols for experiments involving zebrafish (*Danio rerio*) have been reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at Washington University.

2.7.2 Zebrafish husbandry

Zebrafish (*Danio rerio*) were raised and maintained using standard methods (Westerfield, 1995). Zebrafish were housed and handled using protocols approved by the Institutional Animal Care and Use Committee (IACUC) at Washington University.

2.7.3 TALEN (Transcription activator-like effector nuclease) generation of *smyhc1* mutants

Left and right TALENs were designed to target the sites flanking codon 673 of *smyhc1*, on exon 16. TALENs were generated using the Golden Gate method as previously described (Cermak et al, 2011). Plasmids were acquired from the Golden Gate TALEN and TAL Effector Kit (AddGene) and RVD repeat arrays were cloned into pCS2TAL3DD and pCS2TAL3RR (AddGene). Plasmids were transformed into DH5α competent cells (Invitrogen) and isolated using the QIAprep Spin Miniprep Kit (Qiagen).

Subsequently, plasmids encoding the scaffold and RVD arrays were linearized via restriction digest and 5′-capped mRNA was generated by in vitro transcription using the mMESSAGE mMACHINE SP6 Transcription Kit (Life Technologies). Capped mRNA was purified using the RNeasy Mini Kit (Qiagen). WT embryos were collected, and 40–50 pg of pooled left and right TALENS at equal concentrations were injected into the yolk of 1-4 cell stage embryos. Embryos were raised to adulthood, and sperm and progeny of these fish were collected and sequenced with MiSeq to screen for germline transmission.

2.7.4 Western blot of Smyhc1 protein

For each sample, 20 embryos were collected and dechorionated in a 37°C 1 mg/mL pronase egg water solution. Embryos were deyolked by agitating in deyolking buffer (55mM NaCl, 1.25 mM NaHCO$_3$, 1.8 mM KCl, 2.7 mM CaCl$_2$). Embryos were then washed twice in wash buffer (10 mM Tris pH 8.5, 110 mM NaCl, 3.5 mM KCl, 2.7 mMCaCl$_2$). Embryos were then agitated in RIPA with complete protease inhibitor cocktail (Roche Applied Sciences). 50 µg of each sample was loaded onto a sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). Proteins were transferred to PVDF transfer membrane (Immobilon) for immunoblotting. Membranes were blocked in 5% skim milk for 1 hour and incubated with antibody in 1% skim milk overnight at 4°C. The embryonic myosin heavy chain antibody (F59)(Elworthy et al, 2008) (Developmental Studies Hybridoma Bank) was used to stain Smyhc1. A GAPDH antibody (Thermo Scientific) was used as a loading control. Membranes were washed in 0.1% PBST (0.1% Tween-20 in PBS) and incubated with a secondary anti-mouse antibody conjugated to horseradish peroxidase and developed with a chemiluminescent substrate enhanced chemiluminescence (GE Healthcare).

2.7.5 Genotyping of *smyhc1^{R673H}* mutants

Individual fish were genotyped via PCR amplification of the *smyhc1* mutant region, with subsequent restriction digest targeting to identify the R673H allele. DNA was extracted from tissue samples in lysis buffer (0.1M Tris, 0.005M EDTA, 0.03% SDS, 0.2 M NaCl, 1% Proteinase K [New England Biolabs, P8107S]), incubated at 50°C for 16 hours, and 95°C for 10 minutes. A fragment of DNA containing the R673H mutation was PCR amplified. The amplified DNA was incubated with the restriction enzyme ApaLI in CutSmart buffer (New England BioLabs; R0507S, B7204S) for 16 hours at 37°C, which cleaves *smyhc1* R673H mutant DNA fragments, allowing

wild-type, homozygous, and heterozygous individuals to be identified after separation on an agarose gel.

2.7.6 Alizarin red/Alcian blue skeletal staining

Adult zebrafish (1–2 years postfertilization, male and female) were euthanized and fixed in 4% paraformaldehyde in PBS at room temperature under agitation for 7 days. They were then incubated in acetone overnight under agitation at room temperature. After rinsing in water, they were then incubated overnight under agitation in staining solution (0.015% Alcian Blue, 0.005% Alizarin Red, 5% Glacial Acetic Acid, 59.5% Ethanol). The fish were rinsed once and then for 30 min under agitation at room temperature in water. The fish were incubated in 1% KOH under agitation at room temperature until the soft tissue became transparent, 1–3 weeks depending on the size of the fish. The 1% KOH solution was replaced periodically. The fish were stored in glycerol, and the scales removed manually if they had not fallen off during KOH clearing.

2.7.7 Phalloidin-rhodamine/DAPI/Immunohistochemistry muscle staining and imaging

Zebrafish larvae 1–3 dpf were euthanized and fixed in 4% paraformaldehyde in PBS overnight under agitation at 4°C. The larvae were washed briefly and then twice for 30 min in PBD (0.1% Tween-20, 0.3% Triton X-100, 1% DMSO in PBS) under agitation at room temperature. The embryos were then incubated in block solution (0.1% Tween-20, 10% FBS, 2% BSA) at room temperature under agitation for 60 min. Primary antibodies were added to the block solution at the appropriate dilutions [F59 (Iowa Developmental Studies Hybridoma bank) 1:10; anti-Caspase 3 (Sigma-Aldrich AB3623) 1:100; anti-α-actinin (Sigma-Aldrich A7811) 1:800; anti-phospho-SMAD2 (Sigma-Aldrich AB3849) 1:100] and incubated overnight under agitation for 4°C. The solution was removed, and the embryos were washed once quickly in PBD and twice for 30 min under agitation at room temperature. The embryos were then incubated in block solution

for 60 min under agitation at room temperature. The appropriate secondary antibody was added to the block solution [Goat Anti-Rabbit IgG H&L (Alexa Fluor® 488) (Abcam ab150077) 1:500, Goat Anti-Mouse IgG (H&L) (Adsorbed Against: Hu., Bv., Hs.) (DyLight® 488) (Leinco M1331) 1:500], and the embryos were incubated at 4°C overnight under agitation in the dark. The solution was removed, and the embryos were washed once quickly in PBD and incubated with DAPI (Sigma-Aldrich D9542) at a dilution of 1:15,000 in PBST for 30 min under agitation at room temperature. The solution was removed, and the embryos were incubated in rhodamine phalloidin (Invitrogen 2009720) at a dilution of 1:1,000 in PBD for 30 min at room temperature. The embryos were then washed in PDB for 30 min at room temperature and mounted on slides in glycerol under cover slips before imaging under confocal microscopy. Confocal images of stained muscle were analyzed in ImageJ to measure myosepta distance.

2.7.8 Microscopy

Confocal images were taken on an Olympus BX61WI fixed stage microscope with Olympus Fluoview FV1000 confocal laser scanning. Images were processed in ImageJ.

2.7.9 Drug treatment

Embryos were dechorionated prior to drug treatment. Embryos were immersed in 40 ug/mL tricaine methanesulfonate in egg water (5 mM NaCl, 0.17 mM KCl, 0.33 mm $CaCl_2$, $MgSO4$ in H_2O), 25 uM blebbistatin, or 25 uM para-aminoblebbistatin. Para-aminoblebbistatin in DMSO was mixed with egg water to facilitate dissolution of blebbistatin in aqueous buffer, generating 0.25% DMSO, 25 uM blebbistatin in egg water. For the tricaine treatment, the egg water was replaced every 24 hours for 5 days. All incubations were performed in glass-bottom dishes. Images of the embryos were taken periodically.

2.7.10 Noldus movement quantification

The Noldus DanioVision and EthoVision software were used to record and quantify movement of zebrafish larvae at 6 dpf. Larvae were acclimated in the DanioVision box for 5–10 min before using swimming data for analysis. Embryos were loaded into the DanioVision in 12-well cell culture plates in replicates and placed in wells at random. Five minutes of recorded behavioral data was analyzed for the distance traveled. The EthoVision software tracked and recorded fish movement data.

2.7.11 Light stimulated motor activity

Light stimuli were administered to zebrafish embryos which were individually embedded in their chorions in an agarose mold. They were kept in an opaque box in a 28.5ºC incubator between timepoints. At 17, 21, 24, 26, 28, 30, 32, 34 and 48 hpf, each mold was placed on a dissecting microscope stage, and a video recording was captured. The embryos were subjected to two cycles of 30 seconds in the dark and 30 seconds under light stimulus, with a final 30 second dark period afterwards. The number of times each embryo twitched was counted in each recording.

2.7.12 Swim tunnel quantification

Zebrafish were loaded into a swim tunnel with a steadily increasing flow. Once an individual fish fatigued, it lay against the back grate of the tunnel and was removed from the experiment to a separate tank, and the time to fatigue was recorded. After acclimating with no flow, the fish were subjected to 10 cm/s flow for 1 min. The flow then increased by 2 cm/s every minute until every fish fatigued.

2.8.13 Statistical analysis

Differences in number of twitches, distance traveled, α-actinin intervals, and myoseptal intervals were quantified using nonparametric tests (Wilcoxon rank-sum test), so normal

distribution was not assumed in the statistical tests. An estimate of variation was not included in the statistical analysis, but seems to vary between compared groups. R was used to analyze data and generate graphs.

Sample size was determined based on experience from previous findings and animal availability. Depending on the nature and complexity of the assay, the maximum number of animals that could be tested was used to increase statistical power as much as possible. No samples were excluded from analysis. Zebrafish were chosen at random as subjects before genotype was determined. When determining survivability of zebrafish genotypes, entire clutches were genotyped to avoid selective bias. The investigator was not blinded before assessing skeletal, muscular, or gross anatomical abnormalities of zebrafish.

Chapter 3: A higher-throughput method of targeted mutagenesis to functionally test variants *in vivo*

3.1 Abstract

While inducing a targeted deletion into the genome using CRISPR is straightforward and efficient, insertion of a mutation via homologous recombination is much less probable and often out of frame. As discussed previously, we generated an *smyhc1^R673H* zebrafish line to functionally test the variant *in vivo*. However, to create this stable line, thousands of zebrafish embryos were injected and screened before germline transmission was detected. To identify our founder with the correct single base pair substitution, we genotyped the gametes of many individual F0 fish, which was very time and labor intensive. Many SNPs are reported in human patients. However, without an *in vivo* model, it is difficult to functionally test the pathogenicity and functional effects of these mutations. Using the previous method to produce and test these variants is unfeasible. Therefore, a higher throughput method of mutagenesis is necessary. We developed a plan to create a founder zebrafish line that would allow us to easily detect embryos in which an in-frame knock-in had occurred. Specifically, we inserted a GFP tag to the end of *smyhc1* and deleted a coding region of interest, where many SNPs cluster. This knocks the GFP out of frame, ceasing expression of GFP. The deleted region provides a location for introducing new single nucleotide variants. Upon accurate recombination and insertion of a variant in frame, the GFP will be become in frame, inducing GFP expression as a reporter that enables efficient screening of F0 embryos. This may drastically streamline the process of generating missense mutants in a gene of interest in zebrafish, allowing for functional testing of many variants of unknown significance.

3.2 Introduction

Identification of variants in a candidate disease gene is an important step in diagnosis, and can improve our understanding of the development of congenital disorders (van der Sijde et al, 2014). Guidelines have been established to help estimate the probability of pathogenicity for clinical variant interpretation. First, the allele frequency of a variant in the general population can be used to is an indication of variant pathogenicity. Evolutionarily, it is well known that rare variants are more likely to be detrimental than common ones, as those have undergone selection and would have been bred out of a population if deleterious. Second, segregation of a variant with disease phenotype increases the confidence that a given variant is linked with a specific congenital disorder. Modeling can also be used to predict the effect of a SNP on protein structure, stability, and chemical properties (Ioannidis et al, 2016; Kircher et al, 2014; Rentzsch et al, 2019; Thusberg et al, 2011; Visscher et al, 2012). However, these computational methods are limited in determining the functional effect of variants on protein behavior, how they interact with other proteins, and the effect on cell structure and tissue as a whole. Thus, the gold standard is often to test the functionality of a gene variant *in vitro* or *in vivo*.

Direct comparison of wild type and mutant gene function is crucial to understanding the mechanisms by which genetic variants cause congenital phenotypes. *In vitro* testing can examine chemical properties of wild type and mutant proteins, including ATPase activity and interaction with ligands (Echwald et al, 2002; Han et al, 2000; Racca et al, 2015; Silva et al, 2012; Walklate et al, 2016). However, when genes have non-cell autonomous effects, *in vivo* assays may be more useful in exploring a disease mechanism. For example, a vertebrate model is necessary to understand how a gene expressed in muscle may affect skeletal development and morphology. In

addition, *in vivo* assays allow us to test allele dosage effects, as well as functional effects of interaction between wild type and SNP mutant proteins.

Functional testing of *MYH3* variants is especially important due to the large number of variants of uncertain significance reported in ClinVar (Table 3.1). Many missense mutations in *MYH3* are reported in association with DA and other musculoskeletal disorders, but the mechanism by which these different variants alter protein function and lead to disease phenotype has not been robustly explored.

Development of bone, joints, and muscle are interlinked, and the full effect of mutant *MYH3* function is difficult to study without established vertebrate models. Patient tissue is scarce, and *in vitro* methods of protein and tissue characterization are incomplete, as the skeleton is necessary in the model to fully grasp the mechanism by which these mutations alter joint and bone structure. To functionally test the effects of a reported *MYH3* variant, we generated a zebrafish model for a missense mutation commonly associated with DA (*smyhc1^{R673H}*) (Whittle et al, 2020). While we were able to characterize the muscle structure and function, and later reverse the disease phenotype, this represents only one example of a pathogenesis mechanism of a DA variant. The current methods for generating mutant lines with single nucleotide variants, such as this, are not very efficient, and most are both time and labor intensive.

Here, we outline a method for higher throughput variant generation in zebrafish using a GFP reporter. GFP is inserted at the end of the *smyhc1* gene, and subsequently knocked out of frame as a region of interest in *smyhc1* is removed. Upon accurate in-frame repair with a template containing a single nucleotide variant, the GFP will again be in frame, which allows efficient screening of F0 embryos for successful variant integration. This strategy will allow us to functionally test a broad array of mutations much more efficiently.

Table 3.1 MYH3 variants reported in ClinVar. Many missense variants of unknown significance have been reported in *MYH3*. (Accessed on ClinVar 10 August 2022)

Pathogenicity	Missense	Nonsense	Frameshift	Splice site	UTR
Benign	15	0	0	0	1
Likely Benign	24	0	0	0	1
Uncertain Significance	141	6	2	2	2
Likely Pathogenic	20	3	5	6	0
Pathogenic	10	2	0	5	0
Conflicting Interpretations	29	1	1	1	2
Total	239	12	8	14	6

3.3 Generation of *smyhc1*-2A-GFP mutants

Initial generation of *smyhc1*-C-terminal GFP fusion mutants yielded a homozygous lethal allele (data not shown). GFP protein tags have been previously observed to alter protein function (Oshikane et al, 2018). There are also cases described where the location of a GFP tag has differing effects on protein localization and cell fitness (Skube et al, 2010; Weill et al, 2019). Because this protein dimerizes and bundles into thick filaments, it is possible that the GFP tags affected function when congregated into these large complexes. It is also possible that the tag altered localization of Smyhc1. Being such a large protein, (1969 residues, 226.78 kDa [Ensembl]), errant localization could have drastically deleterious effects on the cells and tissue. Furthermore, as seen in our *smyhc1*^{R673H} mutants, alteration of the function of this gene may also result in homozygous lethality.

Any phenotype causing homozygous lethality will likely confound the phenotype we attempt to observe in our induced missense mutants. Therefore, we pursued a new construct with a 2A self-cleaving peptide linker between *smyhc1* and GFP at the C terminus, resulting in cytosolic GFP in cells expressing Smyhc1 (allele *smyhc1*-2A-GFP) (Figure 3.1, 3.2) (Donnelly et al, 2001; Minskaia et al, 2013; Ryan et al, 1991). We expect that this will eliminate the lethality that occurred with GFP directly added to the C-terminus of Smyhc1.
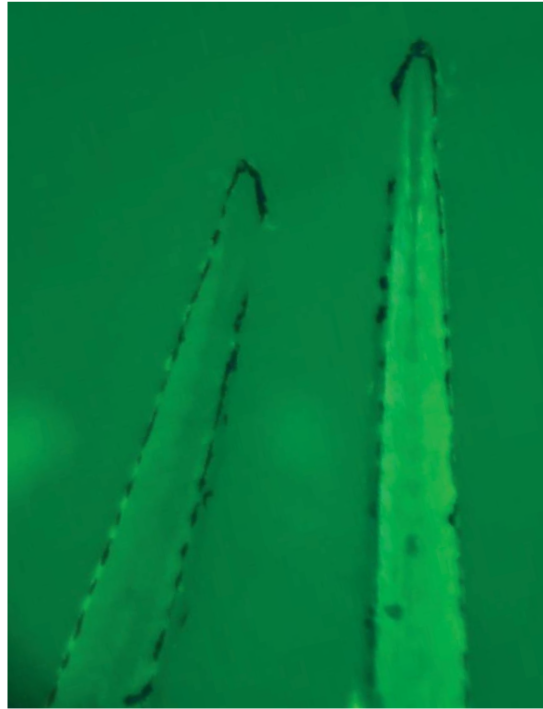
Figure 3.1 Expression of GFP in *smyhc1*-2a-GFP zebrafish larvae. *smyhc1*-2A-GFP F0 mosaic fish were crossed to wild-type. Comparison of resulting *smyhc1*-2A-GFP / *smyhc1*⁺ F1 larva (right) and wild-type sibling (left).
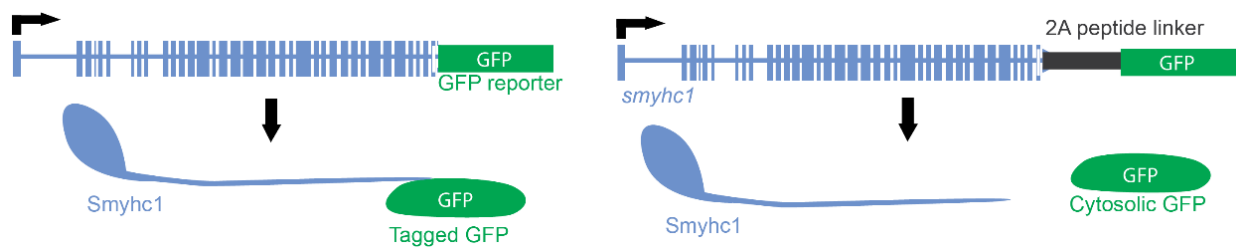
Figure 3.2 Effect of inclusion of 2A peptide self-cleaving linker between

smyhc1 **and GFP.** The original construct yielded a Smyhc1 protein that we tagged

with GFP at the C-terminus (left), vs the 2A peptide linker, which still allows both

Smyhc1 and GFP to be translated from the same transcript, but enables cleavage of

GFP from Smyhc1, resulting in cytosolic GFP.

3.4 Future directions

We plan to use these zebrafish mutant lines to more efficiently induce and screen for variants of interest in exons 2-12 of *smyhc1*, the *MYH3* analog in fish (*smyhc1*). This collaborative effort between the Christina Gurnett and Aaron Johnson labs will be continued by other lab members because the lengthy breeding strategy necessary to complete this project was outside the scope of this thesis.

3.4.1 Future method of high-throughput screening for in-frame mutagenesis

With the GFP reporter, we can efficiently screen for in-frame variant integration. Because pathogenic variants cluster in the motor domain of *MYH3*, we plan to target the analogous region of *smyhc1* for variant testing. This region of interest, encoded by exons 2-12, will be knocked out via a 2-guide CRISPR approach. This will generate a large deletion and frameshift in *smyhc1*, and also put the GFP out of frame (allele *smyhc1*-2A-GFP ^{del}).

Once this knockout (*smyhc1*-2A-GFP ^{del}) is established as a stable line, we will design a single CRISPR gRNA to target the deletion site. Co-injecting a CRISPR cocktail with this gRNA and a repair vector including a variant of interest will induce in-frame integration of the variant at some efficiency (Table 3.2). This accurate in-frame insertion will also put the GFP into frame. Fish with successful integration can be efficiently screened via GFP expression for analysis (Figure 3.3). Fish with high GFP expression (many cells expressing GFP) are more likely to have germline integration of the variant as well.

The benefit of this approach is the ability to screen F0 embryos for correct homologous repair. Using the older, previously described method of variant mutagenesis, all F0 fish needed to be grown to adulthood to test for successful in-frame integration of a variant in the germline via high-throughput sequencing. Using this new innovative method, not only are F0 fish able to be

screened much sooner, but we can also detect individual cells with apparent in-frame insertion. Therefore, there is the potential for phenotype analysis in the F0 mosaic fish as well.

3.4.2 Future method of F0 variant phenotype analysis using GFP reporter

We can specifically target the *smyhc1*-2A-GFP^{del} allele by designing a gRNA targeting the newly created deletion site. *smyhc1*-2A-GFP^{del} / *smyhc1*$^+$ heterozygotes can be injected with this gRNA and a repair vector in a CRISPR cocktail, where only the *smyhc1*-2A-GFP^{del} allele will be targeted. Cells where this variant integration is successful will express GFP, allowing these cells to be examined individually for a heterozygous phenotype (Figure 3.4).

In embryos, slow muscle cells where *smyhc1* is expressed are mononucleate (do not fuse together), ensuring that green cells are not resulting from leaked GFP from neighboring cells (Hromowyk et al, 2020; Shi et al, 2017).

Previously, we observed significantly altered muscle morphology in our previously described heterozygous *smyhc1*$^{R673H/+}$ mutants. The individual muscle fibers of these mutant fish were malformed with shortened sarcomeres as early as 1 dpf (Whittle et al, 2020). In addition, DA-associated *MYH3* variants are almost always *de novo* heterozygous mutations in human patients (Beck et al, 2013; Scala et al, 2018; Toydemir et al, 2006; Zhao et al, 2022). This suggests that the GFP-expressing *smyhc1*-2A-GFP^{VoI} / *smyhc1*$^+$ cells that we generate in F0 fish may have pathogenic muscle phenotypes even at early stages of development.
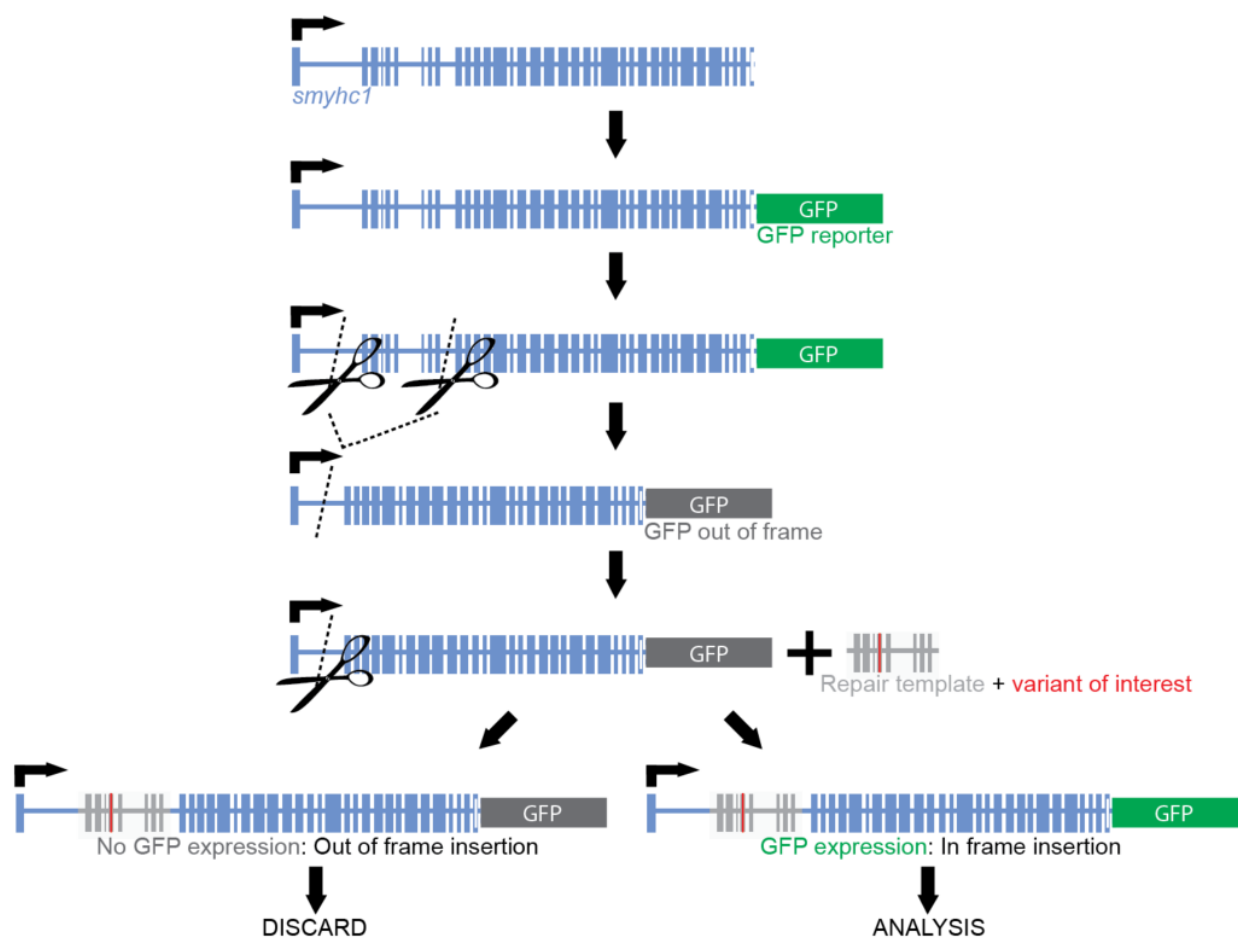
Figure 3.3 Workflow schematic of high throughput variant mutagenesis and screening. A GFP reporter is inserted to the end of *smyhc1*. Region of interest is removed via CRISPR using two gRNAs and non-homologous end joining (NHEJ). A gRNA targeting the deletion site is co-injected with a repair template containing a variant of interest (VoI). In-frame homology-directed repair (HDR) integrating the repair template sequence will put the GFP back into frame, allowing for efficient screening of F0 embryos.
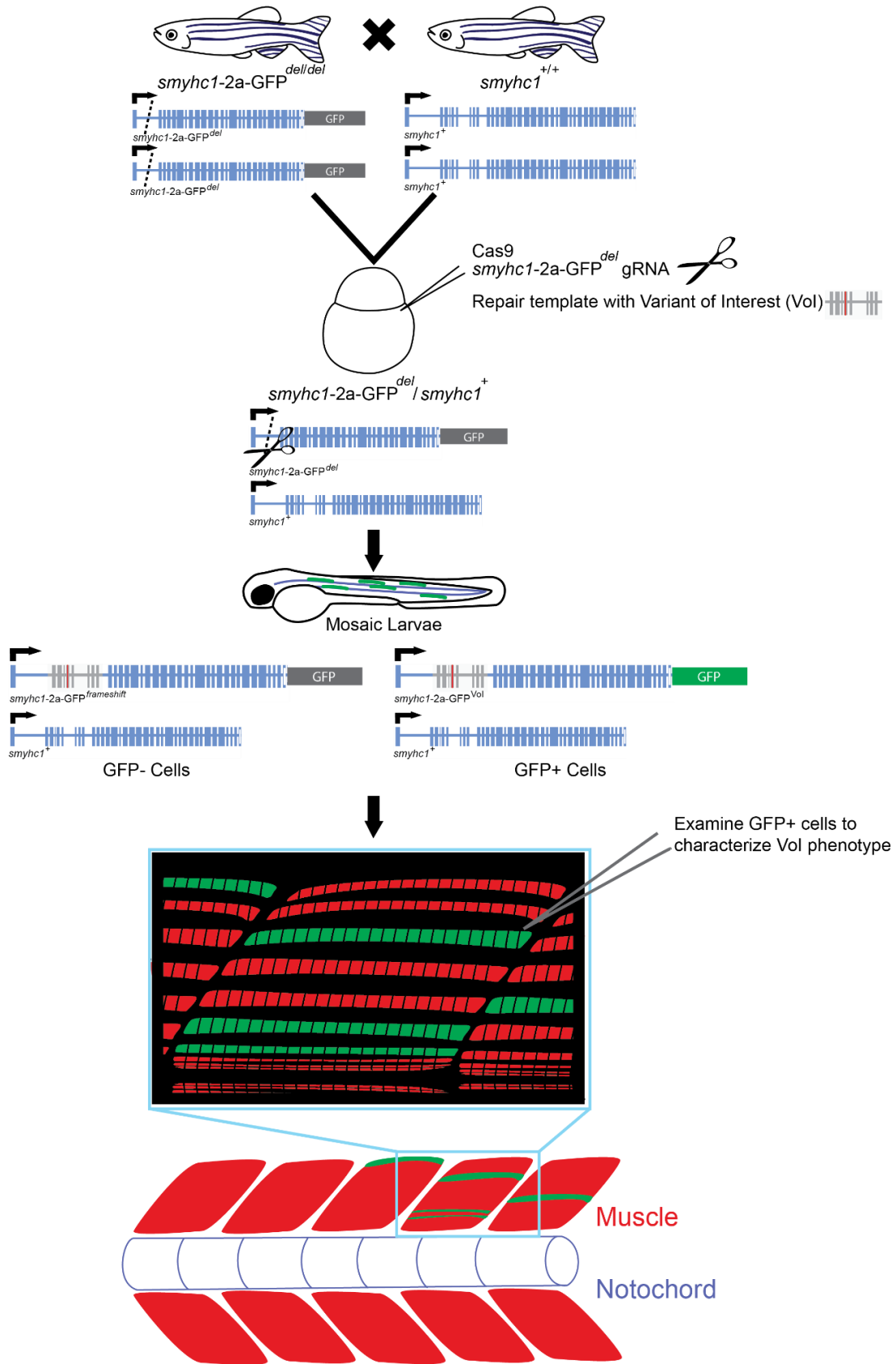
Figure 3.4 Workflow schematic of F0 heterozygous variant phenotype analysis in mosaic larvae. After outcrossing *smyhc1*-2a-GFP*del/del* homozygotes to *smyhc1+/+*, we can target the *smyhc1*-2a-GFP*del* allele for CRISPR mutagenesis. When co-injected with a repair template containing a variant of interest, the *smyhc1*-2A-GFP allele will be cut at the region of deletion, allowing for homology-directed repair (HDR) integrating the repair template into the allele. Upon in-frame integration of the template, the GFP will be knocked into frame at some level of efficiency. The mosaic embryos will express GFP in cells with in-frame integration of the repair template. These GFP-expressing cells can be examined and characterized in an efficient analysis of F0 embryos.

Table 3.2 Candidate variants for functional testing. (Toydemir et al 2006), ClinVar

MYH3 Variant	Associated disease	Pathogenicity	Analogous *smych1* mutation	Accession #
P83R	DA2A, DA2B	Conflicting reports	P82R	VCV000321773
M91L	NA	Likely pathogenic	M90L	VCV000452380
T178I	DA2A, DA2B	Pathogenic	T177I	VCV000014140
S261F	DA2B	Likely pathogenic	S259F	VCV000694358
S292C	DA2B	Conflicting reports	S290C	VCV000211557
M363L	HCM	Likely benign	M361L	VCV000220731
E375K	DA2B	Conflicting reports	E373K	VCV000014143
Y387C	DA2B	Pathogenic	Y385C	VCV000631471
R404T	DA2A,DA2B	VUS	R402T	VCV000321762
D462G	DA2B	VUS	D460G	VCV000014144

3.5 Materials and methods

3.5.1 Ethical statement

Husbandry and protocols for experiments involving zebrafish (*Danio rerio*) have been reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at Washington University.

3.5.2 Zebrafish husbandry

Zebrafish (*Danio rerio*) were raised and maintained using standard methods (Westerfield, 1995). Zebrafish were housed and handled using protocols approved by the Institutional Animal Care and Use Committee (IACUC) at Washington University.

3.5.3 Generation of *smyhc1*-2A-GFP Zebrafish Mutants

We used a dual gRNA approach for this CRISPR protocol (dgRNA). 10 uL of 100 µM Alt-R ® tracrRNA (IDT 1072534) was added to 10 uL 100 uM crRNA (IDT custom [/AlTR1/rArA rCrA rArA rCrA rArA rCrU rCrG rUrU rArU rArC rGrU rUrU rUrA rGrA rGrC rUrA rUrG rCrU/AlTR2/]). This was incubated at 95 °C for 5 minutes, then cooled to 25 °C at a rate of -0.1 °C/second. It was kept at 25 °C for 5 minutes, then stored at -20 °C in 1 uL aliquots (50 uM). 1 uL dgRNA, 1 uL Cas9 protein (25 uM), 0.5 uL phenol red, and 2.5 uL KCl (300 mM), were mixed and incubated at 37 °C for 5 minutes. Next, 2 uL *smyhc1*_2A_GFP (100 ng/uL) repair vector and 1 uL NU7441 (400 mM) were added. This cocktail was injected into 1-cell stage zebrafish embryos.

Embryos were screened between 1 and 5 dpf for GFP expression in F0 and F1 embryos.

Chapter 4: *FNDC1*-associated scoliosis and zebrafish model

4.1 Preface

Parts of this chapter were adapted from the following manuscript:

Antunes, L., Haller, G., **Whittle, J.**, Avery, A., Morcuende, J.A., Giampietro, P. , Raggio,C. , Hadley-Miller, N., Strande, N.T., Seeley, M., Bodian, D.L., Sepich, D. , Dobbs, M.B., Gurnett, C.A."Severe adolescent idiopathic scoliosis is associated with ultra-rare missense variants in FNDC1 and other core matrisome genes." *In Preparation*

Author Contributions:

LA and CAG designed the study and wrote the manuscript. LA carried out data processing and statistical analysis. LA, GH and CAG analyzed and interpreted the statistical results. LA, JW, and CAG designed the laboratory experiments (zebrafish work). NTS and BLD performed and interpreted all analyses pertaining to the Geisinger dataset. MS reviewed scoliosis phenotypes and ICD codes extracted from the Geisinger HER. All authors reviewed and approved the final version.

4.2 Abstract

Adolescent idiopathic scoliosis (AIS) is the most common pediatric spinal deformity. Major spinal fusion surgery is often required for patients with progressive AIS. Although this disorder runs in families, the genetic risk factors for severe disease are largely unknown. To explore the role of rare variants in severe AIS, a gene burden analysis of 1,221 cases revealed an enrichment of ultra-rare missense variants in core matrisome genes (p=7.19 x 10^{-5}, OR =1.1). Most significant enrichment was observed in *fibronectin type III domain containing 1* (*FNDC1*), which was previously associated with joint hypermobility. We found *FNDC1* harbored variants in 3.9% of

AIS cases compared to 1.8% of controls (p=6.63 x 10^{-4}, OR=2.2). The association of *FNDC1* rare variants with scoliosis replicated in the Geisinger MyCode cohort (p=0.0002, OR=3.6). To explore the functional effects of *FNDC1* deficiency, we characterized an *fndc1* zebrafish mutant. While we did not observe a gross phenotype, disruption of the *fndc1* locus in zebrafish caused abnormalities in the vertebrae. MicroCT analysis revealed increased bone mineral density, cross-sectional area, and altered the pMOI score, suggesting a possible mechanism of pathogenesis. In summary, our data implicate *FNDC1* as a novel gene contributing to severe AIS and support a key role for core matrisome gene variation in scoliosis pathogenesis.

4.3 Introduction

Scoliosis is the most common pediatric spinal deformity. In more than 80% of patients, scoliosis is idiopathic and appears during adolescence. Commonly referred to as adolescent idiopathic scoliosis (AIS), it presents without any apparent congenital spinal deformities. Risk factors for curve progression include curve type, age, sex, and lack of joint hypermobility (Czaprowski, 2014; Czaprowski et al, 2011; Miller, 1999). Although this disorder runs in families, the genetic factors that contribute to severe idiopathic scoliosis are largely unknown.

To identify genetic risk factors for severe AIS, we generated whole exome sequence (WES) data from a large multicenter cohort of 1,221 cases with severe AIS, defined as spinal curvatures >35 degrees. Here, we performed burden analyses to investigate the genetic contribution of ultra-rare missense variants and identified a strong association of variants in *FNDC1* and other core matrisome extracellular matrix genes with AIS.

Following this discovery, we then examined *fndc1* nonsense mutant zebrafish to determine how this gene may contribute to AIS progression. We observed no gross phenotype in the fish,

and performed further analysis with microCT. Here, we found subtle differences in the vertebrae between *fndc1* mutants and their wild type siblings. While we did not observe a scoliosis phenotype in our model, our findings suggest alterations in vertebral bone structure may act as risk factors for scoliosis progression in human patients.

4.4 Results

4.4.1 Study samples and quality control

To identify the role of genetic variants in severe AIS susceptibility, we generated and analyzed whole exome sequence (WES) data from a multicenter cohort of unrelated individuals of European ancestry with severe juvenile or adolescent-onset idiopathic scoliosis, defined as a Cobb angle of at least 35 degrees. Overall, this severe AIS cohort was 84% female and had a mean Cobb angle of 55° and a mean age of 23.8 years (Supplemental Table 4.1). In-house control WES data was generated using the same methodologies and consisted of individuals who were ascertained for conditions other than scoliosis, including 83% with Alzheimer's disease or amyotrophic lateral sclerosis.

Following WES data processing and quality control, we analyzed data from 2,618 unrelated individuals of European ancestry that included 1,221 severe AIS cases and 1,397 in-house controls that were matched by ancestry. Multiple steps were taken to minimize spurious associations, including: (1) joint processing of read alignment, variant calls, variant QC, and annotation for AIS cases and in-house controls using the same methods and tools; (2) confirmation of genetic ancestry homogeneity using principal component analysis; (3) confirmation that all individuals were unrelated using pairwise relationship analysis; and (4) control for potential

confounders, including Hardy Weinberg equilibrium, genotype call rate and missingness (Data not shown).

4.4.2 Rare variants in core matrisome genes are enriched in severe AIS cases

To explore the role of rare variants in the susceptibility to severe AIS, we performed a gene-based rare variant burden analysis on WES data from 1,221 severe AIS cases and 1,397 in-house controls. Our AIS cohort was ascertained due to an extreme severe scoliosis phenotype, with the majority of cases having had spinal fusion surgery. This aggressive treatment occurs in only 1 in 10,000 individuals (Martin et al, 2014). To our knowledge, no replication cohort of severe scoliosis cases exists. Therefore, to maximize the likelihood of detecting true enrichment of rare variants within a single gene, (and minimize the risk of spurious association due to small number of individuals and lack of a traditional replication cohort), we repeated the burden analysis by comparing our AIS case data to a second, much larger control cohort of 125,748 individuals from gnomAD global population using previously described methods (Guo et al, 2018).

By limiting our analysis to ultra-rare nonsynonymous missense variants, defined as a minor allele count (MAC) ≤ 3 in case and control cohorts, we sought to identify extremely rare variants contributing to the familial clustering often seen in AIS. Analysis was restricted to genes with a greater burden of ultra-rare variants in cases compared to controls (OR >1). There were 119 genes with burden analyses yielding p-values < 0.05 compared to both in-house controls and gnomAD global population controls. No single gene surpassed exome-wide significance in this analysis. Two of the top 10 genes with the most significant enrichment of ultra-rare nonsynonymous variants in AIS cases, *FNDC1* and *FBN1*, belong to the core matrisome gene set (Naba et al, 2016; Naba et al, 2012; Naba et al, 2017) (Table 4.1).

4.4.3 Association of AIS with *FNDC1* rare variants

Within the group of 262 core matrisome genes, the strongest enrichment of rare variants in AIS cases was seen in *FNDC1 (fibronectin type III domain containing 1)*(Table 4.1 and Figure 4.1), which harbored ultra-rare missense variants in 3.9% of severe AIS cases (48/1221) compared to 1.8% of in-house controls (25/1397)(p=6.63 x 10^{-4}, OR = 2.2). Enrichment of rare variants in *FNDC1* in AIS cases was also strongly confirmed in comparison to 125,748 gnomAD global population controls (p=5.76 x 10^{-4}, OR = 2.3). Clinical features including age, Cobb angle, height, and Beighton hypermobility scores did not differ between AIS patients harboring *FNDC1* rare variants and those without (Supplemental Table 4.1,4.2).

A similar association was observed between rare *FNDC1* nonsynoymous missense variants and scoliosis in patients of European ancestry (n = 84,717) in the Geisinger MyCode cohort (p=0.027, OR=1.5), a healthcare population with paired exome and clinical data (Table 4.2).

4.4.4 *fndc1* stop-gain effects in zebrafish

To understand the mechanism by which *FNDC1* contributes to AIS severity, we investigated the functional effects of a nonsense allele in the zebrafish line, *fndc1^{sa23734}* (Figure 4.4). Morphological analysis of *fndc1$^{+/+}$*, *fndc1$^{+/ sa23734}$*, and *fndc1$^{sa23734/sa23734}$* zebrafish did not reveal increased rates of scoliosis, spinal fusions, or other gross abnormalities. To further examine the fish, we used quantitative analysis of micro-computed tomography (microCT). This is a technique wherein X-ray projection images are taken at multiple angles from a fixed sample, generating a stack of visual slices. From this, a 3-dimensional image can be recreated and various measurements can be extracted (Cheng et al, 2011). We isolated three individual vertebral centra

from each fish for analysis (vertebrae 5th, 6th, and 7th from the tail). This allowed us to individually the parameters of specific bones (Figure 4.2).

First, we examined cross-sectional area of the vertebrae. While cross-sectional area of bone was only slightly different between *fndc1*^{sa23734/sa23734} and wild type siblings, total cross-sectional area was significantly different between fish carrying the *fndc1*^{sa23734} allele and wild type siblings (Figure 4.3). This suggests that the hollow within the centra is larger in mutant fish.

Next, we examined tissue mineral density in the bone of the vertebrae. This measures the density of HA (calcium-hydroxyapatite) per unit volume in the bone tissue (Mei et al, 2017). HA is the form in which calcium crystalizes in the bones (O'Keefe et al, 2016). This data revealed that zebrafish carrying *fndc1*^{sa23734} alleles had significantly increased bone mineral density (Figure 4.3).

Given that vertebrae are hollow, we also measured the average thickness of this bone. While this measurement was not significantly different between genotypes, the standard deviation of bone thickness was significantly different in *fndc1*^{sa23734} homozygotes (Figure 4.3). While the average thickness of the vertebrae was the same in each genotype, it appears that *fndc1*^{sa23734} homozygotes have more variability of bone thickness in each vertebra than siblings.

We also used this data to calculate polar moment of inertia (pMOI) of each vertebrae. This is a calculation of bone strength via spatial parameters of the bone (Wergedal et al, 2003). The calculation determines the distribution of bone around the center of the cross section. This calculation is inversely related to shear stress in the bone induced by a given level of torque (Schoenau et al, 2001). This is a predictor of torsional stiffness based on shape (Martin & Burr, 1984). We found significantly higher pMOI in fish carrying the *fndc1*^{sa23734} allele (Figure 4.3).

4.4.5 Investigating *fndc1^sa23734* expression

Currently, there are no Fndc1 antibodies to our knowledge available to test for protein translation in zebrafish, via Western blot or IHC. Therefore, we opted to test for *fndc1* transcript in the *fndc1^sa23734* line. We performed reverse-transcriptase PCR (RT-PCR) on RNA extracted from *fndc1^sa23734/sa23734* fish and wild type sibling controls. Unfortunately, we were able to amplify RNA (data not shown), indicating that the stop-gain mutation may not have induced complete nonsense-mediated decay, and there may be truncated protein produced in cells.

4.4.6 Generation of true *fndc1* knockout

FNDC1 has several alternative splice variants in humans, which may provide one explanation for this observation. Therefore, to further study the role of *fndc1* in skeletal development, we designed a CRISPR guide to induce a frameshift mutation near the N-terminus of the gene, as well as in reported splice variants (Figure 4.4). This would allow us to corroborate our mutant phenotype with a true knockout zebrafish *fndc1* line.

Preliminary high-throughput sequencing via MiSeq of F0 gRNA target site 1 crispants has returned a high degree of deletion induction, which is encouraging in estimation of gRNA efficiency (Table 4.3). We observed an average deletion efficiency of 0.523126 in whole embryos, where an efficiency of 1 indicates total DNA cutting at the gRNA site.

This line is currently being generated, and its characterization will be performed by other members of the Gurnett Lab in the future.
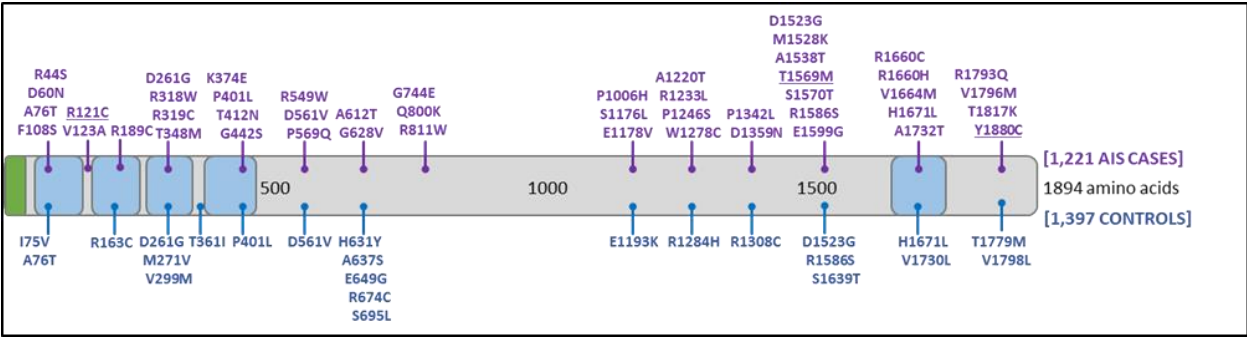
Figure 4.1. *FNDC1* **ultra-rare nonsynoymous missense variants are enriched in AIS cases.** Variants observed in AIS cases (purple) and in-house controls (blue). Green, N-terminal signal peptide. Blue, fibronectin type 3 domain.
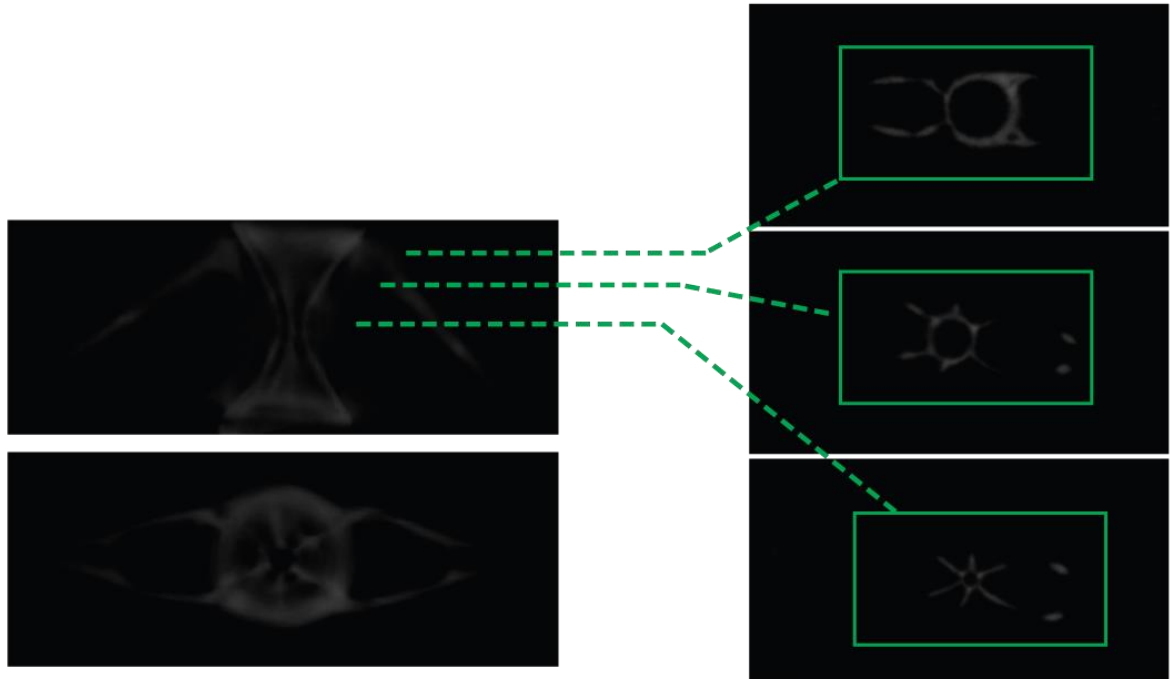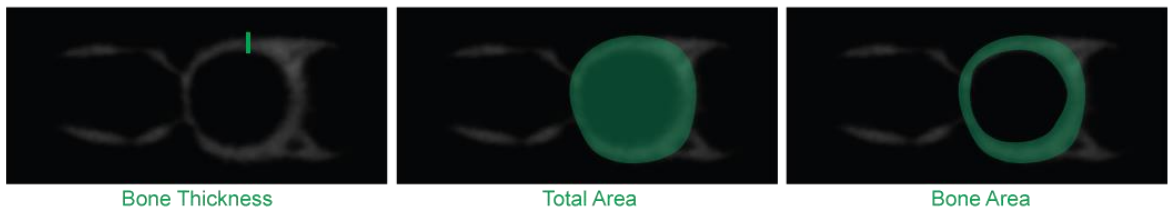
Figure 4.2 Representative microCT images of isolated adult zebrafish vertebrae. (A) Lateral (top left) and rostral (bottom left) microCT images of zebrafish vertebra. Images of corresponding cross-sections along length of vertebra shown to the right. (B) Images representing source of measurements taken from microCT.
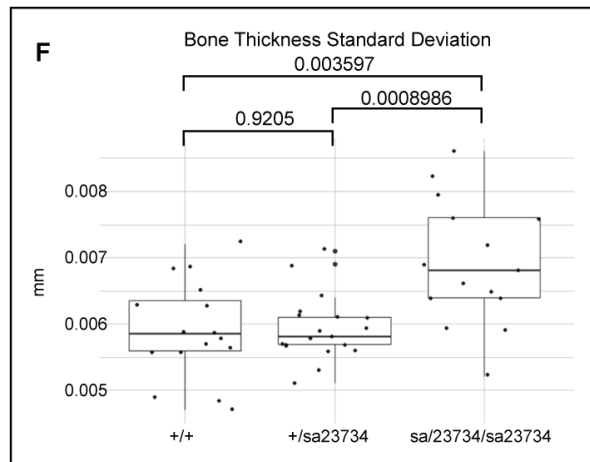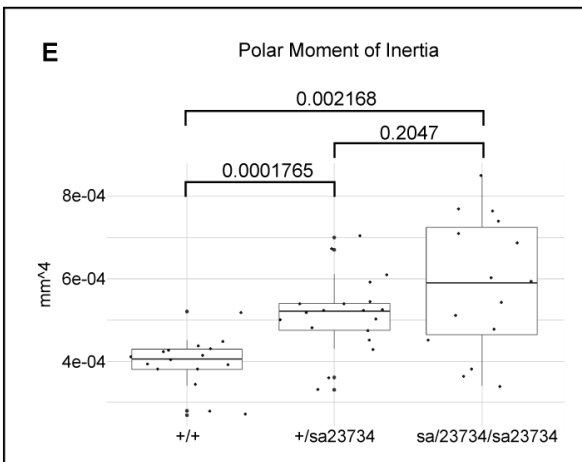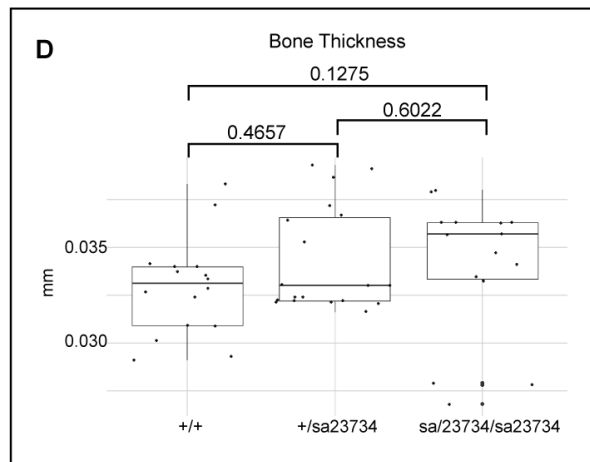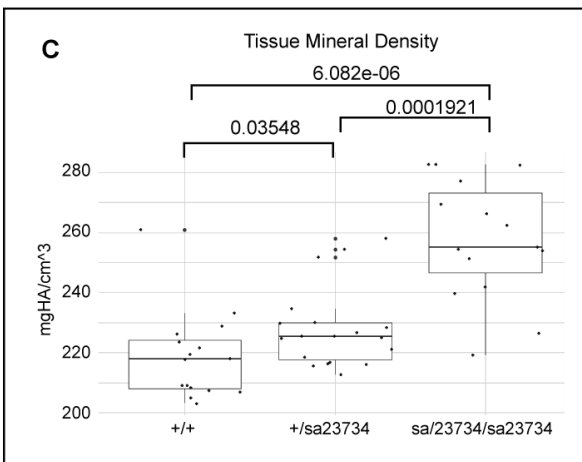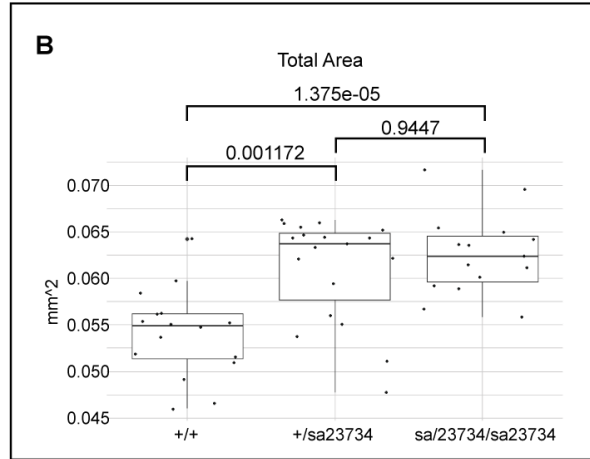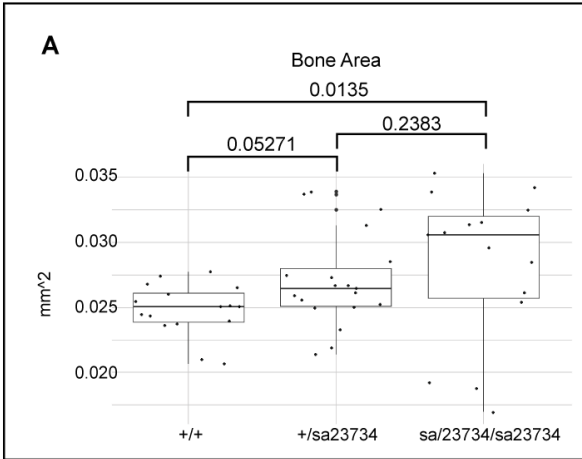
Figure 4.3 The *fndc1* nonsense allele (*fndc1sa23734*) is associated with differences in vertebral centra. MicroCT scans of zebrafish (N=5 for each genotype) were analyzed for various parameters. 3 vertebrae from each fish were examined individually. P-values comparing each genotype are displayed. (A) Mutant fish may have slightly larger area of bone in cross-section. (B) Mutant fish have significantly increased total vertebral cross-sectional area compared to wild-type siblings. (C) Mutant fish have significantly higher tissue mineral density in vertebrae, scaling with *fndc1sa23734* allele dosage. (D) There is no significant difference in average cross-sectional bone thickness between siblings. (E) Mutant heterozygotes and homozygotes have higher polar moment of inertia compared to wild type. (F) Mutant homozygotes have a higher bone thickness standard deviation compared to siblings, indicating more variability in bone thickness. P-values were calculated using Wilcoxon rank-sum test, two-sided. Points represent individual vertebrae.

Figure 4.4 Schematic of *fndc1* zebrafish gene and splice variant. *fndc1^{sa23734}* nonsense allele indicated in red. gRNA target sites indicated in orange.

Table 4.1 Ultra-rare nonsynoymous missense variants in core matrisome genes are enriched in severe AIS. Top 15 core matrisome genes with variant counts and percentages (in parenthesis) for severe AIS cases and in-house controls having p-values ≤ 0.05. Combined all 262 tested core matrisome genes, combined top 15 core matrisome genes, and combined core matrisome for Collagens, ECM Glycoproteins and Proteoglycans respectively. P-values were calculated using Fisher's exact, one-sided.

	Core Matrisome Type	Severe AIS cases (N=1,221)	In-house controls (N=1,397)	OR	P-value	Severe AIS cases (N=1,221)	gnomAD controls (N=125,748)	OR	P-value
FNDC1	ECM Glycoproteins	48 (3.9%)	25 (1.8%)	2.2	6.63E-04	21 (1.7%)	945 (0.8%)	2.3	5.76E-04
FBN1	ECM Glycoproteins	45 (3.7%)	25 (1.8%)	2.1	1.97E-03	26 (2.1%)	1123 (0.9%)	2.4	7.30E-05
COL2A1	Collagens	28 (2.3%)	13 (0.9%)	2.5	3.97E-03	13 (1.1%)	597 (0.5%)	2.3	6.99E-03
PXDNL	ECM Glycoproteins	34 (2.8%)	18 (1.3%)	2.2	4.63E-03	6 (0.5%)	700 (0.6%)	0.9	0.67
PAPLN	ECM Glycoproteins	42 (3.4%)	26 (1.9%)	1.9	7.95E-03	8 (0.7%)	545 (0.4%)	1.5	0.17
OMD	Proteoglycans	5 (0.4%)	0	-	0.02	1 (0.1%)	192 (0.2%)	0.5	0.85
HAPLN1	Proteoglycans	13 (1.1%)	5 (0.4%)	3.0	0.03	6 (0.5%)	178 (0.1%)	3.5	9.17E-03
AGRN	ECM Glycoproteins	46 (3.8%)	34 (2.4%)	1.6	0.03	11 (0.9%)	936 (0.7%)	1.2	0.31
EMID1	ECM Glycoproteins	11 (0.9%)	4 (0.3%)	3.2	0.03	2 (0.2%)	186 (0.1%)	1.1	0.54
MFAP5	ECM Glycoproteins	8 (0.7%)	2 (0.1%)	4.6	0.03	4 (0.3%)	98 (0.1%)	4.2	0.02
TNR	ECM Glycoproteins	38 (3.1%)	27 (1.9%)	1.6	0.04	16 (1.3%)	688 (0.5%)	2.4	1.59E-03
LAMB4	ECM Glycoproteins	28 (2.3%)	18 (1.3%)	1.8	0.04	7 (0.6%)	953 (0.8%)	0.8	0.82
COL14A1	Collagens	26 (2.1%)	17 (1.2%)	1.8	0.05	11 (0.9%)	832 (0.7%)	1.4	0.19
COL26A1	Collagens	10 (0.8%)	4 (0.3%)	2.9	0.05	2 (0.2%)	182 (0.1%)	1.1	0.53
OTOG	ECM Glycoproteins	61 (5%)	51 (3.7%)	1.4	0.05	-	-	-	-
Core matrisome genes (N=262)		4799 (1.5%)	5088 (1.4%)	1.1	7.19E-05	1403 (0.5%)	125874 (0.4%)	1.1	2.5E-07
Top core matrisome genes (N=15)		443 (2.5%)	269 (1.3%)	1.9	2.60E-17	134 (0.8%)	8155 (0.5%)	1.7	1.4E-08
Core - Collagens genes (N=42)		1046 (2.1%)	1183 (2.1%)	1.0	0.40	318 (0.6%)	28052 (0.5%)	1.2	3.8E-03
Core - ECM Glycoproteins genes (N=188)		3323 (1.5%)	3493 (1.3%)	1.1	2.3E-04	966 (0.4%)	86098 (0.4%)	1.2	6.4E-06
Core - Proteoglycans genes (N=32)		430 (1.1%)	412 (0.9%)	1.2	5.3E-03	119 (0.3%)	11724 (0.3%)	1.0	0.33

Table 4.2 Ultra-rare *FNDC1* nonsynoymous missense variants are associated with scoliosis in the Geisinger MyCode cohort.

Diagnosis Codes in EHR	*FNDC1* Positive (N=481)	*FNDC1* Negative (N=84,248)	OR	P-value
Any Scoliosis Diagnosis	36 (7.48%)	4509 (5.35%)	1.43	0.0290
Multiple Scoliosis Diagnoses	18 (3.74%)	1555 (1.85%)	2.07	0.0043
Scoliosis Diagnosis Before 18 y/o	9 (1.87%)	466 (0.55%)	3.46	0.0017

Table 4.3 Deletion in duction efficiency in *fndc1* target site 1 crispant embryos.
Whole F0 embryos were lysed and sequenced via Miseq. From the total amplified fragment read list, we analyzed each sequence for the gRNA target sequence, and a sequence in the gene upstream of this sequence. The ratio of reads not containing the gRNA target sequence : upstream sequence was used to estimate deletion efficiency in each embryo. The average efficiency across all embryos was 0.523126.

Total Read Counts	gRNA Target Sequence counts	Upstream counts	Ratio deletion	Total Read Counts	gRNA Target Sequence counts	upstream counts	ratio deletion
121430	28531	105763	0.730236	39132	16573	35855	0.537777
149253	57379	125991	0.544579	52689	21538	45458	0.5262
142683	65454	135707	0.517681	36620	15601	33884	0.539576
6933	2654	6572	0.596166	76082	24689	54134	0.543928
29	16	27	0.407407	54384	18186	48226	0.622901
96	28	88	0.681818	49805	18642	48520	0.615787
159	39	121	0.677686	58958	28124	54468	0.48366
230704	79688	177077	0.549981	42133	23800	38878	0.387829
220365	79353	144424	0.450555	20695	7893	19756	0.600476
630	384	605	0.365289	30708	16594	26805	0.380936
8624	4561	8390	0.456377	29032	12119	26028	0.534386
6	0	4	1	33611	19045	30993	0.385506
334919	122576	294568	0.583879	34072	19565	32181	0.392033
215043	59756	202030	0.704222	190366	122093	175615	0.304769
296929	125182	281893	0.555924	190215	87958	168334	0.477479

Total Read Counts	gRNA Target Sequence counts	upstream counts	ratio deletion	Total Read Counts	gRNA Target Sequence counts	upstream counts	ratio deletion
77719	36573	73237	0.500621	180016	84825	157858	0.46265
135836	48487	132178	0.633169	67670	40924	66106	0.380934
123123	73367	118059	0.378556	91109	56378	81245	0.306074
114418	55413	110070	0.496566	53378	28682	49065	0.415429
159603	37706	146803	0.743152	10688	3682	9889	0.627667
40577	15170	36919	0.5891	30614	8388	24246	0.654046
14705	7612	12093	0.370545	43675	16558	38269	0.567326
32924	15739	30874	0.490218	39132	16573	35855	0.537777
41667	15911	26857	0.407566	52689	21538	45458	0.5262
36995	21559	34757	0.379722	36620	15601	33884	0.539576
63492	27830	62071	0.551642	76082	24689	54134	0.543928
49805	18642	48520	0.615787	54384	18186	48226	0.622901

4.5 Discussion

By leveraging a large, multicenter collaborative cohort comprising individuals with severe AIS, most of whom have had spinal fusion surgery, we identified a key role for extracellular matrix proteins (the "matrisome") in its etiology. The core matrisome is a curated set of 274 genes encoding extracellular matrix proteins, including glycoproteins (fibronectins, laminins, tenascins, thrombospondins, fibrillins, fibulins), collagens, and proteoglycans (Guo et al, 2018; Naba et al, 2016; Naba et al, 2012). The matrisome itself consists of more than 1000 genes including the core matrisome genes, as well as extracellular matrix associated proteins, secreted factors, and extracellular matrix regulators (Naba et al, 2012). Many core matrisome genes are responsible for connective tissue disorders that cause scoliosis. In prior smaller studies, we and others have shown that rare variants in some of these genes (i.e. *FBN1*) are risk factors for generalized AIS (Buchan et al, 2014; Haller et al, 2016; Wise et al, 2020). In the current study, we present strong genetic evidence linking these genes to severe AIS, specifically *fibronectin type III domain containing 1 (FNDC1)*.

FNDC1 is a compelling candidate gene for AIS- a nonsynonymous variant (rs62432291) in this gene is strongly associated with joint hypermobility as self-reported by 23andme participants (Pickrell et al, 2016). Children, adolescents, and females have a higher prevalence of joint hypermobility, which are risk factors for AIS (Hakim et al, 2004). Many connective tissue disorders associated with joint hypermobility also have a high incidence of scoliosis comorbidity (Cheng et al, 2015). Generalized joint hypermobility is also a risk factor for AIS (Czaprowski, 2014; Czaprowski et al, 2011). Other than its association with joint hypermobility and otitis media, relatively little is known about *FNDC1* except that it is ubiquitously expressed, with higher expression in skeletal muscle, bone, and spinal cord (Nagase et al, 2001; Pickrell et al, 2016; Van

Ingen et al, 2016). The *FNDC* gene family contains 9 members, with some evidence that *FNDC1* and *FNDC5* regulate bone mineral density (Colaianni et al, 2017; Singhal et al, 2014; Xiao et al, 2019). Specifically, *FNDC1* expression was elevated in an osteoporotic rat model, whereas rutin treatment suppressed *FNDC1* expression and increased bone mineral density (Xiao et al, 2019).

MicroCT analysis of *fndc1* nonsense zebrafish confirm its role in regulating bone mineral density (Figure 4.3). Several human conditions are associated with increased bone mineral density, including osteopetroses and sclerosing bone dysplasias which may also present with severe scoliosis (Gregson & Duncan, 2020; Laine et al, 2011). Bone mineral density is a function of calcium and phosphorus deposition in the form of HA (Almeida Paz & Bruno, 2006). Being an ECM protein, it is possible that *fndc1* plays a role in HA deposition.

The shape of vertebrae also appears to be affected in our *fndc1* nonsense mutants. A higher cross-sectional area in vertebrae, but not in bone area, indicates that the hollow centra is larger in our mutants. We also observed a higher standard deviation of bone thickness in homozygotes, indicating greater variability in the thickness of the walls of the vertebrae. Furthermore, we observed a higher pMOI in *fndc1*[sa23734] carriers- because this metric is a function of distribution of tissue, this also supports that the vertebral shape is altered in our mutants (Jepsen et al, 2003).

While the mechanism is not yet clear, the effect of *fndc1* mutation on vertebral composition and morphology supports its role in vertebral development. While we did not observe a gross phenotype in our zebrafish, it is possible that mutation of this gene acts as a risk factor in AIS onset as opposed to a direct cause. Indeed, this is consistent with our exome data analysis- we show that *FNDC1* variants function as risk factors contributing to a ~2-fold increased risk of severe AIS. As expected for risk factors for complex disease, *FNDC1* variants show incomplete penetrance and imperfect familial segregation (Supplemental Figure 4.1).

Further, missense mutations such as those in our exome analysis often have wildly different effects on phenotype and protein function than loss-of-function mutations (Whittle et al, 2020). In the future, the *fndc1* loss-of-function mutants may be a useful tool in identifying mutations that, when combined with other mutations, result in a deleterious effect on gross morphology. In addition, functional testing of reported AIS-associated *FNDC1* SNPs may be useful in directly addressing the mechanisms by which *FNDC1* mutations result in AIS, and which variants are pathogenic.

Despite inclusion of more than 1000 AIS patients in this multicenter study, rare variant burden analyses remain challenging due to their low frequency and the resultant low power to detect single gene associations. To mitigate the possibility of identifying spurious associations, we confirmed our rare variant burden analysis results by also comparing our case data to the gnomAD dataset, which provides a much larger control cohort (Karczewski et al, 2020). Our burden analysis was strengthened by our replication of the association of *FNDC1* rare variants with scoliosis in the Geisinger MyCode cohort. Once confident in the association of *FNDC1* and severe AIS, our zebrafish model confirmed its role in vertebrae morphology via various parameters. Overall, our results demonstrate a compelling association of AIS with alteration of *FNDC1,* which may contribute to severe scoliosis by regulating joint hypermobility, vertebral morphology, and bone mineral density.

4.6 Supplemental materials
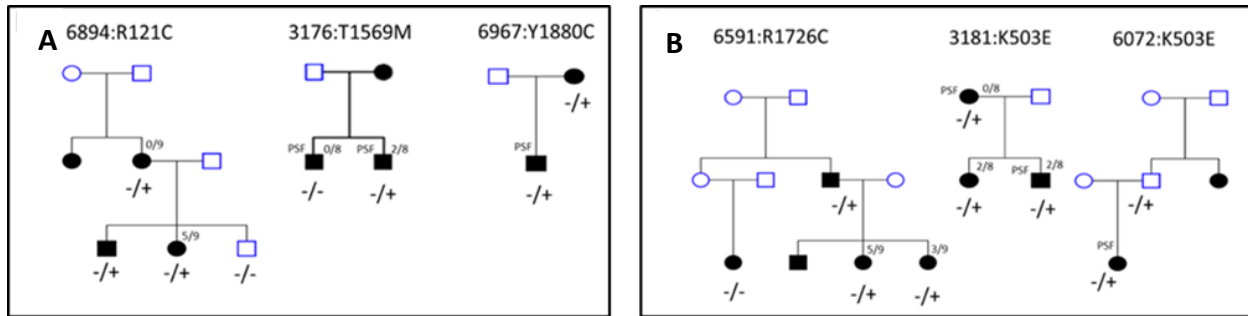
Supplemental Figure 4.1 *FNDC1* **ultra-rare nonsynoymous missense variants partially segregate with incomplete penetrance in AIS in families.** (A) Segregation of *FNDC1* ultra-rare missense variants in families with probands who were included in the gene burden analysis. (B) Segregation of *FNDC1* variants in additional AIS families that were not included in the gene burden analysis. Black is AIS.

Supplemental Table 4.1: Demographics of severe AIS cohort.

Characteristics		AIS (N=1,221)	Female AIS (N=1,022)	Male AIS (N=199)
Age	Mean (Median) ± S.D.	23.8 (16) ± 16.5	24.2 (16) ± 16.9	22.1 (17) ± 14.3
Cobb Angle	Mean (Median) ± S.D.	55.5° (55°) ± 13.0	55.3° (54°) ± 13.0	56.3° (55°) ± 13.2
Height (th percentile)	Mean (Median) ± S.D.	53th (55th) ± 29.6	53th (55th) ± 29.5	51th (48th) ± 30.6
Beighton	Mean (Median) ± S.D.	1.97 (2) ± 2.03	2.11 (2) ± 2.07	1.19 (0) ± 1.53
Beighton score ≤ 3 pts	Patients (%)	538 (44%)	449 (44%)	89 (45%)
Beighton score ≥ 4 pts	Patients (%)	164 (13%)	149 (15%)	15 (7%)

Supplemental Table 2: Demographics of severe AIS patients with *FNDC1* ultra-rare missense variants.

Characteristics		AIS (N=48)	Female AIS (N=38)	Male AIS (N=10)
Age	Mean (Median) ± S.D.	24.1 (16) ± 17.2	25.5 (16) ± 18.3	16.2 (16) ± 3.0
Cobb Angle	Mean (Median) ± S.D.	56.4° (50°) ± 15.5	55.9° (50°) ± 15.2	58.1° (63°) ± 17.2
Height (thpercentile)	Mean (Median) ± S.D.	50th (50th) ± 26.8	53th (54th) ± 26.5	35th (24th) ± 26.1
Beighton	Mean (Median) ± S.D.	1.77 (1) ± 2.16	2.04 (2) ± 2.24	0.4 (0) ± 0.09
Beighton score ≤ 3 pts	Patients (%)	24 (50%)	19 (50%)	5 (50%)
Beighton score ≥ 4 pts	Patients (%)	7 (15%)	7 (18%)	0

4.7 Materials and Methods

4.7.1 Cohort description

Severe AIS cases were recruited from a multicenter collaborative effort, including Washington University School of Medicine (WUSM; n=855), University of Wisconsin (UW; n=28), University of Colorado (UC; n=58) and Texas Scottish Rite Hospital (TX; n=280) (Supplementary Table 6)(n= number whose sequencing data passed QC). We included unrelated individuals of European ancestry with severe spinal curvature with a Cobb angle of at least 35 degrees, juvenile or adolescent onset, and excluded cases with congenital or infantile scoliosis, known or suspected underlying disorder or syndrome. The institutional review board at each institution approved this study and all patients and/or parents provided informed consent. In-house control subjects consisted of healthy individuals or patients ascertained for conditions other than scoliosis (i.e Alzheimer's disease, Amyotrophic Lateral Sclerosis, or male infertility).

4.7.2 Whole-exome sequencing, data processing and quality control

Exome libraries were prepared using Agilent's SureSelect Human All Exon kits V5 or IDT xGen Exome Panel V1 capture. Samples were sequenced at the McDonnell Genome Institute on Illumina HiSeq 2000/4000 with paired-end reads. Exome data from AIS cases is available at dbGaP (Study Accession: phs001677). Sequencing reads (i.e. fastq files) from all severe AIS case and in-house control samples were aligned to the human genome Reference (GRCh37) using BWA-MEM (v0.7.15). Aligned sequencing data (i.e. BAM files) were sorted and marked for duplication using Picard MarkDuplicates (v2.9.0). All case and in-house control BAM files were processed following GATK's (v3) Best Practices[47]. Variant calling was performed using HaplotypeCaller (GATKv3) individually, producing an intermediate vcf file (i.e. GVCF file) for

each WES sample. All GVCF files were combined using CombineGVCFs (GATKv3) and then variant joint calling were performed for all cases and in-house control samples using GenotypeGVCFs (GATKv3). Multiallelic variants were decomposed into biallelic variants. Variant Quality Score Recalibration (VQSR) was performed separately for SNPs and INDELs using VariantRecalibrator (GATKv3). Variants were evaluated and only SNPs above the confidence threshold of 99.7, as indicated by VQSR, were considered for further analysis. We removed all X-chromosome variants from further analysis due to the sex imbalance in our cases and controls. Variants within low complexity and repetitive regions were also removed. Additional hard filters implemented for variant genotype quality control comprise of variants with depth (DP) of at least 10 reads with genomic quality (GQ) of at least 20, and allele balance (AB) for heterozygous calls between 0.2 and 0.8. Samples with genotype call rate < 90% or whose genotype data indicated a sex discordant from the clinical database were excluded.

4.7.3 Variant annotation and statistical analysis

Gene and variant functional annotation was performed using VEP tool with GENCODE (v19), gnomAD exomes (r2.1.1), and CADD scores (v1.4) database (Frankish et al, 2019; Karczewski et al, 2020; McLaren et al, 2016; Rentzsch et al, 2019). Principal component analysis (PCA) was calculated to verify ancestry and restrict our analysis to only individuals of European ancestry (Supplementary Figure 1). Individual and familial relatedness was confirmed using kinship inference analysis KING and only unrelated individuals were kept for further analysis (Supplementary Figure 2) (Manichaikul et al, 2010). Single-variant association analysis was done using PLINK (v1.9). We used a minor allele frequency cutoff of 0.05 (MAF $\geq$ 0.05) and with genotype call rate $\geq$ 90% in cases and controls. We removed variants that were in deviation from Hardy Weinberg equilibrium (p $< 1 \times 10^{-6}$). Fisher Exact test with Lancaster's mid-p adjustment

(Plink –assoc fisher-midp) was used to estimate association between single variants and severe AIS. The established threshold of 5×10^{-8} was used to identify genome-wide significance. Quantile-quantile and Manhattan plots were created using R.

Associations at the gene level were tested using R and Rvtests Exact CMC model which collapse and combine variants, and then perform Fisher's Exact test. We focused on ultra-rare missense variants with minor allele count cutoff of 3 (MAC $\leq$3) and with genotype call rate of at least 90% in cases and in-house controls. For gene-based association analysis of severe AIS cases compared with gnomAD database as controls, using TRAPD methods (Guo et al, 2018). Pathway enrichment analysis was performed with Metascape (Zhou et al, 2019).

4.7.4 Geisinger cohort description and data analysis

Exome sequence data for MyCode participants generated through the Geisinger-Regeneron DiscovEHR collaboration, for which details of the sequencing methods are described in Staples et al, were evaluated to identify individuals harboring ultra-rare *FNDC1* missense variants (MAC $\leq$3 in gnomad exomes.r2.1.1 and MAC $\leq$ 3 in the DiscovEHR unrelated European subset) (Staples et al, 2018). Genomic variants were annotated using VEP (v100) and were required to meet the following thresholds: GQ $\geq$ 20, DP $\geq$ 10, AB $\geq$ 0.25 and alternate depth $\geq$ 3 (Van der Auwera et al, 2013). Additionally, we excluded variants located in tandem repeats, homopolymer runs, low complexity regions, and exon 1 (due to inconsistent coverage) (Krusche et al, 2019; Li, 2014). We further limited our analysis to 85,164 unrelated participants (to the second degree as computationally determined) of European ancestry computed using principal components calculated as previously described (Staples et al, 2018). After applying the aforementioned variant filters and removing individuals with unavailable EHR data, we identified 84,717 participants eligible for inclusion in the data analysis: 469 harboring ultra-rare *FNDC1* (transcript

ENST00000297267.14) missense variants and 84,248 without ultra-rare variants. Scoliosis phenotypes were determined by extracting relevant scoliosis diagnosis codes from the EHR as specified by an orthopedic surgeon (Supplementary Table 9). Diagnosis codes only counted towards a scoliosis phenotype if they were observed in two different EHR encounters. After meeting this criterium, only a single EHR occurrence of any additional scoliosis diagnosis codes was required to count towards a the category of "multiple scoliosis diagnoses." These filters were implemented to minimize the risk of including false positives due to ICD code use related to rule out procedures and testing. It is common practice to implement a minimum number of occurrences for ICD code usage when using this information to define phenotypes. Typically, individuals with a true diagnosis of a medical condition that requires continuous follow-up will continually have the relevant ICD codes documented in their EHR for the follow-up care visits. Individuals counted as having scoliosis before 18 years of age include those with a relevant diagnosis code used before the age of 18 or the use of any scoliosis code with the terms "infantile," "juvenile," or "adolescent" in the diagnosis name (Supplementary Table 9). This study was approved by the Geisinger Institutional Review Board and all participants provided informed consent.

4.7.5 Zebrafish husbandry

Zebrafish (*Danio rerio*) were raised and maintained using standard methods (Westerfield, 1995). Zebrafish were housed and handled using protocols approved by the Institutional Animal Care and Use Committee (IACUC) at Washington University. *fndc1^sa23734^* embryos were acquired through the Zebrafish International Research Center (ZIRC).

4.7.6 Zebrafish *fndc1^sa23734^* Genotyping

Fish were in-crossed and genotyped for *fndc1ˢᵃ²³⁷³⁴* via Sanger sequencing using the following primers:

Forward: `TAATTGGCCTTCCTGCACCTA`

Reverse: `GTATGCCTTCAGTTGCTTCACC`

4.7.7 MicroCT Preparation

Adult fish were fixed for 7 days at room temperature under agitation in 4% paraformaldehyde in PBS, and washed with 70% ethanol. Whole-animal micro-computed tomography (microCT) was conducted on a Scanco μCT-50 instrument. Scanning and data extraction were performed by the microCT core at Washington University in St Louis School of Medicine. Samples were mounted in 19-mm sample holders packed with Kimwipes, dampened with PBS to prevent drying.

Statistical analysis on zebrafish microCT data were performed in R, via Wilcoxon-rank sum test, two-sided for each genotype group (n=5). Graphs were generated in R.

4.7.8 RT-PCR

Initial RNA was extracted using the Qiagen RNeasy kit. Initial RT-PCR was performed using the extracted RNA, then previously described PCR performed on the resulting product.

4.7.9 Zebrafish CRISPR mutagenesis

CRISPR/Cas9 gRNA design and mutagenesis was performed using previously described methods (Hoshijima et al, 2019). We used a dual gRNA approach for this CRISPR protocol (dgRNA).

To generate gRNA-Cas9 complexes for target site 1, 10 uL of 100 uM Alt-R ® tracrRNA (IDT 1072534) was added to 10 uL 100 uM crRNA (IDT custom [/AltR1/rUrG rUrCrU rGrCrA rArUrG rArCrU rCrCrU rGrCrG rGrUrU rUrUrA rGrArG rCrUrA rUrGrC rU/AltR2/]). This was incubated at 95 ºC for 5 minutes, then cooled to 25 ºC at a rate of -0.1 ºC/second. It was kept at 25 ºC for 5 minutes, then stored at -20 ºC in 1 uL aliquots (50 uM). 1 uL dgRNA, 1 uL Cas9 protein (25 uM), 0.5 uL phenol red, and 2.5 uL KCl (300 mM), were mixed and incubated at 37 ºC for 5 minutes. Next, 1 uL NU7441 (400 mM) was added. This cocktail was injected into 1-cell stage zebrafish embryos.

Larvae were collected for DNA extraction at 5 dpf for CRISPR efficiency analysis. They were genotyped via MiSeq using the following primers:

Forward: CTAGAGAACTTCCGTCGCATTT

Reverse: AGAAACAGTTTGACCTACCCGA

Chapter 5: Conclusions and Future Directions

5.1 Perspective

Musculoskeletal disorders can severely impact the quality of life of affected patients. Skeletal muscle comprises about 40% of human body mass, and 50-75% of individual proteins (Frontera & Ochala, 2015). Alteration in the function or structure of its elements can drastically influence daily life. For example, activities as simple as walking can become painful and cumbersome for patients with congenital disorders of the musculoskeletal system (Kimber et al, 2012).

Many genes work in tandem to regulate the structure and function of muscle, joints, and bones. Genes in muscle, nerve tissue, and connective tissue can affect the musculoskeletal system. In addition, genes expressed extracellularly from any of these tissues can interact and impact the musculoskeletal system. Because these tissues are so interconnected, mutations in genes expressed in widely divergent tissues can produce similar phenotypes. For example, mutations in both muscle and nerve genes have been discovered as causative in Distal Arthrogryposis (DA) (Chong et al, 2020; Coste et al, 2013; Gurnett et al, 2010; McMillin et al, 2013; Toydemir et al, 2006; Whittle et al, 2021). Even within the same gene, different mutations can have differing effects on protein function (Del Angel et al, 2020). However, it may be difficult to predict how these divergent effects may affect gross morphology or tissue structure without studying humans with these variants, or functional testing of mutant protein in a model system.

Currently, musculoskeletal disorders are primarily treated conservatively with bracing or physical therapy, and occasionally with surgical intervention (Bamshad et al, 2009). This is a "one size fits all" approach to disorders that are caused by mutations in different genes with vastly

differing disease mechanisms. Alternatively, studying the functional effects of individual genes or mutations in the same gene will allow us to approach patients suffering from these disorders individually. This precision medicine-based approach will allow us to address the underlying issues in the proteins themselves instead of correcting secondary, downstream phenotypes.

Determining the functional effects of mutations often requires directly testing mutant protein function. While single-molecule studies are useful for determining molecular qualities of protein function, animal models allow us to study these functional effects on organ systems. In particular, animal models allow us to study how mutations affect cellular and tissue structure and function.

The objective of this thesis was to develop and test vertebrate models of two musculoskeletal disorders- DA and Adolescent Idiopathic Scoliosis (AIS). We leveraged the benefits of using zebrafish as vertebrate models, with the primary focusing being on the genes *smyhc1* and *fndc1*.

5.2 *smyhc1* zebrafish mutants as models for distal arthrogryposis

In Chapter 2, we describe a *smyhc1^{R673H}* missense zebrafish mutant. This mutation is a direct analog to the most common mutation *(MYH3^{R672H})* associated with DA2A (Freeman-Sheldon Syndrome) (Toydemir et al, 2006; Whittle et al, 2020). DA2A is considered the most severe form of DA, presenting with severe contractures in the hands, feet, and face. Our mutant fish phenocopied the human disease, with kinks in notochord being caused by muscle hypercontraction that developed into vertebral fusions and scoliosis. The phenotype severity increased with allele dosage, with homozygous lethality and severe body curvatures. This mutation

is inherited in an autosomal dominant manner in humans as well, with no homozygous cases of this allele having been described.

We characterized the *smyhc1^R673H* mutant by examining muscle structure, swim capacity, and skeletal structure. We demonstrate that *smyhc1^R673H* mutants displayed shortened sarcomeres and muscle fibers, as well as diminished swim capacity in larval stages and adulthood. This mutation affects both structure and function of the muscle, with non-autonomous effects on skeletal structure. We hypothesized that increased tension produced by the muscle led to these phenotypes.

To test the hypothesis that increased tension produced by the muscle led to these phenotypes, we used pharmacological inhibition of many of the pathways required for muscle contraction. First, neural anesthetic failed to rescue this phenotype, suggesting that this errant muscle function was constitutive and independent of neural stimulation. Indeed, intervention with para-aminoblebbistatin, which directly inhibits actin-myosin interaction through myosin ATPase inhibition, prevented body curvature and kinking of the notochord. In the future, specific pharmacologically inhibition of embryonic skeletal muscle contraction in developing human fetuses with this mutation may prevent DA from developing or progressing.

In Chapter 3, we describe a higher throughput method of *smyhc1* mutagenesis and screening using a GFP reporter which is an ongoing collaboration between the Gurnett and Johnson labs. While we were able to functionally examine the effects of the *MYH3^R672H* mutation using the zebrafish *smyhc1^R673H* model, many more variants have been reported in this gene in human patients. Many of these variants are of unknown significance, and without functional testing it is difficult to determine the pathogenicity of these mutations. Because variant mutagenesis is so time and labor intensive, a higher throughput method of variant induction is necessary. With our

novel approach, a GFP reporter was introduced at the C-terminus of the *smyhc1* gene. Removal of a region of interest, where many of the reported variants cluster, via CRISPR puts the GFP out of frame. Upon subsequent successful in-frame integration of a variant of interest in this region, GFP expression is restored, allowing for efficient screening of F0 embryos. This new screening method will be enable efficient identification of successful integration events, as it will use GFP as a visual cue instead of multiple rounds of high-throughput sequencing. Embryos with high GFP expression in many cells will be more likely to carry the variant in the germline, allowing for efficient screening of germline candidates as well. In addition, F0 fish can be analyzed, as the mononucleate cells expressing GFP will carry the integrated variant. Overall, this method will allow for more streamlined screening for successful mutagenesis, drastically reducing the time and labor in generating mutant zebrafish lines. A large number of reported *MYH3* variants can then be characterized in much less time.

5.3 *fndc1* mutant effect on Zebrafish skeletal structure, and the link between FNDC1 and AIS Progression

In Chapter 4, we describe the gene burden analysis that identified *FNDC1* as a significant genetic risk factor for AIS progression. *FNDC1* ultra-rare variants were significantly enriched in severe AIS patients compared to 2 control sets. To study the mechanism by which *FNDC1* variants may contribute to scoliosis pathogenesis, we characterized *fndc1* nonsense zebrafish mutants to examine the effect of this gene on skeletal morphology. Without observing a gross phenotype, we examined the vertebrae more closely using microCT. Here, we found that *fndc1* mutants had higher tissue mineral density in the vertebral bones with a differing vertebral shape, which was proportional to allele dosage. Alteration of tissue mineral density may be one of the mechanisms by which *FNDC1* variants contributes to AIS risk.

Consistent with the absence of a scoliosis phenotype in *fndc1* zebrafish null mutants, *FNDC1* variants did not perfectly segregate with AIS in our family pedigrees. This may suggest that *FNDC1* variants may act as risk factors for AIS, as opposed to a highly penetrant fully dominant pathogenic variants. To further test this, generation of models with multiple mutations may be used to determine the additive effects that result in AIS.

5.4 Conclusions

Overall, the findings presented here expand our understanding of *MYH3*-driven DA, and the potential role of *FNDC1* in severe AIS. These studies lay the groundwork for phenotype characterization of pathogenic skeletal muscle gene alleles in zebrafish, as well as a method for examining many variants of a gene in an efficient manner. Patient phenotypes are difficult to predict. Therefore, an understanding of gene function and how mutations affect that function will enable the development of predictive models for variant pathogenicity. Zebrafish are a useful tool in understanding the disease mechanism of musculoskeletal disorders, as well as exploring treatment methods that may be adapted for human patients. Our work may subsequently lead to precision medicine for patients affected by musculoskeletal disorders.

REFERENCES

Ablain J, Durand EM, Yang S, Zhou Y, Zon LI (2015) A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Developmental cell* **32:** 756-764

Almeida Paz I, Bruno L (2006) Bone mineral density. *Brazilian Journal of Poultry Science*
 8: 69-73

Alvarado DM, Buchan JG, Gurnett CA, Dobbs MB (2011) Exome sequencing identifies an MYH3 mutation in a family with distal arthrogryposis type 1. *The Journal of Bone and Joint Surgery American volume* **93:** 1045

Attili S, Hughes SM (2014) Anaesthetic tricaine acts preferentially on neural voltage-gated sodium channels and fails to block directly evoked muscle contraction. *PloS One* **9:** e103751

Bamshad M, Van Heest AE, Pleasure D (2009) Arthrogryposis: a review and update. *The Journal of Bone Joint Surgery American volume* **91:** 40

Beck AE, McMillin MJ, Gildersleeve HI, Kezele PR, Shively KM, Carey JC, Regnier M, Bamshad MJ (2013) Spectrum of mutations that cause distal arthrogryposis types 1 and 2B. *American Journal of Medical Genetics Part A* **161:** 550-555

Beck AE, McMillin MJ, Gildersleeve HI, Shively K, Tang A, Bamshad MJ (2014) Genotype-phenotype relationships in Freeman–Sheldon syndrome. *American Journal of Medical Genetics Part A* **164:** 2808-2813

Bessarab DA, Chong S-W, Srinivas BP, Korzh V (2008) Six1a is required for the onset of fast muscle differentiation in zebrafish. *Developmental biology*
 323: 216-228

Boehm S, Limpaphayom N, Alaee F, Sinclair MF, Dobbs MB (2008) Early results of the Ponseti method for the treatment of clubfoot in distal arthrogryposis. *The Journal of Bone and Joint Surgery* **90:** 1501-1507

Buchan JG, Alvarado DM, Haller GE, Cruchaga C, Harms MB, Zhang T, Willing MC, Grange DK, Braverman AC, Miller NH (2014) Rare variants in FBN1 and FBN2 are associated with severe adolescent idiopathic scoliosis. *Human molecular genetics*
 23: 5271-5282

Cameron-Christie SR, Wells CF, Simon M, Wessels M, Tang CZ, Wei W, Takei R, Aarts-Tesselaar C, Sandaradura S, Sillence DO *et al* (2018) Recessive Spondylocarpotarsal synostosis syndrome due to compound heterozygosity for variants in MYH3. *The American Journal of Human Genetics* **102:** 1115-1125

Cantor RM, Lange K, Sinsheimer JS (2010) Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *The American Journal of Human Genetics*
86: 6-22

Carapito R, Goldenberg A, Paul N, Pichot A, David A, Hamel A, Dumant-Forest C, Leroux J, Ory B, Isidor B *et al* (2016) Protein-altering MYH3 variants are associated with a spectrum of phenotypes extending to spondylocarpotarsal synostosis syndrome. *European Journal of Human Genetics* **24:** 1746-1751

Carman D, Browne R, Birch JJTJob (1990) Measurement of scoliosis and kyphosis radiographs. Intraobserver and interobserver variation. *The Journal of bone joint surgery American volume*
72: 328-333

Carraro U, Catani CJB (1983) A sensitive SDS-PAGE method separating myosin heavy chain isoforms of rat skeletal muscles reveals the heterogeneous nature of the embryonic myosin. *Biochemical biophysical research communications*
116: 793-802

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research* **39:** e82

Cheng JC, Castelein RM, Chu WC, Danielsson AJ, Dobbs MB, Grivas TB, Gurnett CA, Luk KD, Moreau A, Newton PO (2015) Adolescent idiopathic scoliosis. *Nature reviews disease primers*
1: 1-21

Cheng KC, Xin X, Clark DP, La Riviere P (2011) Whole-animal imaging, gene function, and the Zebrafish Phenome Project. *Current opinion in genetics development*
21: 620-629

Chong JX, Burrage LC, Beck AE, Marvin CT, McMillin MJ, Shively KM, Harrell TM, Buckingham KJ, Bacino CA, Jain M *et al* (2015) Autosomal-dominant multiple pterygium syndrome is caused by mutations in MYH3. *The American Journal of Human Genetics* **96:** 841-849

Chong JX, Talbot JC, Teets EM, Previs S, Martin BL, Shively KM, Marvin CT, Aylsworth AS, Saadeh-Haddad R, Schatz UA (2020) Mutations in MYLPF cause a novel segmental amyoplasia that manifests as distal arthrogryposis. *The American Journal of Human Genetics* **107:** 293-310

Cobb J (1948) Outline for the study of scoliosis. *Instr Course Lect AAOS* **5:** 261-275

Codina M, Li J, Gutierrez J, Kao JP, Du SJ (2010) Loss of Smyhc1 or Hsp90α1 function results in different effects on myofibril organization in skeletal muscles of zebrafish embryos. *PLoS One* **5:** e8416

Colaianni G, Notarnicola A, Sanesi L, Brunetti G, Lippo L, Celi M, Moretti L, Pesce V, Vicenti G, Moretti B (2017) Irisin levels correlate with bone mineral density in soccer players. *Journal of biological regulators homeostatic agents* **31:** 21-28

Coste B, Houge G, Murray MF, Stitziel N, Bandell M, Giovanni MA, Philippakis A, Hoischen A, Riemer G, Steen U (2013) Gain-of-function mutations in the mechanically activated ion channel PIEZO2 cause a subtype of Distal Arthrogryposis. *Proceedings of the National Academy of Sciences* **110:** 4667-4672

Czaprowski D (2014) Generalised joint hypermobility in caucasian girls with idiopathic scoliosis: relation with age, curve size, and curve pattern. *The Scientific World Journal* **2014**

Czaprowski D, Kotwicki T, Pawłowska P, Stoliński L (2011) Joint hypermobility in children with idiopathic scoliosis: SOSORT award 2011 winner. *Scoliosis* **6:** 1-10

Dalgin G, Prince VE (2015) Differential levels of Neurod establish zebrafish endocrine pancreas cell fates. *Developmental biology* **402:** 81-97

Dalgin G, Ward AB, Hao LT, Beattie CE, Nechiporuk A, Prince VE (2011) Zebrafish mnx1 controls cell fate choice in the developing endocrine pancreas. *Development* **138:** 4597-4608

Das S, Kumar P, Verma A, Maiti TK, Mathew SJ (2019) Myosin heavy chain mutations that cause Freeman-Sheldon syndrome lead to muscle structural and functional defects in Drosophila. *Developmental Biology* **449:** 90-98

Del Angel G, Reynders J, Negron C, Steinbrecher T, Mornet E (2020) Large-scale in vitro functional testing and novel variant scoring via protein modeling provide insights into alkaline phosphatase activity in hypophosphatasia. *Human mutation* **41:** 1250-1262

Desai D, Stiene D, Song T, Sadayappan S (2020) Distal Arthrogryposis and Lethal Congenital Contracture Syndrome–An Overview. *Frontiers in Physiology* **11:** 689

Devoto SH, Melançon E, Eisen JS, Westerfield M (1996) Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation. *Development* **122:** 3371-3380

Donnelly ML, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, Ryan MD (2001) Analysis of the aphthovirus 2A/2B polyprotein 'cleavage'mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *Journal of General Virology*
82: 1013-1025

Echwald SM, Bach H, Vestergaard H, Richelsen B, Kristensen K, Drivsholm T, Borch-Johnsen K, Hansen T, Pedersen O (2002) A P387L variant in protein tyrosine phosphatase-1B (PTP-1B) is associated with type 2 diabetes and impaired serine phosphorylation of PTP-1B in vitro. *Diabetes* **51:** 1-6

Ellis K, Bagwell J, Bagnat M (2013a) Notochord vacuoles are lysosome-related organelles that function in axis and spine morphogenesis. *Journal of Cell Biology* **200:** 667-679

Ellis K, Hoffman BD, Bagnat M (2013b) The vacuole within: how cellular organization dictates notochord function. *Bioarchitecture* **3:** 64-68

Elworthy S, Hargrave M, Knight R, Mebus K, Ingham PW (2008) Expression of multiple slow myosin heavy chain genes reveals a diversity of zebrafish slow twitch muscle fibres with differing requirements for Hedgehog and Prdm1 activity. *Development* **135:** 2115-2126

Felsenthal N, Zelzer E (2017) Mechanical regulation of musculoskeletal system development. *Development* **144:** 4271-4283

Fitts RH (2008) The cross-bridge cycle and skeletal muscle fatigue. *Journal of Applied Physiology* **104:** 551-558

Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research*
47: D766-D773

Frontera WR, Ochala J (2015) Skeletal muscle: a brief review of structure and function. *Calcified tissue international*
96: 183-195

Gil-Gálvez A, Carbonell-Corvillo P, Paradas C, Miranda-Vizuete A (2020) Cautionary note on the use of Caenorhabditis elegans to study muscle phenotypes caused by mutations in the human MYH7 gene. *BioTechniques* **68:** 296-299

Gray RS, Wilm TP, Smith J, Bagnat M, Dale RM, Topczewski J, Johnson SL, Solnica-Krezel L (2014) Loss of col8a1a function during zebrafish embryogenesis results in congenital vertebral malformations. *Developmental Biology* **386:** 72-85

Green EM, Wakimoto H, Anderson RL, Evanchik MJ, Gorham JM, Harrison BC, Henze M, Kawas R, Oslob JD, Rodriguez HM *et al* (2016) A small-molecule inhibitor of sarcomere contractility suppresses hypertrophic cardiomyopathy in mice. *Science* **351:** 617-621

Gregson CL, Duncan EL (2020) The genetic architecture of high bone mass. *Frontiers in Endocrinology* **11:** 595653

Gross MK, Moran-Rivard L, Velasquez T, Nakatsu MN, Jagla K, Goulding M (2000) Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. *Development* **127:** 413-424

Guo MH, Plummer L, Chan Y-M, Hirschhorn JN, Lippincott MF (2018) Burden testing of rare variants identified through exome sequencing via publicly available control data. *The American Journal of Human Genetics* **103:** 522-534

Guo Y, Kronert WA, Hsu KH, Huang A, Sarsoza F, Bell KM, Suggs JA, Swank DM, Bernstein SI (2020) Drosophila myosin mutants model the disparate severity of type 1 and type 2B distal arthrogryposis and indicate an enhanced actin affinity mechanism. *Skeletal muscle* **10:** 1-18

Gurnett CA, Desruisseau DM, McCall K, Choi R, Meyer ZI, Talerico M, Miller SE, Ju J-S, Pestronk A, Connolly AM (2010) Myosin binding protein C1: a novel gene for autosomal dominant distal arthrogryposis type 1. *Human molecular genetics* **19:** 1165-1173

Hakim AJ, Cherkas LF, Grahame R, Spector TD, MacGregor AJJA, Rheumatology ROJotACo (2004) The genetic epidemiology of joint hypermobility: a population study of female twins. *Arthritis Rheumatism: Official Journal of the American College of Rheumatology*
 50: 2640-2644

Hall JG (1997) Arthrogryposis multiplex congenita: etiology, genetics, classification, diagnostic approach, and general aspects. *Journal of pediatric orthopedics* **6:** 159-166

Hall JG (2014) Arthrogryposis (multiple congenital contractures): diagnostic approach to etiology, classification, genetics, and general principles. *European journal of medical genetics*
 57: 464-472

Haller G, Alvarado D, Mccall K, Yang P, Cruchaga C, Harms M, Goate A, Willing M, Morcuende JA, Baschal E (2016) A polygenic burden of rare variants across extracellular matrix genes among individuals with adolescent idiopathic scoliosis. *Human molecular genetics* **25:** 202-209

Haller G, McCall K, Jenkitkasemwong S, Sadler B, Antunes L, Nikolov M, Whittle J, Upshaw Z, Shin J, Baschal E (2018) A missense variant in SLC39A8 is associated with severe idiopathic scoliosis. *Nature communications* **9:** 1-7

Han S-Y, Kato H, Kato S, Suzuki T, Shibata H, Ishii S, Shiiba K-i, Matsuno S, Kanamaru R, Ishioka C (2000) Functional evaluation of PTEN missense mutations using in vitro phosphoinositide phosphatase assay. *Cancer research* **60:** 3147-3151

Hirata H, Saint-Amant L, Waterbury J, Cui W, Zhou W, Li Q, Goldman D, Granato M, Kuwada JY (2004) accordion, a zebrafish behavioral mutant, has a muscle relaxation defect due to a mutation in the ATPase Ca2+ pump SERCA1. *Development* **131:** 5457-5468

Hoshijima K, Jurynec MJ, Shaw DK, Jacobi AM, Behlke MA, Grunwald DJ (2019) Highly efficient CRISPR-Cas9-based methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish. *Developmental cell* **51:** 645-657. e644

Hromowyk KJ, Talbot JC, Martin BL, Janssen PM, Amacher SL (2020) Cell fusion is differentially regulated in zebrafish post-embryonic slow and fast muscle. *Developmental biology* **462:** 85-100

Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D (2016) REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics* **99:** 877-885

Janicki JA, Alman BJP (2007) Scoliosis: Review of diagnosis and treatment. *Paediatrics child health* **12:** 771-776

Jepsen KJ, Akkus OJ, Majeska RJ, Nadeau JH (2003) Hierarchical relationship between bone traits and mechanical properties in inbred mice. *Mammalian Genome* **14:** 97-104

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581:** 434-443

Karsch-Mizrachi I, Travis M, Blau H, Leinwand LA (1989) Expression and DNA sequence analysis of a human embryonic skeletal muscle mvosin heavy chain gene. *Nucleic acids research* **17:** 6167-6179

Kiefer J, Hall JG (2019) Gene ontology analysis of arthrogryposis (multiple congenital contractures). In American Journal of Medical Genetics Part C: Seminars in Medical Genetics pp 310-326. Wiley Online Library

Kimber E, Tajsharghi H, Kroksmark AK, Oldfors A, Tulinius M (2012) Distal arthrogryposis: clinical and genetic findings. *Acta paediatrica* **101:** 877-887

Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* **46:** 310-315

Kokel D, Peterson RT (2011) Using the zebrafish photomotor response for psychotropic drug screening. In Methods in Cell Biology pp 517-524. Elsevier

Konieczny MR, Senyurt H, Krauspe R (2013) Epidemiology of adolescent idiopathic scoliosis. *Journal of children's orthopaedics* **7:** 3-9

Kou I, Otomo N, Takeda K, Momozawa Y, Lu H-F, Kubo M, Kamatani Y, Ogura Y, Takahashi Y, Nakajima M (2019) Genome-wide association study identifies 14 previously unreported susceptibility loci for adolescent idiopathic scoliosis in Japanese. *Nature communications* **10:** 1-9

Kou I, Takahashi Y, Johnson TA, Takahashi A, Guo L, Dai J, Qiu X, Sharma S, Takimoto A, Ogura Y (2013) Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nature genetics* **45:** 676-679

Kovács M, Tóth J, Hetényi C, Málnási-Csizmadia A, Sellers JR (2004) Mechanism of blebbistatin inhibition of myosin II. *Journal of Biological Chemistry* **279:** 35557-35563

Krusche P, Trigg L, Boutros PC, Mason CE, De La Vega FM, Moore BL, Gonzalez-Porta M, Eberle MA, Tezak Z, Lababidi S (2019) Best practices for benchmarking germline small-variant calls in human genomes. *Nature biotechnology* **37:** 555-560

Laine CM, Chung B-D, Susic M, Prescott T, Semler O, Fiskerstrand T, d'Eufemia P, Castori M, Pekkinen M, Sochett E (2011) Novel mutations affecting LRP5 splicing in patients with osteoporosis-pseudoglioma syndrome (OPPG). *European Journal of Human Genetics* **19:** 875-881

Li H (2014) Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30:** 2843-2851

Li S, Wen H, Du S (2020) Defective sarcomere organization and reduced larval locomotion and fish survival in slow muscle heavy chain 1 (smyhc1) mutants. *The FASEB Journal* **34:** 1378-1397

Lonstein JE (1994) Adolescent idiopathic scoliosis. *The Lancet* **344:** 1407-1412

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* **26:** 2867-2873

Martin CT, Pugely AJ, Gao Y, Mendoza-Lattes SA, Ilgenfritz RM, Callaghan JJ, Weinstein SL (2014) Increasing hospital charges for adolescent idiopathic scoliosis in the United States. *Spine* **39:** 1676-1682

Martin RB, Burr DB (1984) Non-invasive measurement of long bone cross-sectional moment of inertia by photon absorptiometry. *Journal of biomechanics* **17:** 195-201

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham FJGb (2016) The ensembl variant effect predictor. *Genome biology* **17:** 1-14

McMillin MJ, Below JE, Shively KM, Beck AE, Gildersleeve HI, Pinner J, Gogola GR, Hecht JT, Grange DK, Harris DJ (2013) Mutations in ECEL1 cause distal arthrogryposis type 5D. *The American Journal of Human Genetics* **92:** 150-156

Mei K, Schwaiger BJ, Kopp FK, Ehn S, Gersing AS, Kirschke JS, Muenzel D, Fingerle AA, Rummeny EJ, Pfeiffer F (2017) Bone mineral density measurements in vertebral specimens and phantoms using dual-layer spectral computed tomography. *Scientific reports* **7:** 1-10

Miller NHJOC (1999) Cause and natural history of adolescent idiopathic scoliosis. *Orthopedic Clinics* **30:** 343-352

Minskaia E, Nicholson J, Ryan MD (2013) Optimisation of the foot-and-mouth disease virus 2A co-expression system for biomedical applications. *BMC biotechnology* **13:** 1-11

Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO (2016) The extracellular matrix: Tools and insights for the "omics" era. *Matrix Biology* **49:** 10-24

Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO (2012) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular Cellular Proteomics* **11**

Naba A, Pearce OM, Del Rosario A, Ma D, Ding H, Rajeeve V, Cutillas PR, Balkwill FR, Hynes RO (2017) Characterization of the extracellular matrix of normal and diseased tissues using proteomics. *Journal of proteome research* **16:** 3083-3091

Nagase T, Nakayama M, Nakajima D, Kikuno R, Ohara O (2001) Prediction of the coding sequences of unidentified human genes. XX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA research* **8:** 85-95

O'Keefe JH, Bergman N, Carrera-Bastos P, Fontes-Villalba M, DiNicolantonio JJ, Cordain L (2016) Nutritional strategies for skeletal and cardiovascular health: hard bones, soft arteries, rather than vice versa. *Open Heart* **3:** e000325

Ogura Y, Kou I, Miura S, Takahashi A, Xu L, Takeda K, Takahashi Y, Kono K, Kawakami N, Uno K (2015) A functional SNP in BNC2 is associated with adolescent idiopathic scoliosis. *The American Journal of Human Genetics* **97:** 337-342

Oshikane H, Watabe M, Nakaki TJPE (2018) Facilitation of yeast-lethal membrane protein production by detoxifying with GFP tagging. *Protein Expression Purification* **148:** 40-45

Park JH, Hogrebe M, Grüneberg M, DuChesne I, Ava L, Reunert J, Schlingmann KP, Boycott KM, Beaulieu CL, Mhanni AA (2015) SLC39A8 deficiency: a disorder of manganese transport and glycosylation. *The American Journal of Human Genetics* **97:** 894-903

Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA (2016) Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics* **48:** 709-717

Racca AW, Beck AE, McMillin MJ, Korte FS, Bamshad MJ, Regnier M (2015) The embryonic myosin R672C mutation that underlies Freeman-Sheldon syndrome impairs cross-bridge detachment and cycling in adult skeletal muscle. *Human molecular genetics* **24:** 3348-3358

Racca AW, Beck AE, Rao VS, Flint GV, Lundy SD, Born DE, Bamshad MJ, Regnier M (2013) Contractility and kinetics of human fetal and human adult skeletal muscle. *The Journal of physiology* **591:** 3049-3061

Rao DS, Kronert WA, Guo Y, Hsu KH, Sarsoza F, Bernstein SI (2019) Reductions in ATPase activity, actin sliding velocity, and myofibril stability yield muscle dysfunction in Drosophila models of myosin-based Freeman–Sheldon syndrome. *Molecular Biology of the Cell* **30:** 30-41

Rauch G, Lyons D, Middendorf I, Friedlander B, Arana N, Reyes T, Talbot W, Friedlander B, Reyes-Robles T, B M (2003) Submission and curation of gene expression data**. *ZFIN Direct Data Submission***

Rauscher AÁ, Gyimesi M, Kovács M, Málnási-Csizmadia A (2018) Targeting myosin by blebbistatin derivatives: optimization and pharmacological potential**. *Trends in Biochemical Sciences* 43:** 700-713

Ravenscroft G, Clayton JS, Faiz F, Sivadorai P, Milnes D, Cincotta R, Moon P, Kamien B, Edwards M, Delatycki M (2020) Neurogenetic fetal akinesia and arthrogryposis: genetics, expanding genotype-phenotypes and functional genomics**. *Journal of medical genetics***

Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M (2019) CADD: predicting the deleteriousness of variants throughout the human genome**. *Nucleic acids research* 47:** D886-D894

Risch NJ (2000) Searching for genetic determinants in the new millennium**. *Nature* 405:** 847-856

Roy S, Wolff C, Ingham PW (2001) The u-boot mutation identifies a Hedgehog-regulated myogenic switch for fiber-type diversification in the zebrafish embryo**. *Genes & Development* 15:** 1563-1576

Ryan MD, King AM, Thomas GP (1991) Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence**. *Journal of General Virology* 72:** 2727-2732

Saint-Amant L, Drapeau P (1998) Time course of the development of motor behaviors in the zebrafish embryo**. *Journal of Neurobiology* 37:** 622-632

Scala M, Accogli A, De Grandis E, Allegri A, Bagowski CP, Shoukier M, Maghnie M, Capra V (2018) A novel pathogenic MYH3 mutation in a child with Sheldon–Hall syndrome and vertebral fusions**. *American Journal of Medical Genetics Part A* 176:** 663-667

Schiaffino S, Reggiani C (1996) Molecular diversity of myofibrillar proteins: gene regulation and functional significance**. *Physiological Reviews* 76:** 371-423

Schiaffino S, Rossi AC, Smerdu V, Leinwand LA, Reggiani C (2015) Developmental myosins: expression patterns and functional significance**. *Skeletal Muscle* 5:** 1-14

Schoenau E, Neu C, Rauch F, Manz F (2001) The development of bone strength at the proximal radius during childhood and adolescence**. *The Journal of Clinical Endocrinology Metabolism***

86: 613-618

Sharma S, Londono D, Eckalbar WL, Gao X, Zhang D, Mauldin K, Kou I, Takahashi A, Matsumoto M, Kamiya N (2015) A PAX1 enhancer locus is associated with susceptibility to idiopathic scoliosis in females**. *Nature communications* 6:** 1-10

Shi J, Bi P, Pei J, Li H, Grishin NV, Bassel-Duby R, Chen EH, Olson EN (2017) Requirement of the fusogenic micropeptide myomixer for muscle formation in zebrafish**. *Proceedings of the National Academy of Sciences* 114:** 11950-11955

Sieck GC, Prakash Y, Han Y-S, Fang Y-H, Geiger PC, Zhan W-ZJJoap (2003) Changes in actomyosin ATP consumption rate in rat diaphragm muscle fibers during postnatal development**. *Journal of applied physiology* 94:** 1896-1902

Silva S, Alves A, Patel D, Malhó R, Soutar A, Bourbon M (2012) In vitro functional characterization of missense mutations in the LDLR gene**. *Atherosclerosis* 225:** 128-134

Singhal V, Lawson EA, Ackerman KE, Fazeli PK, Clarke H, Lee H, Eddy K, Marengi DA, Derrico NP, Bouxsein ML (2014) Irisin levels are lower in young amenorrheic athletes compared with eumenorrheic athletes and non-athletes and are associated with bone density and strength estimates**. *PloS one* 9:** e100218

Skube SB, Chaverri JM, Goodson HV (2010) Effect of GFP tags on the localization of EB1 and EB1 fragments in vivo**. *Cytoskeleton* 67:** 1-12

Smerdu V (2002) Expression of myosin heavy chain transcripts within distinct types of human muscle fibres**. *Slovenia: University of Ljubljana***

Sorrells S, Toruno C, Stewart RA, Jette C (2013) Analysis of apoptosis in zebrafish embryos by whole-mount immunofluorescence to detect activated Caspase 3**. *JoVE*:** e51060

Staples J, Maxwell EK, Gosalia N, Gonzaga-Jauregui C, Snyder C, Hawes A, Penn J, Ulloa R, Bai X, Lopez AE (2018) Profiling and leveraging relatedness in a precision medicine cohort of 92,455 exomes**. *The American Journal of Human Genetics* 102:** 874-889

Stefl S, Nishi H, Petukh M, Panchenko AR, Alexov E (2013) Molecular mechanisms of disease-causing missense mutations**. *Journal of molecular biology* 425:** 3919-3936

Stevenson DA, Carey JC, Palumbos J, Rutherford A, Dolcourt J, Bamshad MJ (2006) Clinical characteristics and natural history of Freeman-Sheldon syndrome**. *Pediatrics* 117:** 754-762

Tajsharghi H, Kimber E, Kroksmark A-K, Jerre R, Tulinius M, Oldfors A (2008) Embryonic myosin heavy-chain mutations cause distal arthrogryposis and developmental myosin myopathy that persists postnatally. *Archives of neurology* **65:** 1083-1090

Takahashi Y, Kou I, Takahashi A, Johnson TA, Kono K, Kawakami N, Uno K, Ito M, Minami S, Yanagida H (2011) A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. *Nature genetics* **43:** 1237-1240

Thusberg J, Olatubosun A, Vihinen M (2011) Performance of mutation pathogenicity prediction methods on missense variants. *Human mutation* **32:** 358-368

Toydemir RM, Rutherford A, Whitby FG, Jorde LB, Carey JC, Bamshad MJ (2006) Mutations in embryonic myosin heavy chain (MYH3) cause Freeman-Sheldon syndrome and
Sheldon-Hall syndrome. *Nature genetics* **38:** 561-566

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics* **43:** 11.10. 11-11.10. 33

van der Sijde MR, Ng A, Fu J (2014) Systems genetics: From GWAS to disease pathways. *Biochimica et Biophysica Acta -Molecular Basis of Disease* **1842:** 1903-1909

Van Ingen G, Li J, Goedegebure A, Pandey R, Li YR, March ME, Jaddoe VW, Bakay M, Mentch FD, Thomas K (2016) Genome-wide association study for acute otitis media in children identifies FNDC1 as disease contributing gene. *Nature communications* **7:** 1-7

Várkuti BH, Képiró M, Horváth IÁ, Végner L, Ráti S, Zsigmond Á, Hegyi G, Lenkei Z, Varga M, Málnási-Csizmadia A (2016) A highly soluble, non-phototoxic, non-fluorescent blebbistatin derivative. *Scientific Reports* **6:** 26141

Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. *The American Journal of Human Genetics* **90:** 7-24

Walklate J, Vera C, Bloemink MJ, Geeves MA, Leinwand L (2016) The most prevalent Freeman-Sheldon Syndrome mutations in the embryonic myosin motor share functional defects. *Journal of Biological Chemistry* **291:** 10318-10331

Wang X, Chong M, Wang X, Wang H, Zhang J, Xu H, Zhang J, Liu D (2015) Block the function of nonmuscle myosin II by blebbistatin induces zebrafish embryo cardia bifida. *In Vitro Cellular* **51:** 211-217

Weill U, Krieger G, Avihou Z, Milo R, Schuldiner M, Davidi D (2019) Assessment of GFP tag position on protein localization and growth fitness in yeast. *Journal of molecular biology*
 431: 636-641

Wergedal JE, Veskovic K, Hellan M, Nyght C, Balemans W, Libanati C, Vanhoenacker FM, Tan J, Baylink DJ, Van Hul W (2003) Patients with Van Buchem disease, an osteosclerotic genetic disease, have elevated bone formation markers, higher bone density, and greater derived polar moment of inertia than normal. *The Journal of Clinical Endocrinology Metabolism*
 88: 5778-5783

Westerfield M (1995) The zebrafish book: a guide for the laboratory use of zebrafish (Brachydanio rerio): University of Oregon press

Whittle J, Antunes L, Harris M, Upshaw Z, Sepich DS, Johnson AN, Mokalled M, Solnica-Krezel L, Dobbs MB, Gurnett CA (2020) MYH 3-associated distal arthrgryposis zebrafish model is normalized with para-aminoblebbistatin. *EMBO molecular medicine*
 12: e12356

Whittle J, Johnson A, Dobbs MB, Gurnett CA (2021) Models of distal arthrgryposis and lethal congenital contracture syndrome. *Genes*
 12: 943

Wieczorek DF, Periasamy M, Butler-Browne GS, Whalen RG, Nadal-Ginard B (1985) Co-expression of multiple myosin heavy chain genes, in addition to a tissue-specific one, in extraocular musculature. *The Journal of cell biology* **101:** 618-629

Wise CA, Sepich D, Ushiki A, Khanshour AM, Kidane YH, Makki N, Gurnett CA, Gray RS, Rios JJ, Ahituv N (2020) The cartilage matrisome in adolescent idiopathic scoliosis. *Bone research* **8:** 1-13

Wynne-Davies R (1968) Familial (idiopathic) scoliosis: a family survey. *The Journal of bone joint surgery British volume*
 50: 24-30

Xiao Y, Wei R, Yuan Z, Lan X, Kuang J, Hu D, Song Y, Luo J (2019) Rutin suppresses FNDC1 expression in bone marrow mesenchymal stem cells to inhibit postmenopausal osteoporosis. *American Journal of Translational Research*
 11: 6680

Yip YL, Scheib H, Diemand AV, Gattiker A, Famiglietti LM, Gasteiger E, Bairoch A (2004) The Swiss-Prot variant page and the ModSNP database: a resource for sequence and structure information on human protein variants. *Human mutation* **23:** 464-470

Zhang J, Zhao J, Jiang W-j, Shan X-w, Yang X-m, Gao J-g (2012a) Conditional gene manipulation: Cre-ating a new biological era. *Journal of Zhejiang University Science B* **13:** 511-524

Zhang Z, Miteva MA, Wang L, Alexov E (2012b) Analyzing effects of naturally occurring missense mutations. *Computational mathematical methods in medicine* **2012**

Zhao S, Zhang Y, Hallgrimsdottir S, Zuo Y, Li X, Batkovskyte D, Liu S, Lindelöf H, Wang S, Hammarsjö A (2022) Expanding the mutation and phenotype spectrum of MYH3-associated skeletal disorders. *NPJ genomic medicine* **7:** 1-11

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications* **10:** 1-10

Zhuo C, Zhang J, Lee J-H, Jiao J, Cheng D, Liu L, Kim H-W, Tao Y, Li M (2021) Spatiotemporal control of CRISPR/Cas9 gene editing. *Signal Transduction Targeted Therapy* **6:** 1-18

Zieba J, Zhang W, Chong JX, Forlenza KN, Martin JH, Heard K, Grange DK, Butler MG, Kleefstra T, Lachman RS *et al* (2017) A postnatal role for embryonic myosin revealed by MYH3 mutations that alter TGFβ signaling and cause autosomal dominant spondylocarpotarsal synostosis. *Scientific Reports* **7:** 41803

Julia N. Whittle

4917 McPherson Ave, St Louis, MO, 63108

(845)551-9908

juliawhittle@wustl.edu

EDUCATION

Ph.D. Washington University in St Louis; St Louis, MO   2016-Anticipated 2022
    Division of Biology and Biomedical Sciences
    Developmental, Regenerative, and Stem Cell Biology Program
    Laboratory of: Christina Gurnett, Department of Neurology
    Thesis: "Modeling Developmental Musculoskeletal Disorders in Zebrafish"

B.S. State University of New York College at Geneseo; Geneseo, NY   2012-2016
    Major: Biology
    Minor: Biomathematics

RESEARCH EXPERIENCE

Washington University in St Louis, Department of Neurology; St Louis, MO   2017-present
    Graduate Student, Laboratory of Christina Gurnett, M.D., Ph.D.
        - Developed and characterized zebrafish model for Distal Arthrogryposis,
          elucidating mechanism of disease and exploring treatment options
        - Developing higher-throughput method of variant mutagenesis in zebrafish
        - Mentored undergraduate and high school students in laboratory techniques
          and data analysis

State University of New York College at Geneseo, Department of Biology; Geneseo, NY   2013-2016
    Undergraduate Researcher, Laboratory of Travis Bailey, Ph.D.
        - Identified and sequenced splicing variants of mutant allele of *chaf1b* in
          zebrafish
        - Practiced zebrafish husbandry and line maintenence

PUBLICATIONS

Antunes, L., Haller, G., **Whittle, J.**, Charng, W., Khanshour, A.M., Morcuende, J.A.,
    Giampietro, P., Raggio,C., Hadley-Miller, N., Wise, C., Dobbs, M.B., Gurnett, C.A.
    "Increased burden of ultra-rare missense variants in core matrisome genes are
    associated with severe adolescent idiopathic scoliosis" *In preparation

Whittle, J, Johnson, A., Dobbs, M.B., and Gurnett, C.A. "Models of Distal
    Arthrogryposis and Lethal Congenital Contracture Syndrome." Genes 12,
    no. 6 (2021): 943.

Whittle, J., Antunes, L., Harris, M., Upshaw, Z., Sepich, D.S.,
    Johnson, A., Mokalled, M., Solnica-Krezel, L., Dobbs, M.B., and Gurnett, C.A.
    "MYH 3-associated distal arthrogryposis zebrafish model is normalized
    with para-aminoblebbistatin." EMBO molecular medicine 12, no. 11 (2020): e12356.

Haller, G., McCall, K., Jenkitkasemwong, S., Sadler, S., Antunes, L.,
    Nikolov, M., **Whittle, J.,** Upshaw, Z., Shin, J., Baschal, E. and Cruchaga, C.
    "A missense variant in SLC39A8 is associated with severe idiopathic scoliosis."
    Nature communications 9, no. 1 (2018): 1-7.

GRANTS

CMB T32 Training Grant, National Institute of Health   2017-2019

AWARDS

Rita Levi-Montalcini Prize for Outstanding Student Presentation   September 2021
  Joint Retreat of the Developmental, Regenerative, & Stem Cell Biology Program,
  The Department of Developmental Biology, and the Center for Regenerative
  Medicine

PRESENTATIONS

17th International Zebrafish Society Conference   June 22-26, 2022
  "The effect of adolescent idiopathic scoliosis (AIS)-associated gene FNDC1
    on zebrafish vertebrae morphology"
  Poster
  Centre Mont-Royal, Montreal, Quebec, Canada

Structural Birth Defects Trainee Symposium
  "FNDC1 rare variants are associated with severe adolescent idiopathic scoliosis."   2 November 2021
  Talk
  Online Conference

PROFESSIONAL ASSOCIATIONS

Beta Beta Beta Biology Honor Society, Lambda Kappa Chapter   2015-present

MENTEE RECORD

Washington University in St Louis Undergraduate Students
  Isabel Shrestha   2021-present
  Addie Avery   2020-2021
  Mya Harris   2017-2018
  Zachasry Upshaw   2017-2019

Washington University in St Louis Young Scientist Program High School Students
  Kaylee Green   2019

REFERENCES

Christina Gurnett, M.D., Ph.D.

    Professor of Neurology, Pediatrics; Washington University in St Louis

    A. Ernest and Jane G. Stein Professor of Developmental Neurology Director, Division of Pediatric and Developmental Neurology; Washington University in St Louis

    Neurologist-in-Chief; St Louis Children's Hospital

    (314) 454-6120

    gurnettc@wustl.edu

Liliana Solnica-Krezel

    Alan A and Edith L Wolff Distinguished Professor and Head of Developmental Biology, Department of Developmental Biology; Washington University in St Louis

    (314) 362-8768

    solnical@wustl.edu

Diane Sepich, Ph.D.

    Assistant Professor of Developmental Biology, Department of Developmental Biology; Washington University in St Louis

    (314) 362-9370

    sepichd@wustl.edu

Aaron Johnson, Ph.D.

    Assistant Professor of Developmental Biology, Department of Developmental Biology; Washington University in St Louis

    (314) 273-1834

    anjohnson@wustl.edu