Towards a better understanding of the mechanisms underlying myosin-related congenital myopathies

Hau, Abbi

Awarding institution:
King's College London

Towards a better understanding of the mechanisms underlying myosin-related congenital myopathies

Thesis for the Degree of Doctor of Philosophy

Hoi Ting Abbi Hau

Supervisors:

Dr Julien Ochala (Primary Supervisor)

Prof. Simon M Hughes (Secondary Supervisor)

Submission Date: 24th May 2022

Pages: 203 Words: 56,187

Centre of Human and Applied Physiological Sciences

King's College London, UK

Table Of Contents

Figures

Tables

Abstract

Myosin is a family of proteins that plays a crucial role in generating force and motion by interacting with actin filaments in skeletal muscle. Myosin molecules notably contain heavy chain (MyHC) isoforms that have different functional capabilities. Mutations in one of its isoforms, MyHC I/β (encoded by the *MYH7* gene) have been reported in humans, associated with muscle weakness and have led to two main distinct skeletal muscle diseases, Laing Distal Myopathy (LDM) and Myosin Storage Myopathy (MSM). The pathophysiological mechanisms by which subtle amino acid changes in the LMM region of MyHC I/β molecules leading to such variable skeletal muscle phenotypes in LDM and MSM patients remain poorly understood. Using a wide range of human *MYH7* patient muscle biopsy samples, investigation of primary biophysical defects in the presence of defective MyHC I/β molecules including myosin filament length has revealed no change but rather a shift in myosin head positioning into a disordered relaxed state (DRX). On the road to generating a zebrafish LDM and MSM disease model, several genes were identified to be orthologous to human *MYH7*. Amongst orthologous genes, *smyhc1* was targeted for genome editing using CRISPR/Cas9 to generate a loss of function model. Loss of *smyhc1* led to early developmental defects, however, continued to grow to adulthood with no observable muscle defects. *Smyhc1* null zebrafish are replaced and compensated by *smyhc2* and *smyhc3* in adult zebrafish. Work ongoing to generate large deletion of *smyhc* locus to understand the role of slow MyHC in sarcomere assembly during early developmental stages through to adulthood. It is concluded that in the presence of LDM mutations in the MyBP-C binding domain, myosin heads in the SRX state are destabilised, and zebrafish *smyhc1* is orthologous to human *MYH7* but only functions during the early stages of development. Continued work to generate knockout of smyhc locus may describe the function of *smyhc2* and *smyhc3* in later stages of development in the quest to model the progressive phenotype in LDM and MSM patients.

Acknowledgements

Thank you to all those who have supported me through the four years of my PhD. I thank my primary supervisor Dr Julien Ochala for guiding and supporting me throughout the PhD project. I am hugely grateful for the huge amounts of patience and support through many challenges in my mental health, family as well as the challenges in my PhD project itself. I would like to thank my secondary supervisor, Prof. Simon Hughes, for always showing the highest enthusiasm and continuously inspiring me throughout my MRes and PhD. I thank all Ochala lab members, Dr Yotam Levy, Dr Jacob Ross, Natasha Ranu, Edmund Battey and Dr Hannah Dugdale, and all Hughes and Hinits lab members, Vikki Williams-Ward, Tapan Pipalia, Dr Yaniv Hinits, Dr Ailbhe O'Brien, Dr Michael Attwaters, Dr Jeff Kelu, Kees Wanders and Dr Dan Turner for being such lovely friends and always warm and welcoming in the office and so supportive in the lab. I want to show appreciation to Prof Christer Hogstrand, Prof Wolfgang Maret and Prof Stephen Sturzenbaum for being constant supporters, being amazing guidance and always creating a fun environment whilst teaching the Biomedical and Molecular Science Research course. Thank you, Muscular Dystrophy, for funding my work. Most importantly, I thank my family for always showing unconditional love, I thank mum and dad for always feeding me every day, giving all the hugs needed and supporting me through everything and I thank all my sisters Claire, Sin-Yee and Sarah for being cheerleaders through everything. Finally, I would like to thank my boyfriend Erolle, who has supported me through the final years of my PhD, thank you for always being so loving and caring, encouraging me to be more courageous each day, and being the safe space to be comforted during the biggest challenges and supporting me as a scientist. I would have not been able to complete my PhD without such belief in me to become stronger each day, thank you so much Erolle.

Statement of Disruption

Between March 2020 and June 2020, access to the lab were restricted by King's College London due to the UK government's COVID-19 lockdown. Restricted lab access initiated late June 2020 and limited number of researchers were allowed to occupy the lab until June 2021. During restricted lab access, the fish facility was only open to researchers during early morning at 8-9:45am and after 4pm which greatly reduced the number of experiments performed on zebrafish as well as maintaining fish lines. COVID-19 restrictions did shift and delay experiments, use of equipment and fish maintenance with new booking rules with limited access. This delayed my ability to generate large number of fish lines through CRISPR injection, screening multiple CRISPR lines and maintain already existing fish lines.

Mid-March 2020 I caught COVID-19 and subsequently was diagnosed with long covid. Further disruption of my studies for a further 2 months to recover delayed and put all my experiments to a halt. Returning to the lab to continue experiments after partial recovery from long covid after a total of 5-months disruption did delay my output to generate data and write up my thesis as the symptoms involved impacted my usual workflow pre-pandemic. Thankfully King's College London provided a 5-month extension to my PhD to cover lockdown disruption and for my 2-month time out of my PhD. Although this did not give me enough time to complete all the experiments to generate multiple zebrafish mutant lines (due to restricted access and inability to work to 100% performance pre-covid), it did give me enough time to wrap up any existing experiments that were partially completed and write up my thesis.

Abbreviations

AA, amino acid

AMO, antisense morpholino

ATP, adenosine triphosphate

BBR, 2% Boehringer Blocking Reagent™

BH, blocked head

BLAST, basic local alignment search tool

BTS, N-benzyl-p-toluene sulphonamide

Ca2+, calcium

cDNA, complementary DNA

CRISPR, clustered regularly interspaced short palindromic repeats

crRNA, CRISPR RNA guide sequence

CSA, cross-sectional area

DAPI, 4',6-diamidino-2-phenylindole

DCM, dilated cardiomyopathy

ddH_2O, double distilled water

dH_2O, distilled water

DIG-UTP, digoxigenin-uridine triphosphate

DNAse, deoxyribonuclease, enzyme degradation of DNA

dNTP, deoxyribonucleotide triphosphate

Dpc, days post coitum

Dpf, days post fertilisation

DRX, disordered relaxed state

DSB, double-strand break

EGTA, ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*',*N*'-tetraacetic acid

EJC, exon-junction complex

ELC, essential light chain

Els, embryonic lateralis superficialis

EtOH, ethanol

FH, Free head

Fmyhc, fast myosin heavy chain

gRNA, guide RNA

GS, goat serum

HCM, hypertrophic cardiomyopathy

Hh, hedgehog

HM, horizontal myoseptum

HMM, heavy meromyosin, head region of myosin, consisting of S1 and S2

Hpf, hours post fertilisation

HR, homologous recombination

HRM, high-resolution melt

HRP, Horseradish peroxidase

Icp, infracarinalis posterior

IFM, indirect flight muscle

IHM, Interacting Heads Motif

INDEL, insertion/deletion

Iob, inferior obliquus

ISH, *in situ* hybridisation

K, Potassium

K_2HPO_4, di-potassium hydrogen phosphate

KH_2PO_4, potassium di-hydrogen phosphate

KO, knockout

L50 kDa, Lower 50 kDa

LB, Luria broth

LDM, Laing distal myopathy

LMM, light meromyosin, tail region of myosin

LOF, Loss-of-function

MABTween, MAB in PBS

Mant-ATP, 2'/3'-O-(N-Methyl-anthraniloyl)-adenosine-5'-triphosphate

mATP, Adenosine triphosphate

Mef2, myocyte enhancer factor 2

Mg, Magnesium

Mhc, myosin heavy chain (nomenclature in *D. melanogaster*)

MO, morpholino oligonucleotide

MOPS, 3-(N-Morpholino)propane sulfonic acid

Mpcs, muscle precursor cells

Mpf, months post fertilisation

MSM, myosin storage myopathy

MTJ, Myotendinous Junction

MyBP-C, Myosin Binding Protein C

MYH, Myosin Heavy Chain

MyHC, Myosin Heavy Chain

Myo, myosin (nomenclature in *C. elegans*)

NADH, nicotinamide adenine dehydrogenase

Nc, notochord

NHEJ, nonhomologous end-joining

NMD, nonsense-mediated mRNA decay

Nt, neural tube

Oligo, oligonucleotide

ORF, open reading frame

PAM, protospacer adjacent motif

PBS, phosphate-buffered saline

PBTween, Tween20 in PBS

PCR, polymerase chain reaction

PFA, paraformaldehyde

Pi, Organic phosphate

PMSF, phenylmethylsulfonyl fluoride

PTC, premature termination codon

PTU, 1-phenyl 2-thiourea, to block pigmentation in zebrafish

RLC, regulatory light chain

RNA, ribonucleic acid

rNTP, ribonucleotide triphosphate

S1, subfragment 1 of the myosin head

S2, subfragment 2 of the myosin head

Sca, supracarinalis anterior

SFLS, stress fibre-like structures

sgRNA, single guide RNA

Sh, sternohyoid

SM, slow myosin

Smyhc, slow myosin heavy chain

SRX, Super relaxed state

Ssoligo, short single oligonucleotide

TALENS, transcription activator-like effector nucleases

TCEP, Tris(2-carboxyethyl)phosphine

Tm, melting temperature

tracrRNA, tracr guide RNA

U50 kDa, Upper 50 kDa

UTR, untranslated region

WISH, whole-mount *in situ* hybridisation

WT, wild type

XY, lateral view

XZ, transverse view

ZFN, zinc finger nucleases

Chapter 1

General introduction

1.1. Congenital Myopathies and the MYH7 gene

Myopathies, in general, are defined as disorders where detrimental muscle dysfunction is a prominent feature. Multiple origins for such a group of diseases are known, including: the central nervous system, peripheral muscle system or the skeletal muscle itself. Congenital myopathies are early-onset muscle diseases which and occur at birth or early stages of life (Ravenscroft *et al.*, 2017). Clinical assessment of congenital myopathy includes hypotonia, muscle weakness, the disproportion of muscle fibre types, centralised nuclei, cores and nemaline bodies. A recent increase in genetic diagnosis of many congenital myopathies have been clinically characterised into many subtypes through the discovery of affected muscle genes (Boycott *et al.*, 2013). There are many different types of congenital myopathies as shown in Table 1.1. The congenital myopathies focused upon in the present work are due to sarcomeric gene mutations on one myosin gene, *MYH7*. *MYH7* encodes beta/slow myosin heavy chain (MyHC I) that is known to facilitate muscle contraction in slow skeletal muscle and in the heart ventricle. Currently, there are no curative medicines for *MYH7*-related congenital myopathies and available treatments simply target the various symptoms (Tajsharghi and Oldfors, 2013; Topaloglu, 2020). During my analysis, I first described in my introduction the disease and its related clinical phenotypes and define myosin structure and function concerning the human *MYH7* mutations that are well-described in the literature. Secondly, in my results chapters, I studied the potential underlying molecular and cellular mechanisms leading to pathology in the presence of defective slow myosin molecules and subsequently developed animal models that could be beneficial in the quest for treatment designed for *MYH7*-related diseases.

Table 1.1. Types of congenital myopathies.

Congenital Myopathy	*Description*	*Clinical Phenotype*	*Affected Genes*	*Reference*
Central core disease	• Large, well-demarcated, central cores within numerous myofibres.	• Muscle weakness • Developmental problems • Some may develop malignant hyperthermia (reaction to general anesthesia)	• RYR1 • SELENON • MYH7 • TTN	• (Robinson *et al.*, 2006)
Centronuclear myopathy	• An elevated number of myofibres • centrally or internally located nuclei.	• Muscle weakness • Affects the face, arms, legs, eyes • Breathing difficulties	• CNMX • MTM1 • DNM2 • CNM1 • BIN1 • CNM2	• (Laporte *et al.*, no date; Bitoun *et al.*, 2005; Tosch *et al.*, 2006; Nicot *et al.*, 2007; Koutsopoulos *et al.*, 2013)
Congenital fibre type disproportion myopathy	• Small fibres • The predominance of either fast or slow fibres	• Muscle weakness • Affects the face, neck, arms, leg, and trunk	• ACTA1 • SEPN1 • TPM3 • RYR1 • TPM2 • MYH7	• (Laing *et al.*, 2004, 2005; Sobrido *et al.*, 2005; Clarke *et al.*, 2008; Lawlor *et al.*, 2010; Ortolano *et al.*, 2011)
Nemaline myopathy	• Presence of electron-dense rod-like aggregates within myofibres	• Muscle weakness • Affects the face, neck, arms, and legs • Sometimes cases with scoliosis • May cause breathing and feeding problems	• ACTA1 • CFL2 • TPM2 • TPM3 • TNNT1 • NEB	• (Nowak, Ravenscroft and Laing, 2013; Romero, Sandaradura and Clarke, 2013)
Multi minicore disease	• Presence of multiples cores within a myofibre cross-sections	• Muscle weakness • Affects the arms and legs • Scoliosis	• SEPN1 • RYR1 • MYH7 • TTN • MEGF10 • EMARDD • CACNA1S	• (Muelas *et al.*, 2010; Cullup *et al.*, 2012)
Hyaline body myopathy	• Hyaline bodies found between slow fibres • Hyaline bodies containing protein aggregates	• Muscle weakness • Muscle hypertrophy • Symptoms are quite variable	• MYH7 • FHL1 • NEB	• (Goebel and Blaschek, 2011)

1.2. *MYH7* mutations cause several distinct clinical pathologies

To date, over 200 mutations in the *MYH7* gene have been associated with congenital myopathies, with manifestation, symptoms and severity being variable (Tajsharghi and Oldfors, 2013). *MYH7* mutations have often been associated with either cardiac phenotypes, such as hypertrophic cardiomyopathy (HCM) or dilated cardiomyopathy (DCM), and/or skeletal muscle symptoms, such as Laing Distal Myopathy (LDM) or Myosin Storage Myopathy (MSM). Interestingly, the position of mutation along the slow myosin molecule can dictate which one of the four diseases may be present. Mutations in the *MYH7* gene have primarily been dominant, involving mutations leading to amino acid substitutions and/or deletions (Appendix 1.1). Human *MYH7,* located on chromosome 14, consists of 40 exons encoding the 1935-amino acid long MyHC I protein which is subdivided into the N-terminal heavy meromyosin (HMM) consisting of subfragment-1 (S1) and subfragment-2 (S2) and at the C-terminus, the light meromyosin (LMM) (Fig 1.1A). Each subdivision of the MyHC I protein describes the general structure of the myosin molecule (Fig 1.1B1). The S1 region encodes for the head region where myosin can interact with actin and ATP, the S2 region encodes for the neck of the myosin molecule for head movement and the LMM encodes for the tail for myosin monomers to dimerise into a double head myosin dimer and subsequently interlace into a myosin filament in the sarcomere for muscle contraction (Fig 1.1B2). HCM and DCM-related mutations are mainly concentrated in the S1 and S2 regions (Fig 1.1C) whilst LDM and MSM-related defects are primarily located in the LMM region (Fig 1.1C) (Lamont *et al.*, 2014).

The most common diseases associated with *MYH7* mutations are HCM and DCM and have been widely explored in literature (Tajsharghi and Oldfors, 2013). Since *MYH7* is expressed in ventricular cardiac muscle as well as skeletal muscle, mutations in *MYH7* can lead to cardiomyopathy in the absence of skeletal myopathy and some skeletal *MYH7*-associated myopathies may present with cardiomyopathy (Darin *et al.*, 2007; Overeem *et al.*, 2007; Tajsharghi *et al.*, 2007; Uro-Coste *et al.*, 2009; Homayoun *et al.*, 2011). HCM presents with ventricular hypertrophy, hypercontraction, altered myosin head positioning and cardiac myocyte disorganisation (Maron *et al.*, 1995; Fatkin and Graham, 2002; Frazier *et al.*, 2008; Alamo *et al.*, 2017). Contrastingly, DCM presents with dilated and enlarged ventricles leading to weakened contraction (Walsh *et al.*, 2009; Alamo *et al.*, 2017). Greater understanding of pathology has led to advances in the treatment of HCM treatments such as using Mavacamten to reverse the symptoms in HCM patients (Anderson *et al.*, 2018; Spertus *et al.*, 2021). Even though the underlying mechanism of HCM and DCM are widely explored (Tajsharghi and Oldfors, 2013), the pathophysiology of LDM and MSM are less well understood. When investigating the localisation of the *MYH7* mutations affecting skeletal muscle, mutations in *MYH7* leading to LDM and

MSM are predominantly located in the LMM region and are absent in the S1 and S2 region (Fig. 1.1D). LDM-related residue substitutions are generally found in the earlier segment of the LMM whereas MSM amino acid replacements are primarily observed in the C-terminal end of the LMM region (Fig 1.1D). A clear overlap between mutations leading to either LDM or MSM is, however, present from aa 1600 to aa 1800 (Lamont *et al.*, 2014). Thus, the functional role of each segment within the LMM may describe the pathology in leading to either LDM or MSM.

Mutations leading to HCM, and DCM were found to have cardiac involvement and thus, are easily distinguished from LDM and MSM. Just over half of the patients with HCM and DCM have early onset of disease (Fig 1.2). In my PhD, my main aim is to focus on mutations in *MYH7* leading to skeletal muscle diseases LDM and MSM. From here, I focus my investigation on the main symptoms and details distinguishing between the two skeletal muscle diseases, LDM and MSM. There are several similarities in clinical symptoms between LDM and MSM such as overall skeletal muscle weakness, distal lower limb myopathy, proximal myopathy, hypertrophy of muscle, axial involvement, and abnormal biopsy findings (Fig 1.2, Appendix 1.1). Despite such similarities between LDM and MSM, there are key distinct symptoms that enable each disease to be categorised (Fig 1.2).

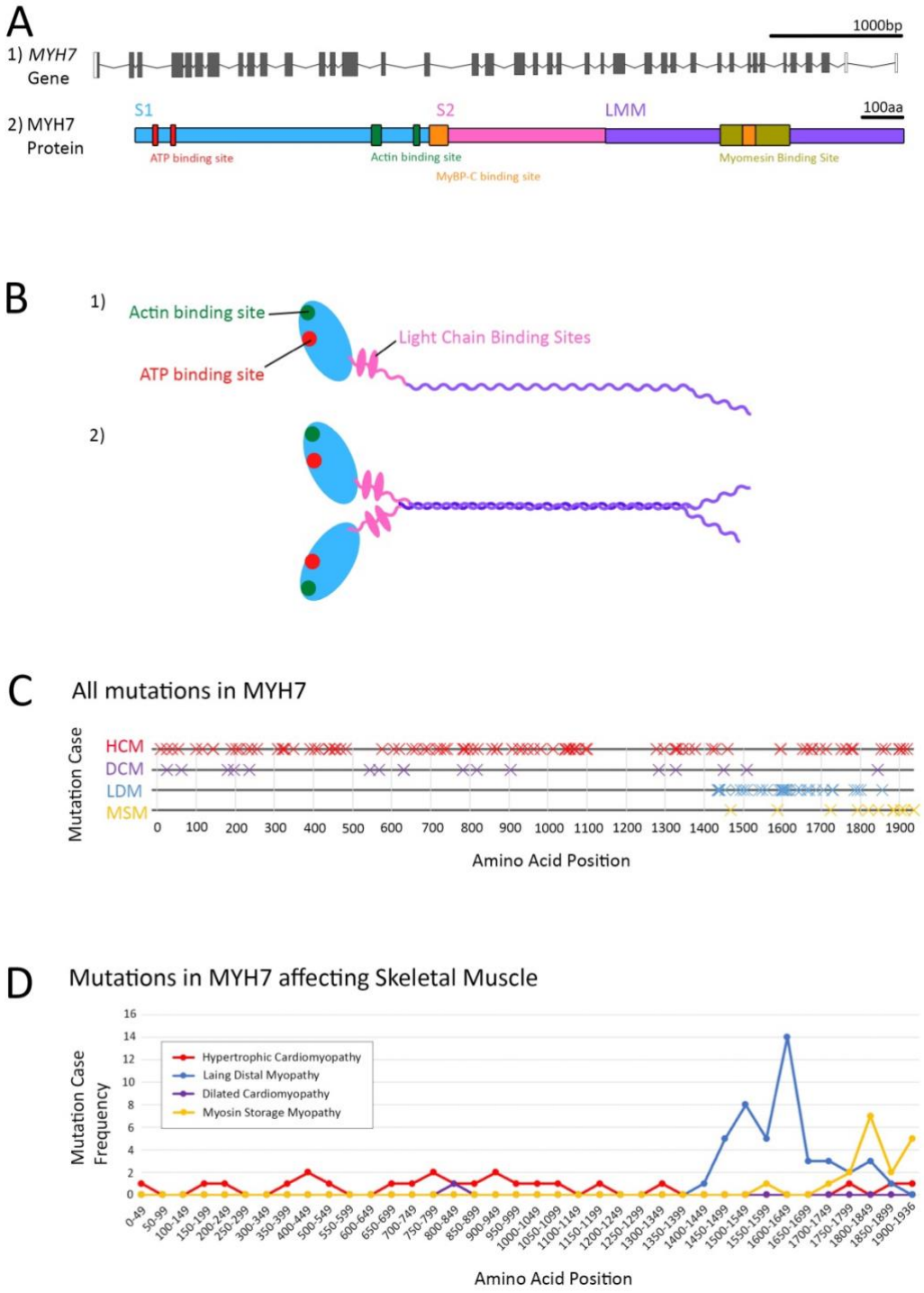


Figure 1.1. Mutations in MYH7 and their corresponding disease.
A) Gene sequence map of *MYH7* with 40 exons, 38 of which are coding exons. The transcript length is 6027 bps. **B)** Protein sequence map of *MYH7* consisting of 1935 amino acid residues. *MYH7* is subdivided into Subfragment-1 (S1) in blue, subfragment-2 (S2) in pink and light meromyosin (LMM) in purple. **C)** All mutations in *MYH7* mapped onto amino acid sequence. Each line represents the different diseases: Hypertrophic cardiomyopathy (HCM) in red, dilated cardiomyopathy (DCM) in purple, Laing distal myopathy (LDM) in blue and myosin storage myopathy (MSM) in yellow **D)** Number of mutations in *MYH7* found in the literature with clinical phenotype in skeletal muscle mapped onto amino acid sequence. Mutation case frequency Hypertrophic cardiomyopathy (HCM) in red, dilated cardiomyopathy (DCM) in purple, Laing distal myopathy (LDM) in blue and myosin storage myopathy (MSM) in yellow. All mutations in *MYH7* are sourced and detailed in Appendix 1.1.

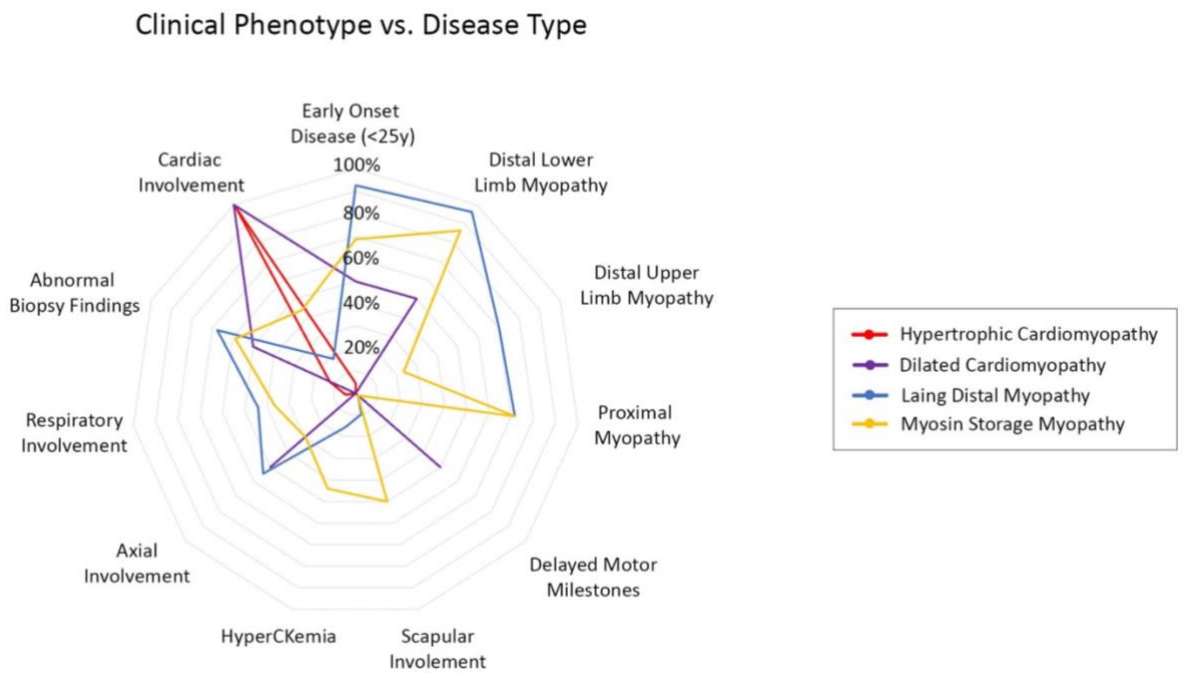


Figure 1.2. Clinical phenotype vs *MYH7* Disease.
Radar chart clinical phenotype and their percentage prevalence between the four main diseases associated with *MYH7* mutations: Hypertrophic cardiomyopathy (red), dilated cardiomyopathy (purple), Laing distal myopathy (blue) and myosin storage myopathy (yellow). Percentage calculated by presence of particular phenotype from one disease category in comparison to the total number of cases with the disease. Clinical phenotypes were obtained from literature to generate this graph and detailed in Appendix 1.1.

1.2.1. Laing Distal Myopathy (LDM)

Both *de novo* and familial cases of LDM have been identified (Lamont *et al.*, 2006). The cases encompass patients with symptoms early in childhood or individuals with phenotypes appearing as after 50 years of age (Laing *et al.*, 1995; Mastaglia *et al.*, 2002; Lamont *et al.*, 2006; Tasca *et al.*, 2012). Overall, muscle weakness in distal limbs (hands and feet) is conspicuous and may expand to other muscles too. Raising all five fingers are challenging exercise as finger extensors are weak, patients show an inability to raise the middle, ring and fifth finger in their attempt but all cases show an ability to raise their index finger (Fig 1.3A). Raising their feet upward is another challenging exercise as their ankle dorsiflexors are weak, patients appear to show slight flexion of their toes in their attempt but are unable to use the ankle to raise their foot (Fig 1.3B). Additionally, patients may experience neck flexion problems and, in some cases scoliosis (Fig 1.3C) (Tajsharghi and Oldfors, 2013). Muscle biopsy findings can include a predominance of fibres expressing MyHC I when cross-sections are stained for NADH, additionally, type I fibres appear smaller than fast fibres, additionally, there are scattered mini-cores (Fig 1.5A), internal nuclei, mitochondrial abnormalities, rimmed vacuoles and necrosis (Mastaglia *et al.*, 2002; Tasca *et al.*, 2012; Tajsharghi and Oldfors, 2013). Despite such a distinguished phenotype, the severity of each diagnostic phenotype will vary from one patient to another, whether they have the same mutation or between the different *MYH7* mutations leading to LDM.

1.2.2. Myosin Storage Myopathy (MSM)

Like LDM, MSM also has *de novo* or familial cases (Cancilla *et al.*, 1971; Barohn, Brumback and Mendell, 1994; Masuzugawa *et al.*, 1997; Bohlega *et al.*, 2003; Tajsharghi *et al.*, 2003; Laing *et al.*, 2005; Shingde *et al.*, 2006; Pegoraro *et al.*, 2007; Uro-Coste *et al.*, 2009; Stalpers *et al.*, 2011). Only one patient show delayed motor milestones, such as difficulty climbing stairs, running or a waddling gate (Tajsharghi and Oldfors, 2013). They are unable to raise all five fingers where only the index and fifth fingers can be raised as patients have muscle weakness in their fingers and their palm (Fig 1.4A). Patients also show muscle wasting in the upper limbs, particularly in the thighs and forearms (Fig 1.4B). A rare set of patients show a severe progression of their symptoms, where patients show additional phenotypes including scoliosis, assisted ventilation and scapular winging (Fig 1.4C) (Bohlega *et al.*, 2003; Stalpers *et al.*, 2011; Tajsharghi and Oldfors, 2013). Muscle biopsy specimens display protein aggregates which are present as clusters between slow fibres and are also described as hyaline bodies (Fig 1.5B, C) (Barohn, Brumback and Mendell, 1994; Bohlega *et al.*, 2003; Shingde *et al.*, 2006; Pegoraro *et al.*, 2007; Tajsharghi and Oldfors, 2013). Protein aggregates between slow fibres in MSM patients are mostly made of filamentous material that can be slow MyHC immunoreactive (Fig 1.5C) (Tajsharghi *et al.*, 2003).

Likewise with LDM, with such distinguished phenotype, the severity of each diagnostic phenotype will vary from one patient to another, whether they have the same mutation or between the different *MYH7* mutations leading to MSM. Since there is high variability in the severity of clinical phenotype in LDM and MSM patients, diagnosing between the two diseases can be difficult without muscle biopsy data. There are some *MYH7* patients that do not fit the initial diagnostic criteria (without muscle biopsy) for LDM or MSM and have been misdiagnosed as cases of limb-girdle syndrome or scapuloperoneal myopathy (Pegoraro *et al.*, 2007; Ortolano *et al.*, 2011). Limb-girdle syndromes predominantly affect the proximal limb muscles especially in the shoulder and hip areas, whereas scapuloperoneal myopathies typically present with symptoms of the scapular, lower leg, and sometimes facial muscles (Thomas, Schott and Morgan-Hughes, 1975; Groen *et al.*, 2007). Despite such variable phenotypes, the main diagnostic criteria to distinguish between LDM and MSM are from analysing their muscle biopsies where biopsies from LDM show fibre type disproportion and small slow fibres and biopsies from MSM patients show the presence of myosin protein aggregates between fibres.

Figure 1.3. The clinical phenotype for Laing Distal Myopathy.
A) Distal lower limb myopathy shown in the hands with weakness in finger extension. Patients are unable to lift exterior fingers in an attempt to raise all fingers upwards. **B)** Weakness shown in the feet when patients were examined for ankle flexion to point feet and toes upward, patients appear unable to flex the foot and exterior toes upward and appear dropped. **C)** Axial involvement where some patients showcase scoliosis. Figure permission granted from Van den Bergh *et al.*, 2014; Oda *et al.*, 2015; Fiorillo *et al.*, 2016.

Figure 1.4. The clinical phenotype for Myosin Storage Myopathy.
A) Distal lower limb myopathy shown in the hands with weakness in finger extension. Patients are unable to lift all exterior fingers. **B)** Thighs show muscle wasting of the posterior compartment in thigh muscles (1-2) in more severe cases (left) and less severe cases (right) and wasting of forearms with abnormal elbow flexion (3-4). **C)** Scapular winging was identified in more severe cases (left) and less severe cases (right). Figure permission granted from Pegoraro *et al.*, 2007; Cullup *et al.*, 2012.

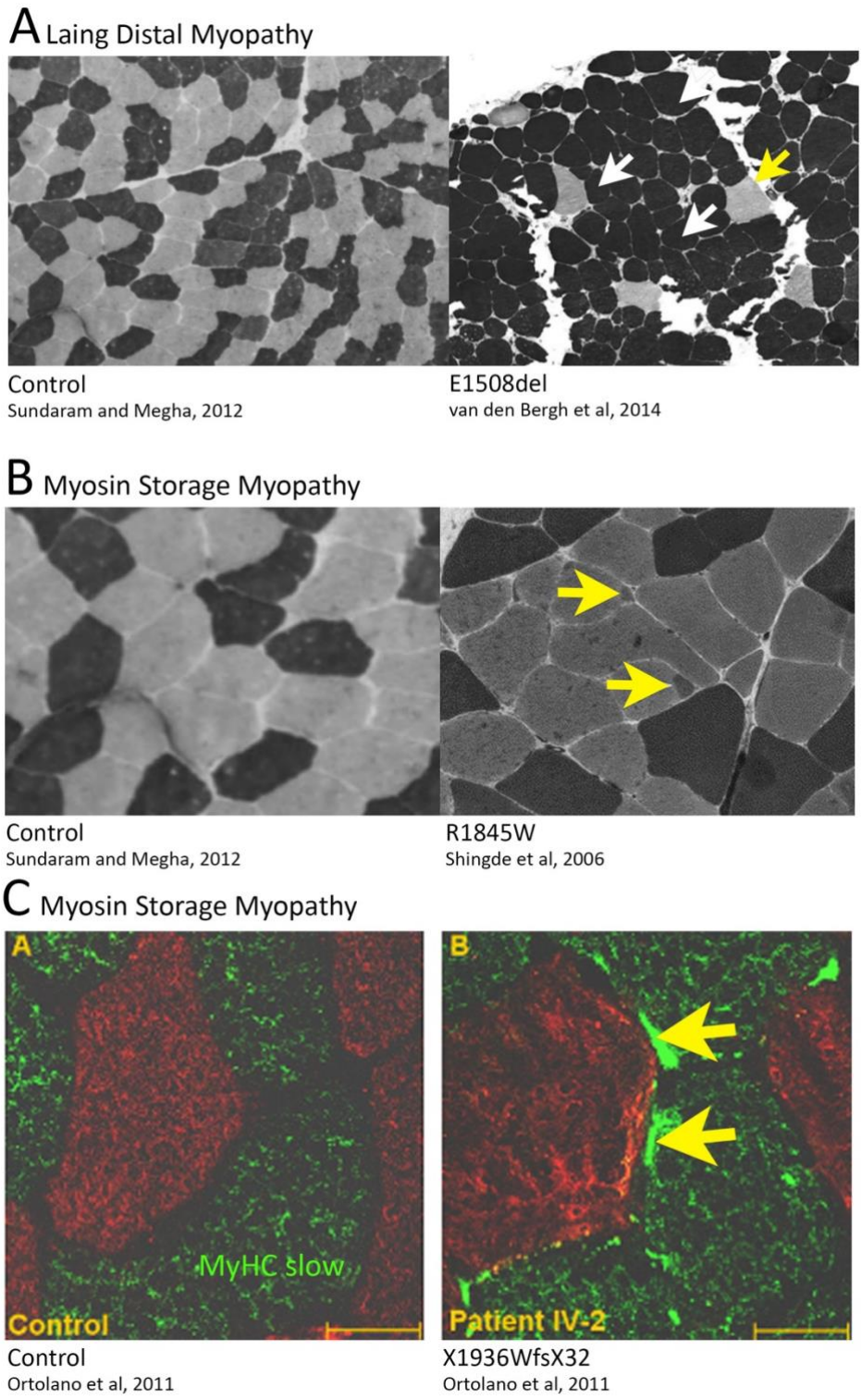


Figure 1.5. Muscle Biopsy Diagnostics for LDM and MSM
A) LDM patient biopsy with NADH staining showing fibre type disproportion. There is type I fibre predominance and with small type I fibres, type I fibres are labelled in dark staining (white arrows). Fibres also show oxidative defects where pale circle patches (mini cores) are found in NADH staining (yellow arrow). **B)** MSM patient biopsy with NADH staining showing the presence of hyaline bodies between fibres. **C)** MSM patient biopsy with slow myosin antibody stain (green) shows hyaline bodies containing aggregates of slow myosin. Fast myosin stain in fast fibres (red) appears normal with no protein aggregation. Figure permission granted from Shingde *et al.*, 2006; Ortolano *et al.*, 2011; Sundaram and Megha, 2012; Van den Bergh *et al.*, 2014.

1.3. Basic muscle physiology and phenotype

During early development, muscle fibres are formed from the fusion of myoblasts, the mesoderm progenitor cells. At neonatal stages, the number of muscle fibres remain constant but grow in size by fusing with postnatal muscle stem cells, known as satellite cells. In adult skeletal muscle, muscle fibres remain constant with only fusion to compensate for muscle turnover from daily use and repair. Muscle can regenerate in response to injury through a series of degeneration and regeneration of tissue, cellular and molecular levels with the presence of satellite cells near the muscle fibres.

Muscle fibre type specification during embryogenesis in vertebrates are governed by the specific spatial and temporal expression of transcription factors MyoD, Myf-5, myogenin and MRF4 which are initiated by several inductive pathways (Cossu *et al.*, 1996; Currie and Ingham, 1996). Myf5 and MRF4 are transcribed in the dorsal medial and ventrolateral ends of the dermomyotome. Cells expressing Myf5 and MRF4 then migrate beneath the dermomyotome and differentiate into first set of mononucleated skeletal muscle cells (Summerbell, Halai and Rigby, 2002; Kassar-Duchossoy *et al.*, 2004).The myotome also have Pax3 and Pax7 expressing stem cells present in the central segment of the dermomyotome. MyoD is expressed in the hypaxial and epaxial progenitors and overlap with Myf5 expression to further develop the myotome alongside Pax3/Pax7 expressing stem cells (Kassar-Duchossoy *et al.*, 2005; Relaix *et al.*, 2005). Later in development, Mrf4 expression is suppressed, myod and myogenin are then expressed to initiate myoblast fusion into multinucleated muscle cells to produce mature myofibres (Tajbakhsh *et al.*, 1997).

Capillaries facilitate the exchange of O_2, substrates and metabolites from the blood to skeletal muscle as well as many organs. Skeletal muscle occupies most capillary beds in the body, especially playing the dominant role for exchange of O_2, glucose, lactate, and fatty acid dynamics during exercise. Chronic diseases such as heart failure, muscle weakness and diabetes have been correlated with impaired capillary function (Klitzman and Duling, 1979; Sarelius and Duling, 1982; Cossu *et al.*, 1996; Frisbee and Barclay, 1999).

Skeletal muscle is regulated through excitation-contraction coupling, a process involving the conversion of electrical activity of muscle fibres to the activation of muscle contraction. The initial steps for excitation-contraction coupling involve the action potential propagation from the spinal cord via motor neurons to the neuromuscular junction. This action potential to the neuromuscular junction triggers the release of acetylcholine (Ach) from nicotinic receptors. Released ACh binds to the post-synaptic receptors causing depolarisation of sarcolemma, depolarisation above the threshold will

initate an action potential that will spread along the surface and into the T-tubules of the muscle fibre (González-Serratos, 1971). Depolarisation down T-tubules activates voltage-sensitive dihydropyridine receptors and thus, open ryanodine receptor channels in the sarcoplasmic reticulum to release stored Ca^{2+} into the muscle fibre cytoplasm. Muscle in the inactive state is stabilised through the troponin/tropomyosin system in which tropomyosin is positioned to block myosin binding sites on thin actin filaments. Ca^{2+} release activates the contractile apparatus by binding to troponin C and subsequently lead to a confirmational change in tropomyosin complex, revealing the myosin binding sites on thin actin filaments to enable cross-bridge formation and force generation powered by ATP activated myosin (Gordon, Homsher and Regnier, 2000). When neural action potential decreases below threshold level, Ca^{2+} ions are transported back into the sarcoplasmic reticulum through the sarcoplasmic reticulum/endoplasmic reticulum ATPase. Lack of Ca^{2+} ions in the cytoplasm lead to tropomyosin returning into its inhibitory conformation and thus block actomyosin binding and cross bridge cycling stops.

1.4. Myosin structure and function

MYH7 is expressed in the heart and slow skeletal muscle to drive the contraction of cardiomyocytes in the ventricles and contraction of slow muscle within the sarcomere (the smallest unit for muscle contraction). As the exact pathogenic mechanisms by which subtle mutations in *MYH7* lead to LDM or MSM remain unclear, it is important to describe what is known about myosin within its basic functional unit, the sarcomere. In the presence of defective myosin molecules, pathology of LDM and MSM may be a result of altered aggregation or impair the function of the myosin in cardiac and/or slow skeletal muscle.

1.4.1. Muscle structure - from whole muscle to sarcomere

Skeletal muscle is made up of muscle fibres and it is important to note that there are two types of skeletal muscle fibres in vertebrates, slow and fast (type 1 and type 2, respectively). Whilst fast fibres consist of three subgroups (2A, 2B and 2X), all of which show rapid contraction speed to produce high force but have low resistance to fatigue. Slow fibres contain more mitochondria than fast fibres and have oxidative metabolism to enable efficient muscle contraction to produce small but frequent forces and are resistant to fatigue. To generate such force in either fast or slow skeletal muscle, muscle fibres contain contractile tissue that is highly conserved in vertebrates and are organised with distinctive features for muscle contraction. Notably, cylindrical bundles of muscle fibres able to contract and relax. Each fibre contains myofibrils made up of small contractile units called sarcomeres that are arranged in series and parallel (Fig 1.6). Sarcomeric MyHCs form the regular filamentous array of

parallel thick myosin filaments interdigitating with parallel arrays of thin actin filaments. These structures are stabilised by M-lines and Z-lines that crosslink myosin thick filaments and actin thin filaments respectively (Howard, 1997; Alberts *et al.*, 2015). During muscle contraction, coordinated events whereby actin filaments slide between myosin filaments causes Z-disk positioning to shorten and thus, causes muscle shortening to generate isometric and concentric muscle contraction.

1.4.2. Sarcomeric assembly

The exact mechanism and sequence of events in the assembly of the sarcomere remain controversial and there are several models describing aspects of sarcomere assembly (Fig 1.7). There have been initial models describing the ability for myosin molecules to self-assemble into thick filaments. Myosin molecules interlock at their C-terminal coiled coil rod domain, known as the assembly competence domain. Thus, enabling myosin molecules to form thick filaments and integrate into the sarcomere. However, the mechanism for integration of myosin molecules into the sarcomere still remain unclear (Atkinson and Stewart, 1991; Sohn *et al.*, 1997; Ikebe *et al.*, 2001; Ojima *et al.*, 2015). One model of sarcomere assembly describes the formation of stress fibre-like structures. Sarcomere assembly involves utilising non-muscle myosin as a template for sarcomere proteins to assemble to form pre-myofibril. Pre-myofibrils containing non-muscle myosin and are then later replaced with muscle myosin such as embryonic, neonatal, fast or slow MyHC to form mature myofibrils (Fig 1.7A) (Rhee, Sanger and Sanger, 1994). Muscle-specific desmin, actinin and Z-disk portion of titin are initially expressed to produce stress fibre-like structures (SFLS) known as premyofibrils containing non-muscle myosin II (Rhee, Sanger and Sanger, 1994; Swailes *et al.*, 2006). These stress fibre-like structures act as a template for the formation of a myofibril (Dlugosz *et al.*, 1984). A second model describes the independent assembly of actin filaments stabilised by z-disks to form I-Z-I bodies (Fig 1.7B). I-Z-I bodies are proposed to be assembled prior to the integration of myosin and this model is known as the "stitching model" of sarcomere assembly (Rhee, Sanger and Sanger, 1994; Holtzer *et al.*, 1997; Van Der Ven *et al.*, 1999; Sanger *et al.*, 2005). Data from cultured skeletal muscle cells describe independent actin filament complex formation from myosin thick filament assembly (Lin *et al.*, 1994). A third model describes the role of titin recruited by α-actinin to bind to the Z-disk region and the M-line to act as a template to regulate the alternating patterning of I-Z-I bodies and myosin filaments (Kelly and Zacks, 1969; Tokuyasu and Maher, 1987; Schwander *et al.*, 2003; Au *et al.*, 2004). The transition from premyofibrils to myofibrils occur when SFLS coupled with titin molecules stretch out whereby Z-disk spacing is increased from 1 to 2 μm (Yang, Obinata and Shimada, 2000). This stretching exposes M-band region of titin for the assembly of myomesin molecules into the M-band. Then the final step is the assembly if sarcomeric myosin filaments to integrate the I-Z-I bodies form the A band

using titin as the molecular ruler for sarcomere assembly (Komiyama, Maruyama and Shimada, 1990; Péault *et al.*, 2007)

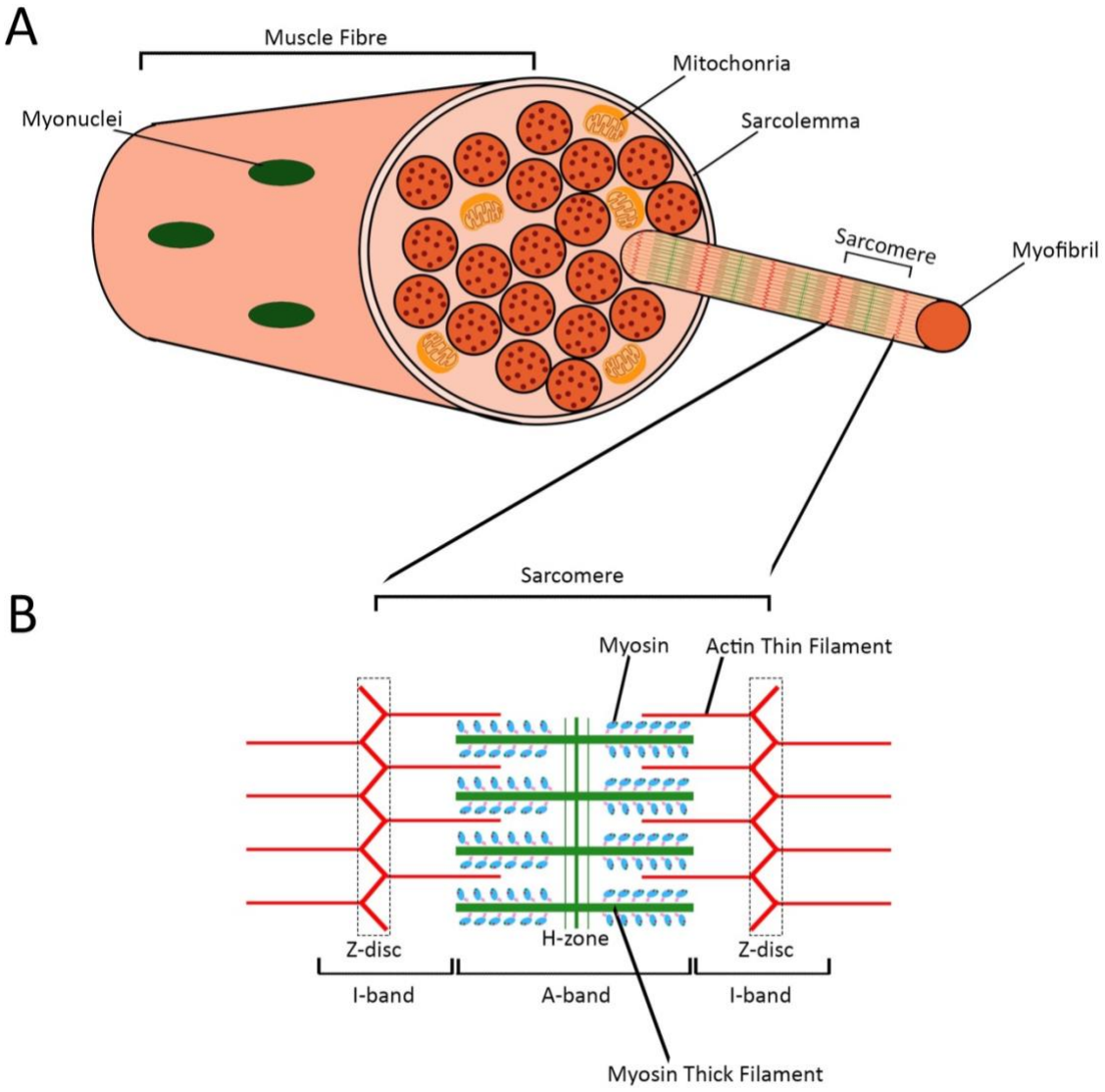


Figure 1.6. Anatomy of skeletal muscle fibre
A) Skeletal muscle is made up of myofibers which are elongated muscle cells. Muscle fibres are in turn made of bundles of smaller myofibrils, mitochondria and surrounded by the sarcolemma. Each myofibril is made up of highly organised repeat structures called sarcomeres. **B)** Schematic showing basic components of the sarcomere. Components labelled are the Z-disk which anchor the actin filaments, I-band, a region only containing actin filaments, A-band, a region only containing myosin filaments and H-zone, where no A-band or I-band overlap. Adapted from Creative Commons Attribution 4.0 International license available online: https://open.oregonstate.education/aandp/chapter/10-2-skeletal-muscle.

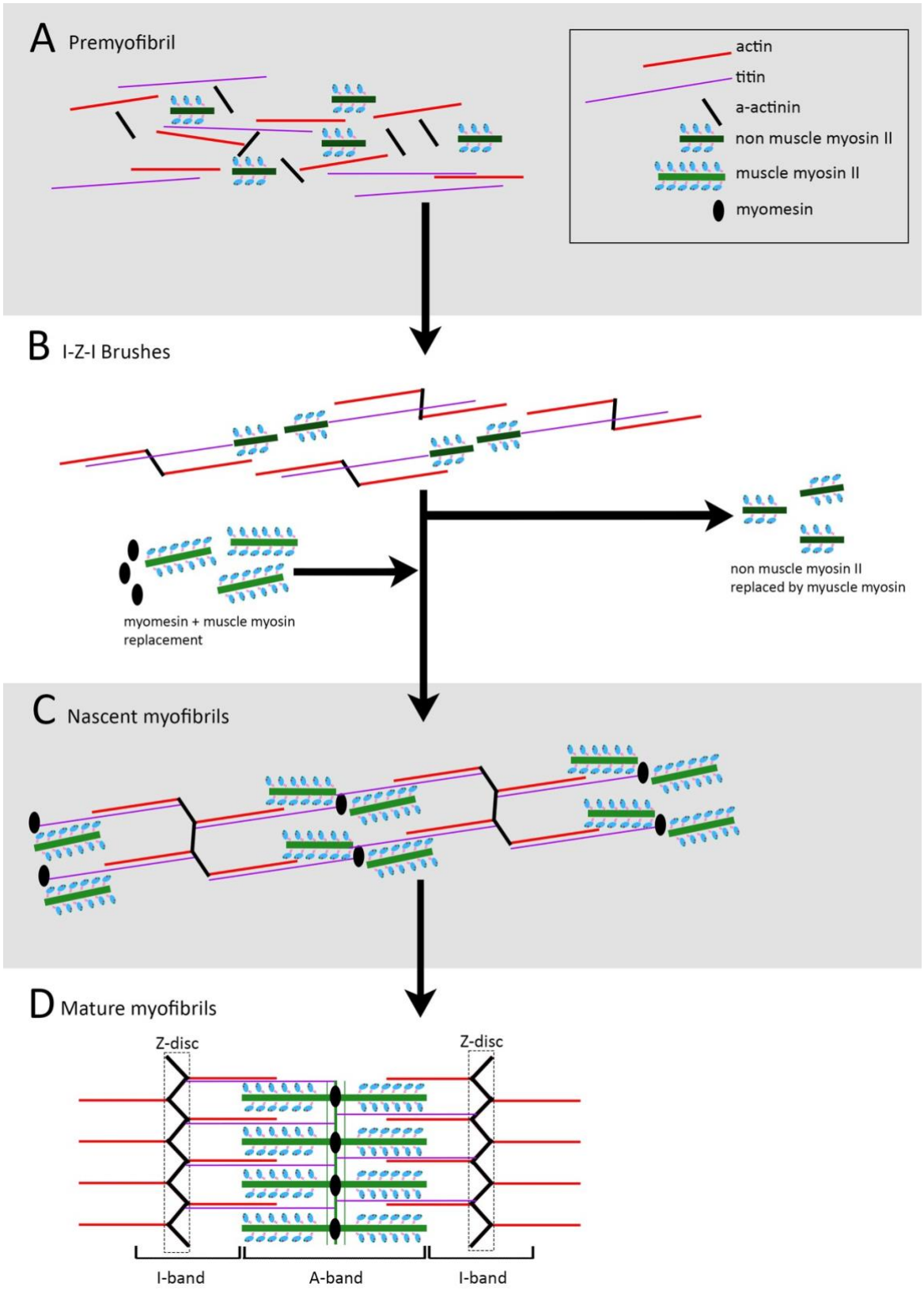


Figure 1.7. Schematic describing sarcomere assembly
A) Premyofibrils form where actin filaments, non-muscle myosin and a-actinin accumulate together at the edge of the muscle cell. **B)** I-Z-I brushes form when premyofibrils assemble and fuse together. Myomesin and muscle myosin are recruited. **C)** Nascent myofibril form when muscle myosin II replaces non-muscle myosin II using titin as a molecular ruler. **D)** Mature myofibril formed with aligned thick filament into A-band and stabilised with M-band proteins, myomesin and C-proteins. Figure adapted from Du *et al.*, 2003.

1.4.3. MyHC I expression

In humans, there are a total of eleven sarcomeric MyHC genes, from an evolutionary standpoint, the oldest of these genes is *MYH16* and was ancestrally expressed for jaw muscles (Rossi *et al.*, 2010). A later duplication event led to the formation of *MYH15* and *MYH14 (MYH7B),* which were the ancestral skeletal and cardiac MyHC genes (Rossi *et al.*, 2010). Currently there are two cardiac MyHC genes, *MYH6* and *MYH7* which are present in tandem on chromosome 14 (Yamauchi-Takihara *et al.*, 1989; Gulick *et al.*, 1991) whereas the fast, embryonic and neonatal skeletal MyHC genes are present in tandem on human chromosome 17. Each MyHC isotype and related gene has specific roles in development and/or physiology (Table 1.2). The human *MYH7* gene (NM_000257) is located on the reverse strand of chromosome 14: 23,412,740-23,435,660 and consists of 38 coding exons with 2 flanking UTRs. *MYH7* encodes the 1935 amino acid MyHC I protein which is expressed both in heart ventricles and in slow skeletal muscle. *MYH7* is closely linked to *MYH6*; they are present next to each other on the same chromosome (Yamauchi-Takihara *et al.*, 1989). Nevertheless, *MYH6* is only expressed in the heart and in the atrial cardiac muscle, whereas *MYH7* has both cardiac ventricle and slow skeletal muscle localisation and the only MyHC isotype for slow skeletal muscle in humans (Mahdavi, Periasamy and Nadal-Ginard, 1982).

Table 1.2. MYH and MYL genes are expressed in developing mammalian skeletal muscle.
Table adapted from Schiaffino *et al.*, 2015 with addition from Schiaffino and Reggiani, 2011, Rossi *et al.*, 2010.

	Gene	***Protein***	***Expression in the development of muscle***	***Expression in adult muscle***
Myosin Heavy Chains	MYH3	MyHC-Emb	Embryonic and fetal	Extraocular, masticatory, laryngeal, muscle spindles
	MYH8	MyHC-Neo	Embryonic and fetal	Extraocular, masticatory, laryngeal
	MYH2	MyHC-IIa	Fetal	Type 2A fast
	MYH4	MyHC-IIb	Postnatal	Type 2B fast
	MYH1	MyHC-IIx/d	Late fetal	Type 2X fast
	MYH7	MyHC-I/β	Embryonic and fetal	Type I slow and heart ventricles
	MYH6	MyHC-α	Embryonic and fetal	Heart atrium
	MYH13	EO-MyHC	Embryonic and fetal	Extraocular
	MYH14/MYH7B	MYH14	Embryonic and fetal	Extraocular
	MYH15	MYH15	Postnatal	Extraocular
	MYH16	MYH16	Embryonic and fetal	Jaw
Essential Light Chains	MYL1	MLC-1fast	Embryonic	Fast
	MYL1	MLC-3fast	Fetal	Fast 2B predominance
	MYL4	MLC-1emb/atrial	Embryonic	Heart Atrium
	MYL3	MLC-1sb	Fetal	Type I slow and heart ventricles
	MYL6B	MLC-1sa	Fetal	Type I slow
Regulatory Light Chains	MYLPF	MLC-2fast	Embryonic and fetal	Fast
	MYL2	MLC-2slow	Embryonic and fetal	Type I slow and heart ventricles

1.4.4. MyHC I in muscle development

Myosin molecules are formed through the dimerization of individual myosin units and stabilised through a coiled-coil structure (Fig 1.1B). The coiled-coil structure was first introduced from X-ray diffraction studies modelling the coiled-coil as a packing mechanism with the presence of a heptad pattern for the two myosin monomers to adhere together (Crick, 1953; Cohen and Holmes, 1963). The heptad repeat is made up of amino acids that are named "*abcdefg*" to label the function of each amino acid playing different functional purposes for the myosin molecules to dimerise and interlace myosin dimers into a larger thick filament (Fig 1.8A) (Lupas, 1996). Amino acids at the *a* and *d* positions are primarily hydrophobic residues and are responsible for binding myosin monomers together to form coiled-coil myosin dimers (Fig 1.8A) (McLachlan and Karn, 1982). Amino acids *e* and *g* are generally

polar or charged residues that may form salt bridges to stabilize myosin dimers (Fig 1.8A) (Lupas, 1996). Amino acids at positions *b, c* and *f* are charged residues that can interact with other myosin dimers (Fig 1.8A) (McLachlan and Karn, 1982), perhaps to facilitate thick filament formation. The heptad repeat pattern continues through and periodically has flexible skip residues which are present within one or two turns of the α-helix (McLachlan and Karn, 1982). In *C. elegans*, there are four skip residues, the first three skip residues are separated by 196 residues and the fourth skip residue by 224 residues (McLachlan and Karn, 1982). The presence of skip residues enhances the skeletal muscle myosin coiled-coil to stabilize together but the main role is to form a larger distribution of alternating positive and negative charges *b, c* and *f* residue (Atkinson and Stewart, 1992) and is highly conserved amongst vertebrates and invertebrates (Rahmani *et al.*, 2021). There are six alternating positively and negatively charged amino acid patterning in a 28 amino acid (aa) repeat throughout the molecule. Within each of the six alternating charged amino acids in the 28 aa patterns, the strongest positive charge is in the first *b* aa (*1b*) and the strongest negative charge is in the 3rd *b* aa (3b) positions (Fig 1.8B). The strongest *1b* and *3b* aa enable multiple myosin molecules to bind together via polar charges to form larger myosin filaments in sarcomeres (Fig 1.8B)(McLachlan and Karn, 1982). Myosin molecules were predicted to pack together into crystalline layers (Squire, 1973) and were further confirmed in more recent cryo-EM myosin filament structures (Hu *et al.*, 2016; Daneshparvar *et al.*, 2020; Rahmani *et al.*, 2021).

Mutations at specific amino acids in the heptad coiled-coil could provide insight into the mechanism for the disease pathology. Mutations in *a* and *d* regions may prevent myosin dimerization, mutations in *e* and *g* may also prevent myosin dimerization but may also alter the stability of the myosin molecule. Another possibility is the mutation in *b, c* and *f* may prevent myosin molecules from forming myosin filaments. Mapping of all LMM mutations from the extant literature was performed and was grouped according to their amino acid position to describe the trend between the position of the mutation in the heptad repeat and their individual diseases (Fig 1.8C, Appendix 1.1). From this analysis, both LDM and MSM patients show many mutations in the hydrophobic *a* and *d* region and the charged amino region *b, c* and *f* (Fig 1.8C, Appendix 1.1). This indicates that most mutations in *MYH7* have the potential to affect either the myosin molecule dimerization or myosin filament formation. Whilst mutations affecting *a* and *d* or *b, c,* and *f* lead to skeletal muscle diseases LDM and MSM, data from amino acid positioning alone does not explain whether mutations affecting particular amino acids in the heptad repeat lead to one disease or the other.

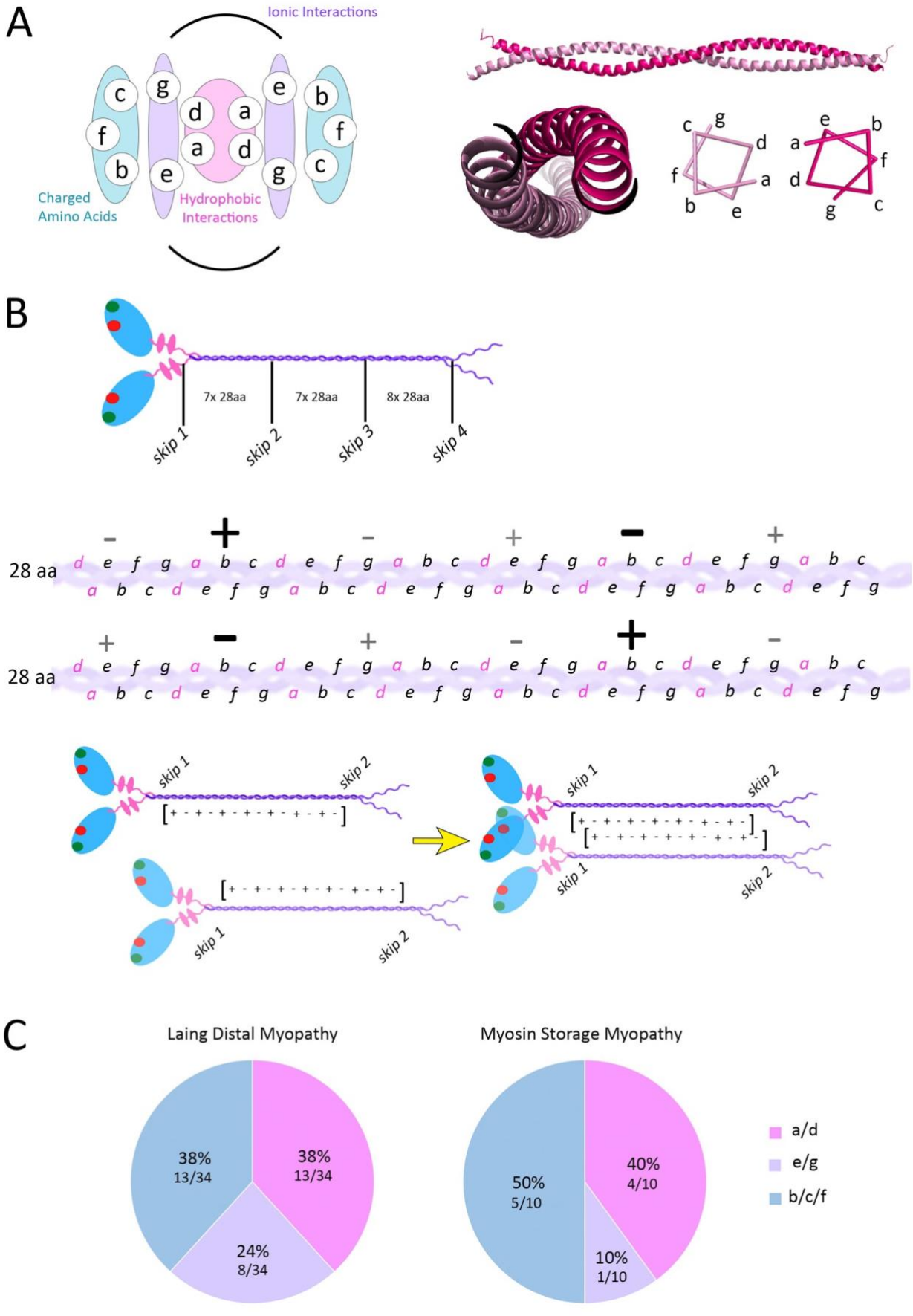


Figure 1.8. Structure of Myosin Class II for thick filament assembly

A) Myosin LMM region forms a heptad amino acid repeat arrangement from *a-g*. Each amino acid plays a role for myosin monomers to dimerise and for myosin dimers to form thick filaments. i) role of amino acids are: *a, d* – hydrophobic interactions (pink), ii) *e, g* ionic interactions (purple) and iii) *b, c, f* – charged amino acids (blue). Both S2 and LMM region show heptad repeat and dimerise through this coiled-coil structure in LMM region **B)** LMM region form a 28 aa pattern of 6 alternating positive and negatively charged amino acids (+ and – symbol). In each 28 aa pattern, the strongest charge occurs in 1b and 3b positions (large bold + and – symbol). There are 7 28aa repeats between skip 1 and skip 2, 7 28aa repeats between skip 2 and 3 and there are 8 28aa repeats between skip 3 and 4. Schematic drawing of strongest charged patterning from 1b and 3b positions of coiled-coil between skip 1 and 2 on myosin LMM enabling myosin molecules to bind together for thick filament assembly through amino acid charges. **C)** Mutations in functional amino acids associated with LDM and MSM. Percentage calculated from the proportion of patients with a mutation in the selected region compared with the total number of disease patients (LDM or MSM).

1.4.5. Role of LMM region for MyHC I head functioning

Myosin filaments function within sarcomeres, myofibrils and myofibres where myosin is present in three main states: i) active, ii) disordered relaxed state (DRX) and iii) super relaxed state (SRX). (Alberts *et al.*, 2015). In the active state, MyHC acts as a molecular motor within a kinetic cycle by interacting with actin and converting chemical energy of ATP hydrolysis into mechanical force and motion, thus generating muscle contraction (Sweeney and Houdusse, 2010). Structural features of MyHC S1 head indicate the functional role for each step in the myosin actin kinetic cycle. Myosin head consists of a large upper 50 kDa cleft where actin and ATP binding sites are found. As ADP is released, this cleft closes and binds to actin tightly (Yengo *et al.*, 1999; Volkmann *et al.*, 2000; Coureux *et al.*, 2003). At the C-terminus of the myosin S1 head, essential light chains bind in this region which elongates the alpha helical structure of the neck, this may aid in neck movement for power stroke for force production (Rayment and Holden, 1994). Myosin functional elements in S1 region include relay loops for connecting the molecule together in addition to two binding regions for both ATP and actin, the former consists of: ATP binding loops and switch 1 and 2, whereas the latter consists of actin binding loops, loop 1-4 and HCM loop (Fig 1.9A). MyHC S1 region can be subdivided into four subunits: the upper 50 kDa (U50 kDa), lower 50 kDa (L50 kDa), N-terminal subdomain and the lever arm. The four subunits in the S1 head connects to S2 region via the lever arm, this can be observed from a side view (Fig 1.9B). These four subdomains are linked by four highly conserved connectors: SH1 helix, relay, switch 2 and strut (Fig 1.9B) (A.T.Geisterfer-Lowrance *et al.*, 1990). The S1 helix and relay domain play an important role to connect to the converter at the start of the S2 region to induce conformational changes for actin binding and release via L50 kDa subunit. Before power stroke, the 50 kDa cleft partially closes near switch 2, this enables the hydrolysed phosphate to be held in place (Fig 1.9B) (Yount *et al.*, 1995; D'Agostino *et al.*, 2011). In the absence of ATP when myosin is in rigor state, both the outer cleft (closest to loop 2) and the inner loop (closest to switch 2) are closed, which enables protein to strongly hold onto actin (Coureux *et al.*, 2003). The cleft is able to open and close across the myosin actin kinetic cycle, this may be controlled through ATP binding site opening and closing or

through alterations in the conformational change in β-sheet transducer (Málnási-Csizmadia *et al.*, 2005).

Although the S1 head is intricate in structure to generate mechanical force for muscle contraction, the LMM region also plays a role in controlling myosin head positioning during muscle contraction and relaxation. When myosin is in the relaxed state, there have been observations of a DRX and SRX (Fig 1.10A) (McNamara *et al.*, 2015). The conventional "J" motif seen in myosin filament structures describes myosin molecules in their SRX state (Fig 1.10B). During the SRX state, myosin heads interact with each other and both heads block actin and ATP binding sites by folding towards the S2 region into a form called the "interacting heads motif" (IHM) (Alamo *et al.*, 2017; Woodhead and Craig, 2020). Myosin is stabilised in the SRX state by a second protein MyBP-C where MyBP-C connects to myosin at two sites, the N-terminus of MyBP-C connects to the myosin head region and C-terminus of MyBP-C connects to the myosin LMM (Luther *et al.*, 2008; Spudich, 2015). During the DRX state, myosin heads spring away from the thick filament backbone and protrude towards the actin filaments (Zhao, Padrón and Craig, 2008; Wilson *et al.*, 2014). The main difference between myosin in the DRX state compared with the SRX state is that myosin in an SRX state consumes ATP at a slower rate in comparison to DRX and much slower than during the contractile state (Fig 1.10A) (McNamara *et al.*, 2015). There has been a link between HCM mutations affecting the N-terminal MyBP-C binding site on *MYH7* that has led to destabilising myosin in the SRX state (Alamo *et al.*, 2017; Toepfer *et al.*, 2020). In chapter 3, I analyse the proportion of SRX and DRX myosin molecules in the presence and absence of *MYH7* mutations to identify whether the C-terminal MyBP-C binding domain also plays a role in stabilising myosin in the SRX state.

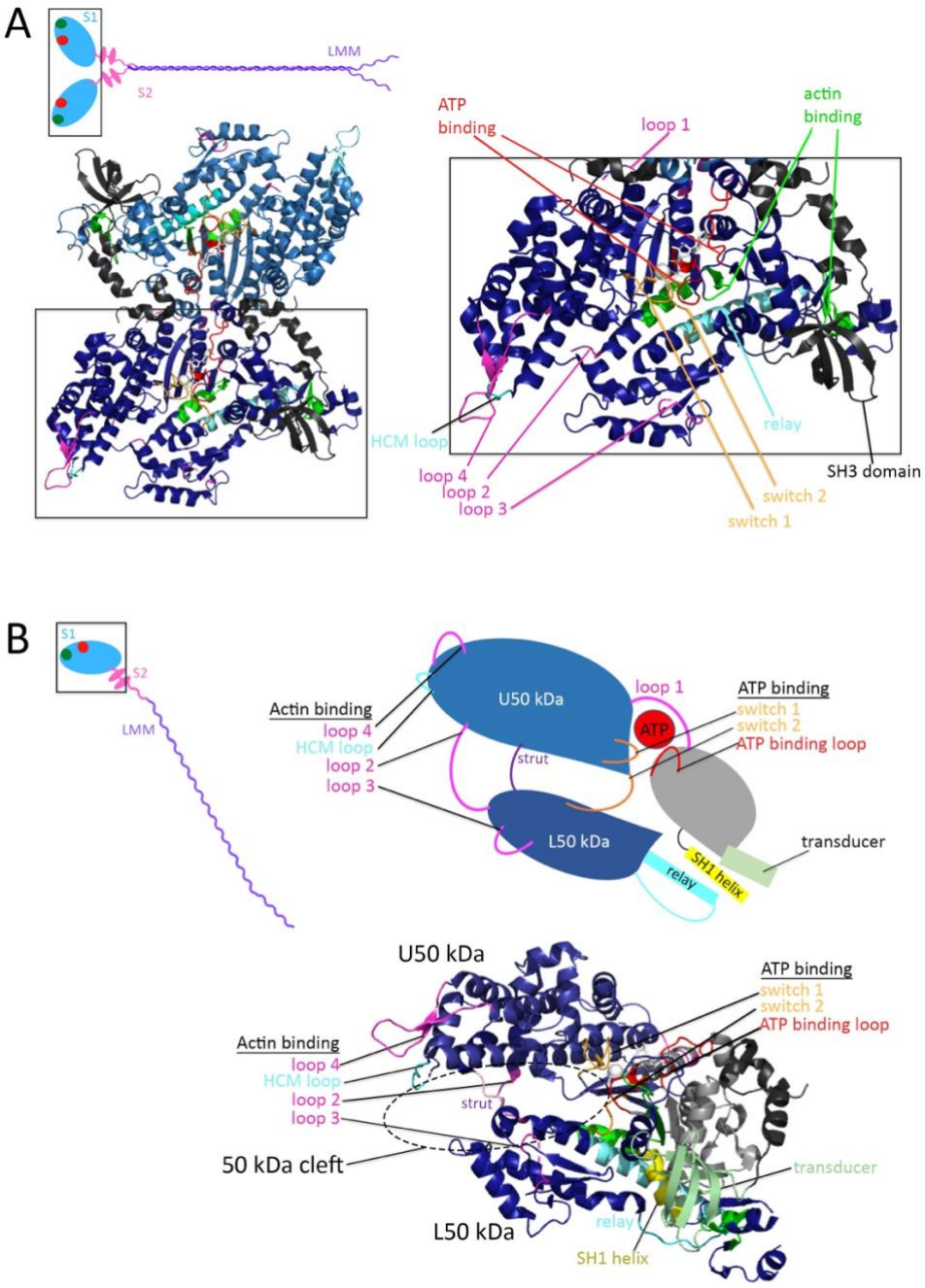


Figure 1.9. Structure of S1 myosin head.
Functional elements of the human cardiac myosin head. **A)** Top view of myosin S1 head. Schematic diagram of myosin molecule on top left and box indicates the location of S1 head for structural ribbon view. Close up of lower myosin S1 head labelled with the positions of functional domains: ATP binding loops (red), HCM loop (cyan) and loop 1-4 to aid actin-binding (pink), main actin-binding loops (green), switch 1 and 2 for ATP binding (orange) and relay loops for connecting molecule together. **B)** Side view of myosin S1 head. Schematic diagram of myosin molecule on top left and box indicates the location of S1 head for schematic diagram of head and its structural ribbon. S1 head can be subdivided into 4 subdomains, 1. Upper 50 kDa subdomain (U50 kDa) and 2. Lower 50 kDa subdomain (U50 kDa) with the presence of the 50 kDa cleft in between where actin binds, 3. N-terminal peptide, 4. Lower connector for lever arm and S2 connection. Actin binding is modulated between loops 2-4, the HCM loop and the main actin-binding loops in (A), ATP binding is modulated between ATP binding loops and switch 1/2. Relay and SH1 helix main function for twisting S2 head for power stroke. Diagram in B adapted from Sweeney and Houdusse, 2010 using protein 4DB1 from Protein data bank (https://www.rcsb.org/structure/4DB1).

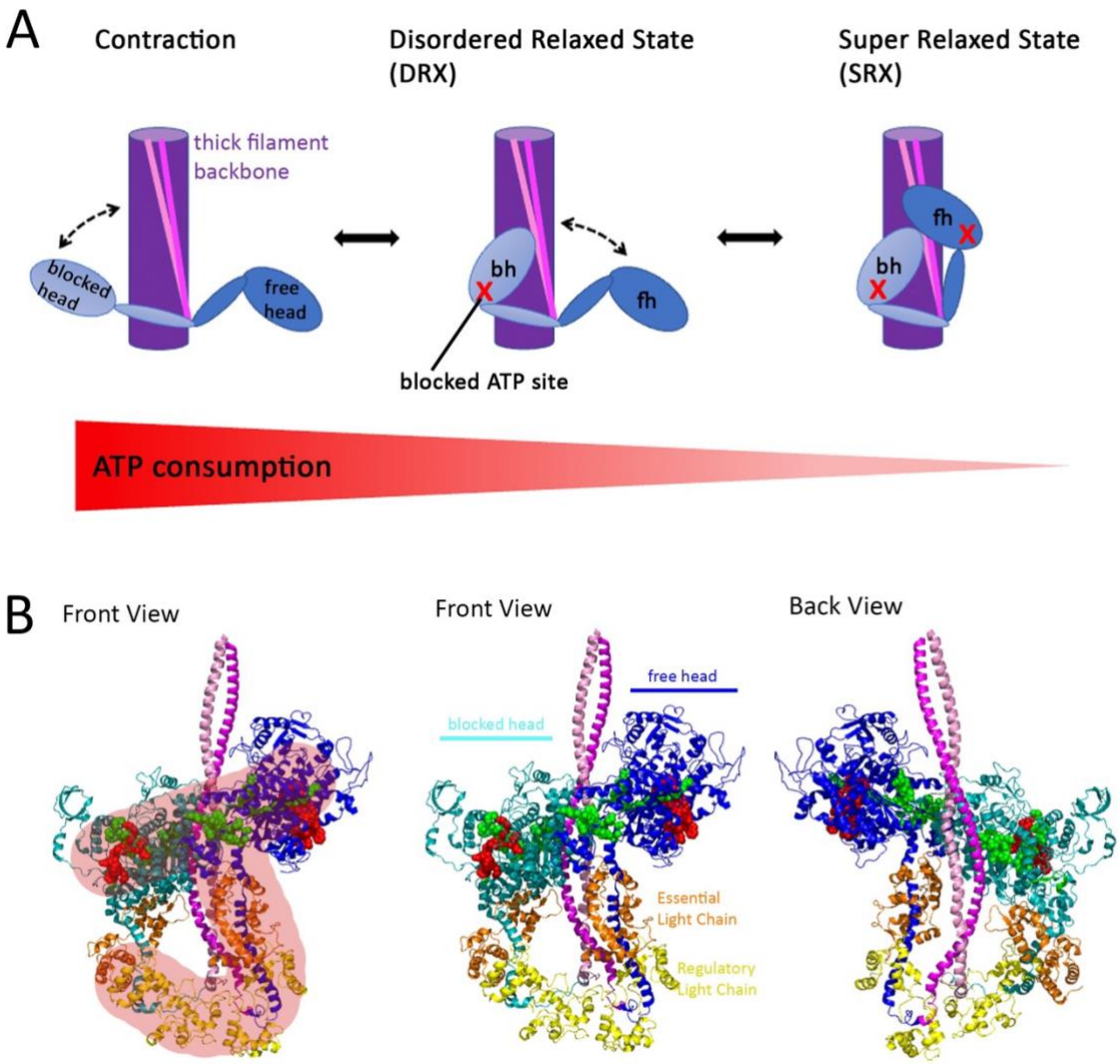


Figure 1.10. Myosin conformation in contracting, DRX and SRX state
A) Schematic diagram showing myosin molecules in 3 states: Contracting, DRX and SRX states. Blocked head represents the blocking of ATP site during SRX and DRX state. During contracting state, both blocked head and free head are available for ATP hydrolysis and ready for actin binding in the myosin actin kinetic cycle. During DRX state, blocked head is bound to the thick filament backbone and unable to hydrolyse ATP, free head remains flexible and able to hydrolyse ATP. When myosin is in SRX, both heads are bound to the thick filament backbone and both are unable to hydrolyse ATP. Energy consumption lowers from transition from contracting myosin to DRX to SRX state as ATP site availability decreases during the transition. **B)** Front and Back view of myosin in SRX state in ribbon representation. Myosin is helically ordered with the appearance of a tilted 'J' motif (left). Free head represents availability of ATP site during DRX but not available during SRX. Essential light chain (orange) and regulatory light chain (yellow) binds at the bottom of the 'J' motif. Panel A adapted from Garfinkel, Seidman and Seidman, 2018 and panel B adapted from Woodhead *et al.*, 2005.

1.5. Zebrafish as a model LDM and MSM

Currently there are a few *in vitro* human cell culture and *in vivo* animal models for LDM and MSM. Cell culture models for MSM have modelled mutations R1845W and H1901L. Cells transfected with R1845W and H1901L show accumulation of MyHC aggregates that did not incorporate into thick filaments and demonstrated that mutations R1845W and H1901L affected the 29 residue assembly complex domain for myosin filament formation (Dahl-Halvarsson *et al.*, 2017). Although transfection

experiments in human cell lines can generate identical *MYH7* mutations from human patients in human *MYH7* gene, experiments have only shown overexpression of mutant *MYH7* to healthy *MYH7* and are not representative of the biological expression level of *MYH7 in vivo* (Dahl-Halvarsson *et al.*, 2017). There are several *in vivo* models such as *C. elegans* and *D. melanogaster*.

C. elegans MSM models have modelled mutations by creating alleles in the *unc-54* gene with mutations R1845, E1883K and H1901L (Dahl-Halvarsson *et al.*, 2017). *Unc-54* is the major MyHC gene expressed in body wall muscles (Tajsharghi, Pilon and Oldfors, 2005). *C. elegans* lacking *unc-54* gene led to paralysis of body wall muscles and could be rescued in the presence of wild type UNC-54. Introducing either of the three mutant alleles of the *unc-54* gene (R1845, E1883K and H1901L) there was partial rescue of motility but not to full rescued effect observed when introduced with wild type *unc-54* (Dahl-Halvarsson *et al.*, 2017). Although *C. elegans* is an easy model organism for the introduction of various alleles to identify the mechanism leading to disease, the *unc-54* gene may not represent as an accurate ortholog to human *MYH7*. *C. elegans* is an invertebrate model organism where there is no skeletal muscle. Although *C. elegans* do contain striated muscle, *unc-54* is not a skeletal MyHC and thus by utilising *unc-54* to represent slow skeletal muscle may be inaccurate. There are alternative MyHCs in *C. elegans* such as *unc-15, myo1, myo2* and *myo3* that are localised in a subset of muscle or thick filament structure but neither of these muscles accurately resemble slow MyHC in humans (Miller, Stockdale and Karn, 1986).

D. melanogaster MSM models have modelled mutations by creating alleles in the *Mhc* gene with mutations L1793P, R1845W, and E1883K (Viswanathan *et al.*, 2017). *Mhc* mutant alleles L1793P R1845W, and E1883K were transgenically expressed in *Mhc* null background to study developmental defects in the presence of defective myosin molecules. Indirect flight muscle (IFM) fibres show the presence of MyHC aggregates from 1-day old flies with a decline in muscle architecture in adult flies but not during early myofibrillogenesis during the pupae stage (Viswanathan *et al.*, 2017). *In vitro* assembly assay to identify myosin polymerisation into thick filaments show that MSM mutant myosin was not able to form thick filaments as efficiently as wild type myosin when subjected to differing salt concentrations (Viswanathan *et al.*, 2017). Additionally, *in vitro* studies subjecting thick filaments to proteolysis have shown that MSM mutant myosin form thick filaments that were less stable in comparison to wild type thick filaments. MSM mutant myosin were therefore unable to form thick filaments as readily and stably as wild type myosin. *D. melanogaster* LDM models have modelled mutations by creating alleles in the *Mhc* gene with mutations K1729del (Dahl-Halvarsson *et al.*, 2018). *Mhc* mutant allele K1729del were generated using CRISPR/Cas9 genome editing to target *Mhc* gene

and use homologous recombination (HR) to insert a short single oligonucleotide (ssoligo) containing mutant allele K1729del to insert mutation into the *Mhc* gene. Homozygous mutants show high death rate and heterozygous larvae show reduced muscle function. Reduced lifespan, lack of flight, impaired jump movement and reduced overall movement were observed (Dahl-Halvarsson *et al.*, 2018). At larvae stage, Z-disks and M-bands appear to be reduced in heterozygous mutants and are difficult to distinguish in homozygous mutants. At older stages, myosin thick filaments show distinct A bands, however appear faint in muscle fibres, however myosin accumulated in certain areas of the thick filament (Dahl-Halvarsson *et al.*, 2018). Thus, *D. melanogaster Mhc* mutant allele K1729del show progression of disease with varying degree of severity, a similar phenotype to clinical diagnostics of human LDM patients. Although both MSM and LDM models in *D. melanogaster* have demonstrated instability of myosin filament formation in the presence of MSM mutations and the progression of muscle disease in the presence of LDM mutations, representation of the single *Mhc* gene in *D. melanogaster* as the ortholog to human *MYH7* may be inaccurate.

To study how mutations in *MYH7* pathologically lead to diseases, I will be studying the mechanisms by which mutations in *MYH7* lead to developmental defects. Currently, there are no vertebrate animal models for MSM or LDM targeting the orthologous gene to human *MYH7*. I will be using zebrafish as my animal model as there are several advantages to studying muscle developmental defects compared to alternative animal models. Zebrafish larvae are transparent with optical clarity, allowing direct visualisation of muscle formation, growth, and function. From 0 to 5 days post fertilisation (dpf), zebrafish larvae do not feed and utilise their yolk as their source of nutrition, thus removing feeding as a factor in muscle development. Zebrafish can frequently produce large clutch sizes of 100-300 embryos which enable large sample sizes to be analysed. Large clutch sizes also enable ease of CRISPR/Cas9 genome editing as injections at the one-cell stage can be made quickly and efficiently (Maves, 2014). In chapter 4, I demonstrate the need to identify the zebrafish ortholog to human *MYH7*. In chapter 5, I determine the role of zebrafish ortholog of human *MYH7* in sarcomere assembly using loss of function experiments in the quest to identify the importance of slow MyHC for the overall muscle structure and function.

1.5.1. Structure of zebrafish skeletal muscle

There are many structural similarities between zebrafish muscle to mammalian muscle. Zebrafish muscle spans from the trunk to the tail and is organised into roughly 30 chevron-shaped blocks called somites separated by connective tissue (Fig 1.11A). Each somite consists of skeletal muscle fibres that bundle to span across the length from posterior to anterior (Fig 1.11B). When viewing the zebrafish

trunk in a transverse cross-sectional view, the somites surround the neural tube, which forms the spinal cord, and notochord, a rod like structure which supports the developing embryo (Fig 1.11C). Zebrafish muscle fibres differentiate during development, muscle fibres are separated into compartments rather than mixed in muscle bundles. Slow fibres are mononucleated and are organised to be on the superficial layer of the somite and the fast muscle is multinucleated and is buried underneath in the deeper tissue (Fig 1.11B). Since fast and slow fibres are in such distinct location, identifying slow MyHC defects in zebrafish prove advantageous as fast and slow muscle fibres can be identified and analysed with ease.

1.5.2. Somatogenesis in zebrafish

In teleosts, to form muscle there is a subdivision step to form the blocks of cells called somites from the anterior to the posterior end of the body. From this subdivision, somites form epithelia to separate from the notochord and neural tube (Chevallier, Kieny and Mauger, 1977; Schröter *et al.*, 2008). At around 10.5 hours post fertilisation (hpf), the first pair of somites are formed. Following this first somite, another pair of somites is added every 15 min until around 8 somites have been, and then every 30 min until 30 somite pairs have been formed by the 24 hpf (Stickney, Barresi and Devoto, 2000). In zebrafish, slow muscle fibres are the first to differentiate and, following this, fast muscle develops later. The first 20 slow fibres are differentiated next to the notochord (van Raamsdonk *et al.*, 1982; Devoto *et al.*, 1996). These cells are signalled through hedgehog (Hh) proteins to differentiate into slow muscle fibres (Blagden *et al.*, 1997; Lewis *et al.*, 1999; Bryson-Richardson *et al.*, 2005). These slow cells first have an elongation step followed by migration to form a superficial monolayer of slow fibres of each somite (Fig 1.11) (Devoto *et al.*, 1996; Blagden *et al.*, 1997; Daggett *et al.*, 2007). At the horizontal myoseptum, near the notochord, there is a subset of cells called the slow muscle pioneers (Fig 1.11) (Waterman, 1969; van Raamsdonk *et al.*, 1982) which lie on the border between the dorsal and ventral divisions of the somite (Felsenfeld, Curry and Kimmel, 1991). Muscle pioneers express engrailed proteins and are also regulated by the Hh signalling present near the horizontal myoseptum (Wolff *et al.*, 2004). Fast muscle fibres are then differentiated through fgf8 signalling (Groves, Hammond and Hughes, 2005). Following formation, muscle fibres fuse and finalise the organisation of the sarcomeres (Kimmel *et al.*, 1995). In adult fish, slow muscle fibres are mainly found in a wedge along the lateral side of the horizontal myoseptum, whereas the fast muscle fibres make up the majority of the myotome. Intermediate muscle fibres are found in between the slow and fast fibres (Stone Elworthy *et al.*, 2008; Nord *et al.*, 2014). There are two groups of muscle stem cells, one set is found in the external cell layer and is responsible for fibre formation at the external layer of the myotome. These muscle stem cells express *pax3, pax7* and *met* and are active for fibre formation after

24 hpf (Hammond *et al.*, 2007; Gurevich *et al.*, 2016; Nguyen *et al.*, 2017). The other set of muscle stem cells is found deeper within the myotome and proliferate at early stages before 24 hpf (Knappe, Zammit and Knight, 2015; Pipalia *et al.*, 2016; Roy *et al.*, 2017).

During the formation of myofibrils, the sequence of events in sarcomere assembly in the developing zebrafish embryo is less well understood. There have been several studies to model aspects of sarcomere assembly described earlier. The initial steps of myofibrillogenesis have been described as an anchoring step at the cell periphery, close to the myotendinous junction (MTJ) (Kelly and Zacks, 1969; Tokuyasu, 1989) where there are integrin adhesion sites to connect thin filaments to the MTJ (Pardo, Siliciano and Craig, 1983; Ervasti, 2003; Quach and Rando, 2006). Thin filaments and Z-disks form I-Z-I bodies accumulate and aggregate at the MTJ (Tokuyasu and Maher, 1987). In chapter 5, I describe the role of slow MyHC in myofibrillogenesis to identify the mechanism behind LDM and MSM in early muscle development (Sanger *et al.*, 2009).

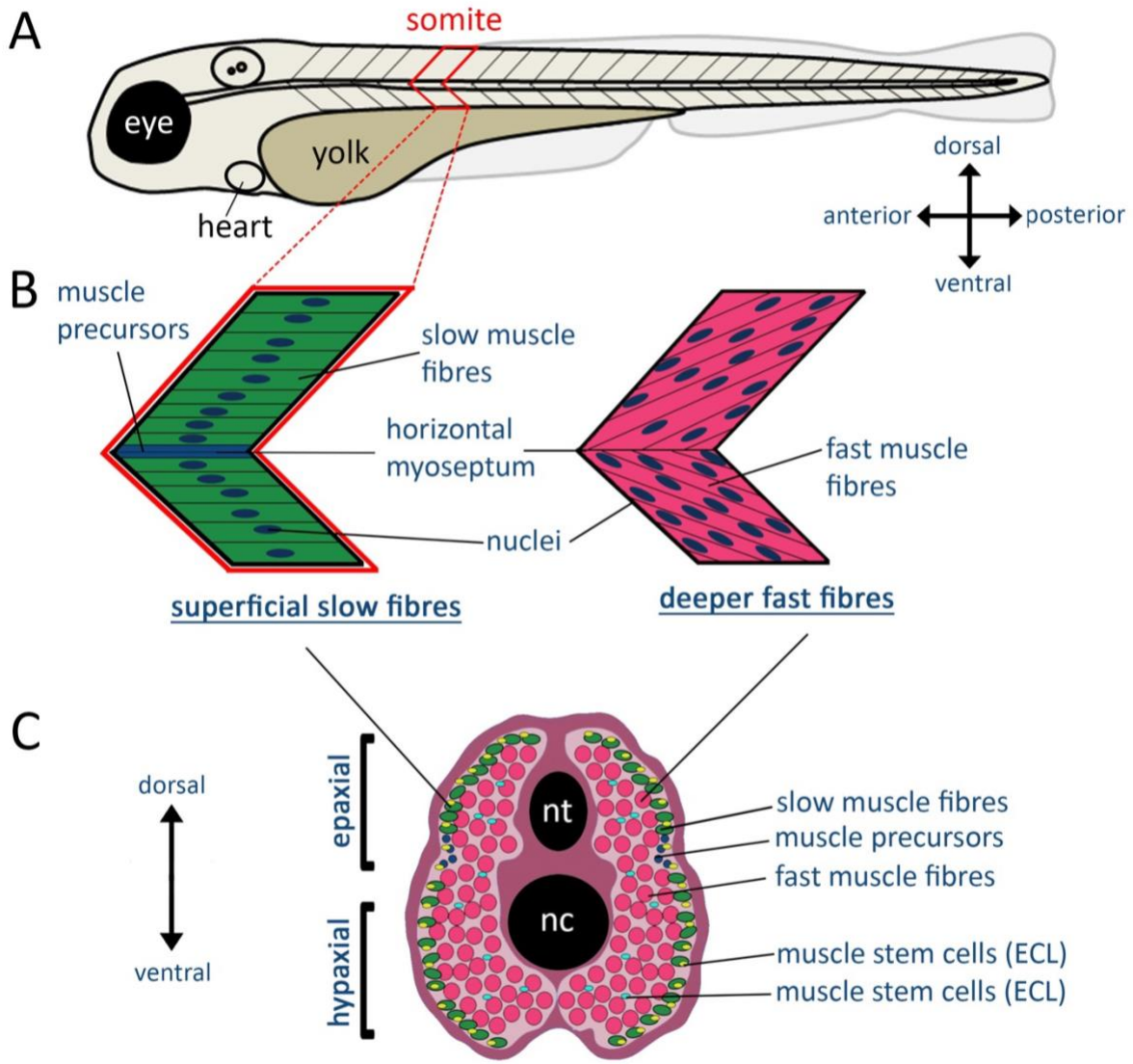


Figure 1.11. Muscle composition of zebrafish larvae
A) Schematic of 3 dpf zebrafish larvae from lateral perspective, anterior left and posterior to the right with dorsal top and ventral bottom. Zebrafish labelled with eye, heart, yolk and somite highlighted in red. **B)** Schematic of zebrafish somite showing superficial slow fibres (left) where they are mono nucleated and arranged in parallel in a horizonal appearance when viewed from lateral perspective. Fast fibres (right) are multinucleated, are found deeper than slow fibres and appear at an angle when viewed laterally. At the centre line of the somite is the horizontal myoseptum where the slow muscle precursors are found. **C)** Cross section (YZ angle) of zebrafish trunk show fast and slow cell populations at their locations. Slow fibres shown at the superficial layer and fast fibres seen in the deeper layers. On the external cell layer (ECL) muscle stem cells can be observed. Another set of muscle stem cells can be observed in the deeper layers, between the fast fibres. Nt, neural tube, nc, notochord.

1.6. Summary

The two congenital myopathies that I have focused on in the present work, Laing Distal Myopathy (LDM) and Myosin Storage Myopathy (MSM) are due to sarcomeric gene mutations in *MYH7* (Lamont *et al.*, 2014; Parker and Peckham, 2020). Although there are currently no curative medicines for *MYH7*-related congenital myopathies and available treatments simply target the various symptoms (*Myosin storage myopathy*, 2016; Topaloglu, 2020). The aim of this thesis was to study the potential underlying molecular and cellular mechanisms leading to LDM and MSM by identifying primary biophysical defects in human fibres obtained from affected patients and developing zebrafish models that investigated developmental defects in the quest for treatment design for *MYH7*-related diseases. In chapter 3, I investigate the primary biophysical defects in the presence of human *MYH7* mutations to assess whether there was a change in myosin filament length or a change in myosin head positioning in the presence of defective myosin molecules. My main findings were the following: 1) There is no overall alteration in sarcomere organisation in the presence of defective myosin molecules. 2) Mutations affecting the *MYH7* MyBP-C binding domain destabilise the SRX state.

In chapter 4, I identify the fish equivalent genes that can be accurately described the orthologous genes to human *MYH7* to target for the generation of an accurate disease model in the future. In chapter 5, I generate *smyhc1* knockout models using CRISPR/Cas9 genome editing to understand the role of zebrafish *smyhc1* in sarcomere assembly. My main findings were the following: 1) Zebrafish genes *smyhc1-5, myh7* and *myh7l* are orthologous to mammalian *MYH7*. 2) Loss of *smyhc1* in zebrafish results in defective sarcomere organisation at the early stages of development, indicating the role of *smyhc1* in sarcomere assembly to elongate the myofiber. 3) Transitional role of *MYH7* from early developmental stages to adulthood remains in question. Overall, current data give early insight into the mechanism for the role of slow myosin in sarcomere assembly. Work ongoing to generate large deletion of *smyhc* locus to understand the role of slow MyHC in sarcomere assembly during early developmental stages through to adulthood.

Chapter 2

Materials and Methods

2.1. Human Muscle Biopsy Samples

All the human muscle biopsy specimens have been taken from various European clinical laboratories.

The use of these samples has been ethically approved (ethics approval REC 13/NE/0373)

Table 2.1. List of human samples from healthy controls and patients with disease mutations in MYH7.
The mutation map of these patients is presented in Fig.2.1.

Patient	*Mutation*	*Gender*	*Age*	*Disease*
C1	None	Male	55	-
C2	None	Female	25	-
C3	None	Female	63	-
C4	None	Male	37	-
C5	None	Female	44	-
P1	p.STOP1936Leu	Male	57	Myosin storage myopathy
P2	p.Thr304Ser	Female	60	Dilated cardiomyopathy and Distal myopathy
P3	p.Arg1845Trp	Male	55	Myosin storage myopathy
P4	p.Leu594Met	Female	30	Distal myopathy
P5	p.Arg453His	Female	47	Hypertrophic cardiomyopathy and Distal myopathy
P6	p.Met982Thr	Male	38	Myosin storage myopathy
P7	p.Ala1882Glu	Female	27	Laing distal myopathy
P8	p.Ala1440del	Female	15	Distal myopathy
P9	p.Glu1508del	Female	68	Laing distal myopathy
P10	p.Ala1603Pro	Female	44	Distal myopathy
P11	p.Glu1610Lys	Female	44	Hypertrophic cardiomyopathy, nemaline myopathy and distal myopathy
P12	p.Lys1617del	Female	22	Laing distal myopathy
P13	p.Ala1636Pro	Male	45	Laing distal myopathy
P14	p.Lys1729del	Female	53	Laing distal myopathy
P15	p.Leu1492Pro	Female	16	Laing distal myopathy
P16	p.Leu1467Pro	Female	35	Distal myopathy
P17	p.Thr441Met	Male	10	Distal myopathy
P18	p.Glu1669del	Female	49	Distal myopathy
P19	p.E1507del	Male	26	Distal myopathy

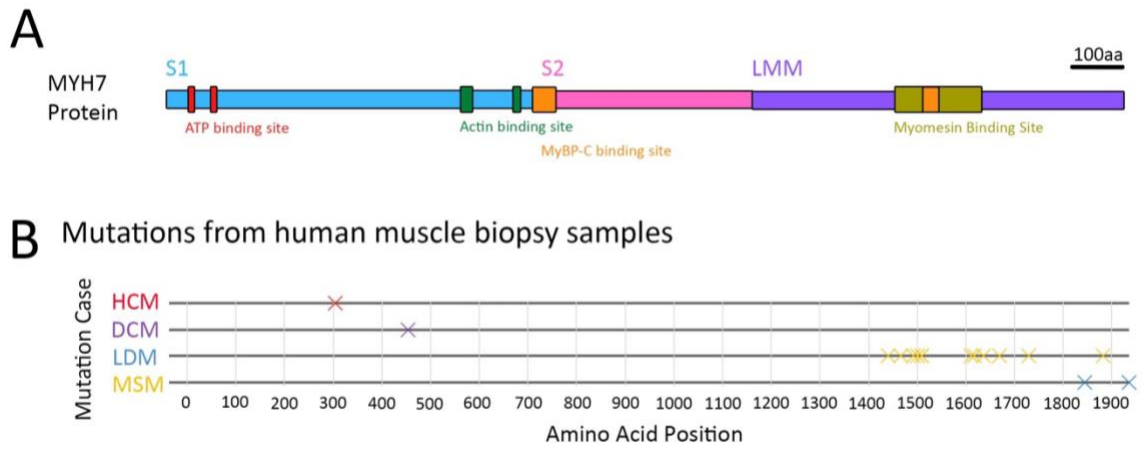


Figure 2.1 Map of MYH7 protein and plotted mutations of patient samples in Table 2.1

Table 2.2. Primary antibodies

Primary Antibody/Small molecule	*Target*	*Type*	*Raised in*	*Raised against*	*Dilution*	*Method Used*
MF20	All MyHC	IgG2b	Mouse	Chicken	1:10	2.2.1.
A4.951	Slow MyHC	IgG	Mouse	Human	1:50	2.2.1.
S58	Slow MyHC	IgA	Mouse	Chicken	1:5	2.8.2.
F59	Slow MyHC	IgG	Mouse	Chicken	1:5	2.8.2.
BA-D5	Slow MyHC	IgG2b	Mouse	Rat	1:5	2.8.2.
α-actinin	α-actinin	IgG1	Mouse		1:500	2.8.2.
F310	Fast MyHC	IgG1	Mouse	Chicken	1:5	2.8.2.
A4.1025	All MyHC	IgG2a	Mouse	Human	1:10	2.8.2.
MF20	All MyHC	IgG2b	Mouse	Chicken	1:500	2.8.2.
Phalloidin488	F-actin				1:50	2.8.2.
Phalloidin405	F-actin				1:50	2.8.2.

Table 2.3. Secondary antibodies

Secondary Antibody	*Target*	*Dilution*
IgG-Alexa488	Goat anti-mouse	1:1000
IgG-Alexa555	Goat anti-mouse	1:1000
IgG1-Alexa555	Goat anti-mouse	1:1000
IgG1-Alexa488	Goat anti-mouse	1:1000
IgG2a-Alexa488	Goat anti-mouse	1:1000
IgG2a-Alexa594	Goat anti-mouse	1:1000
IgG2b-Alexa488	Goat anti-mouse	1:1000
IgG2b-Alexa594	Goat anti-mouse	1:1000
IgA-FITC	Goat anti-mouse	1:500

2.2. Human Biopsy Assays

2.2.1. Measurement of myosin filament length

Muscles were treated with skinning solution (100 mM KCl, 5 mM MgCl2, 5 mM EGTA, 10 mM Imidazole, 50% Glycerol, pH 7.5) at 4°C for 24 hours and then transferred to -20°C. Preparations of single muscle fibres were made by dissecting skinned muscle bundles. Muscle fibres from human biopsy samples were prepared and mounted onto microscope slides at room temperature (Fig 2.2). Each muscle fibre was clamped to half-split copper meshes designed for electron microscopy (SPI G100 2010C-XA, width, 3 mm). Prepared slides were then used for antibody staining (Hooijman, Stewart and Cooke, 2011).

Antibody staining

Fibres were mounted on microscope slides with clamps and stretched between two clamps. Fibres were then fixed with 4% PFA for 15 min at room temperature. After fixing, fibres were washed 3x 5 min in PBS and then permeabilized using 0.1% Triton (in PBS) for 10 min at room temperature. After permeabilization, fibres were washed 3x 5 min in PBS and then blocked in 10% goat serum in PBS for 1 h at room temperature. 10% goat serum was then replaced with the primary antibody in their working dilutions (see Table 2.2) and incubated at 4 °C overnight. Fibres were washed 3x 5 min in PBS followed by incubation with secondary antibody in their working dilutions (see Table 2.3) at room temperature for 3 h. Fibres were washed 3x 5 min PBS to remove excess antibody and then immersed in Fluoromount. Coverslip was placed onto slide without disturbing the position of the fibre and sealed with nail varnish. Confocal images were taken using Cell Voyager at 60x magnification.

Image analysis

Images were analysed using the DDecon plugin for ImageJ, a program that background corrects images by deconvolution and computes filament lengths with a precision of 10-20 nm by measurements of peak positions of fluorescently labelled filaments (Fig 2.2). Thick filament measurements were made by analysing observed scan intensities and modelled intensities (Gokhin *et al.*, 2012).

Statistical Analysis of Myosin Filament Length

If there were a small change detected from my samples, an estimated expectation of change was 20% with an effect size of 1.5. With $\alpha = 0.05$ and power $(1-\beta) = 0.80$, the minimum required sample size was 7 fibres per group (G-power analysis). Measurements taken from one fibre were pooled and averaged by calculating mean value to plot one point on the graph. All statistical computations were performed using Prism 8 (GraphPad). One-way ANOVA was used to identify differences when

comparing fibres between two individuals (controls or patients). Mean filament length measurements from each fibre from each patient were all pooled together and treated as a separate group for comparison to controls as each mutation was different from one patient to another. All data were expressed as mean ± standard deviation (SD).

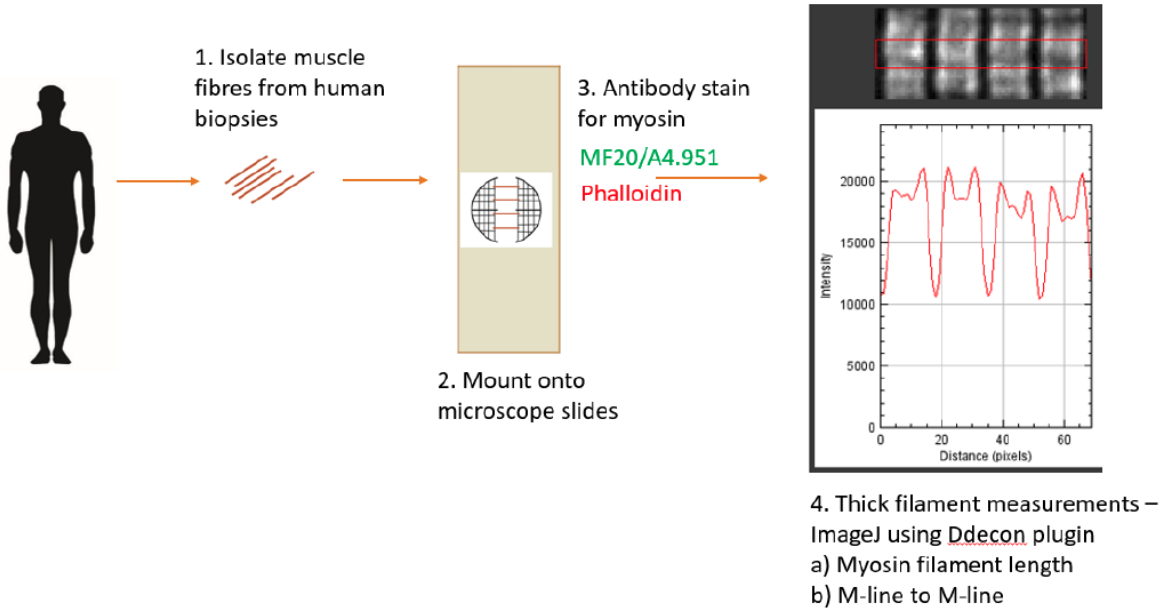


Figure 2.2. Methods for thick filament measurements.
1. Isolating muscle fibres from controls and patients in table 2.1. **2.** Nine muscle fibres per human sample were mounted onto microscope slides. **3.** Antibody staining for either a) fast and slow myosin (MF20) or b) staining for slow specific (A4.951) and actin (phalloidin) were made using antibodies in Table 2.2. **4.** Images were taken on using Cell Voyager at 60x magnification

2.2.2. Identifying proportion of SRX and DRX state in muscle fibres

SRX experiments

Fibres were prepared as Methods 2.2.1 and the experiment was followed as in Toepfer *et al.*, 2018. Fibres were dissected and mounted onto glass coverslip as in methods 2.2.1. The flow chamber was made by using three layers of double-sided sticky tape, fibre side up and a coverslip was placed on top. Before imaging, fibres were washed with 100 uL Rigor buffer (120 mM K acetate, 5 mM Mg acetate, 2.5 mM K_2HPO_4, 2.5 mM KH_2PO_4, 50 mM MOPS, 5 mM, EGTA, 1 mM TCEP, pH 6.8) 5 times over 5 mins to remove ATP, BDM and glycerol. Fibres were then incubated with 100 uL mATP buffer (Rigor Buffer + 250 uM Mant-ATP, pH 6.8) for 5 min. Fibres were imaged at 25 °C using a Zeiss epifluorescence microscope (20x/0.8 Plan Apo objective). Fibre was located using brightfield light to avoid photobleaching of mATP. Fibres were excited at 395 nm (DAPI setting) at 20% laser power, 20% shift and exposure time of 20 ms. At time 0 s, 2 images were taken to set the start point of intensity at 100%. mATP buffer was flushed with ATP buffer (Rigor Buffer + 4 mM ATP, pH 6.8) and fluorescence

decay images were taken using AxioCam Cm1 camera at 5 s intervals for 90 s, then every 10 s until 180 s.

SRX analysis

Images taken from SRX experiments were analysed using ImageJ to record fluorescence intensities (File > Import > Image Sequence > Select the images > OK). A square on the image of the fibre was drawn and the fluorescence intensity of the square was measured (M key to measuring on keyboard). Another square was drawn on the background to measure background fluorescence. To normalise the background, mean background fluorescence was subtracted from the average fibre fluorescence intensity for each image. Then each point was divided by the fluorescence intensity of the final mATP before ATP flush (t=0). Normalised values were exported to Prism GraphPad Software. Decay in fluorescence was then fit into a two-state exponential (Analyse > Nonlinear regression (curve fit) > Exponential > Two-Phase Decay > 95% confidence interval).

$$I = 1 - P_1\left(1-\exp\left(-\frac{t}{T1}\right)\right) - P_2\left(1-\exp\left(-\frac{t}{T2}\right)\right)$$

I = fluorescence intensity
P1 = initial proportion of fluorescence for the fast states
P2= initial proportion of fluorescence for the slow states
T1 = time constants for the lifetime of fast state
T2 = time constants for the lifetime of the slow state

P1 and T1 represent fast decay in the DRX state and P2 and T2 represent slow decay in the SRX state. This assay shows an approximate proportion of 40% of myosin in the DRX state and 60% of myosin in the SRX state (Alamo *et al.*, 2017; Christopher N. Toepfer *et al.*, 2019).

Statistical Analysis for identifying proportion of SRX and DRX myosin states

If there were a small change detected from my samples, there was an estimated expected change of 20% with an effect size of an effect size of 1.5. With $\alpha = 0.05$ and power $(1-\beta) = 0.80$, the minimum required sample size is then 7 fibres per group (G-power analysis) to identify such change. Measurements taken from one fibre were pooled and averaged by calculating mean value to plot one point on the graph. All statistical computations were performed using Prism 8 (GraphPad). One-way ANOVA was used to identify differences when comparing fibres between two individuals (controls or patients). Mean filament length measurements from each fibre from each patient were all pooled

together and treated as a separate group for comparison to controls as each mutation was different from one patient to another. All data were expressed as mean ± standard deviation (SD).

2.3. Zebrafish maintenance

Zebrafish (*Danio rerio*) were obtained from the Zebrafish Facility at King's College London, Guy's Campus. Wildtype embryos (AB background) were obtained from the mass embryos production (MEPs) facility. All genetically altered zebrafish were created on AB and reared on a 14/10 hr light/dark cycle at 28.5 °C, 22% humidity. To collect embryos for experiments, males and females were initially paired in 1-litre breeding tanks and separated with barriers during the evening; this allows the mating pairs to acclimate to each other overnight. At the onset of light exposure, the following morning, the barriers were removed to enable breeding behaviour. Upon spawning, all embryos were collected using a mesh tea strainer and incubated at 28.5°C in petri dishes containing 1% methyl blue system water to prevent possible bacterial growth. Debris in dishes was removed using a 3 mL Pasteur pipette (Kimmel *et al.*, 1995). All experiments were performed in accordance with guidelines and regulations of the UK Animals (Scientific Procedures) Act 1986.

2.4. Generating zebrafish KO lines using CRISPR/Cas9 system

Zebrafish have been used as a classic model organism for their characteristics of fast growth, transparent embryos, and small size. Zebrafish embryos are fertilised after laying eggs, enabling genetic manipulation at the one-cell stage using morpholino or CRISPR/Cas9 (Gutiérrez-Lovera *et al.*, 2017). Genome editing using CRISPR/Cas9 has produced many genetic manipulations in zebrafish. There have been improvements by optimising the Cas9 protein, which was used in creating mutant zebrafish lines in this project (Hwang *et al.*, 2013). Embryos can be collected after fertilisation and can be used to study the early stages of development as they are not developed *in utero.* Internal structures can be visualised as embryos are transparent, therefore, labelled antibodies or RNA probes could be seen easily. (Gutiérrez-Lovera *et al.*, 2017). Thus, utilising zebrafish as a model organism to create disease models can be used to characterise the early developmental defects upon mutation.

2.4.1. CRISPR/Cas9 system

Gene editing tools have been developed to target specific sites in the genomes of many organisms such as mice, drosophila and zebrafish (Ma and Liu, 2015). Traditionally, genome editing tools such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) were used to target genes. These tools require a string of DNA binding proteins that bind to the target gene, this string of proteins is attached with an endonuclease domain whereby the two domains come in close

proximity to then enable the endonuclease to create a double-strand DNA break (DSB) at the target site. CRISPR/Cas9 genome editing is made by creating a DSB by Cas9 at the target gene site (Fig 2.3). CRISPR-Cas was originally discovered in a microbial adaptive immune system utilising 20 nucleotide RNA guide sequences (crRNA) and a guide RNA (tracrRNA) that enables the cleaving of foreign genetic material from invading viruses. The target DNA should contain a PAM sequence (for Cas9 it is NGG). As a genome editing tool, the crRNA and tracrRNA can be fused together to create a single-guide RNA (sgRNA). This sgRNA can then bind to Cas9 to direct the binging towards the target sequence by designing the 20 nucleotide sequence within the sgRNA. Upon binding of Cas9-sgRNA to target DNA, the Cas9 protein undergoes a conformational change and induces a DSB ~3 bp upstream of the PAM sequence. After a DSB is made, pathways for DNA damage repair are activated (Ran *et al.*, 2013).

There are two different pathways: non-homologous end joining (NHEJ) or homologous recombination (HR) (Ran *et al.*, 2013). NHEJ involves Ku proteins to be recruited at the free ends of DSB DNA to enable the recruitment of DNA-PKcs. XRCC4 and Ligase IV are recruited to re-ligate the DNA ends. NHEJ is an error-prone mechanism for DNA repair whereby INDEL mutations are likely to occur (Chang *et al.*, 2017). The alternative repair mechanism is HR where there is a requirement for a template DNA strand from another chromosome. This process ensures that the DNA repair is made without mistakes at the site of damage. The initial steps in HR consist of pre-synapsis whereby double-strand break is detected and resection occurs at the 5’ end to generate 3’ single-stranded DNA ends. Rad51 and Rad52 are recruited at 3’ single-stranded tails which then lead to the homology search and annealing of the complementary template DNA strand from another chromosome. This forms a double Holliday junction between damaged DNA and the complementary template DNA. Template DNA is then used as a reference for repairing damaged DNA. (Jasin and Rothstein, 2013). To make specific mutations, an introduction of an oligonucleotide containing desired could be used and by HR, insert into the genome.

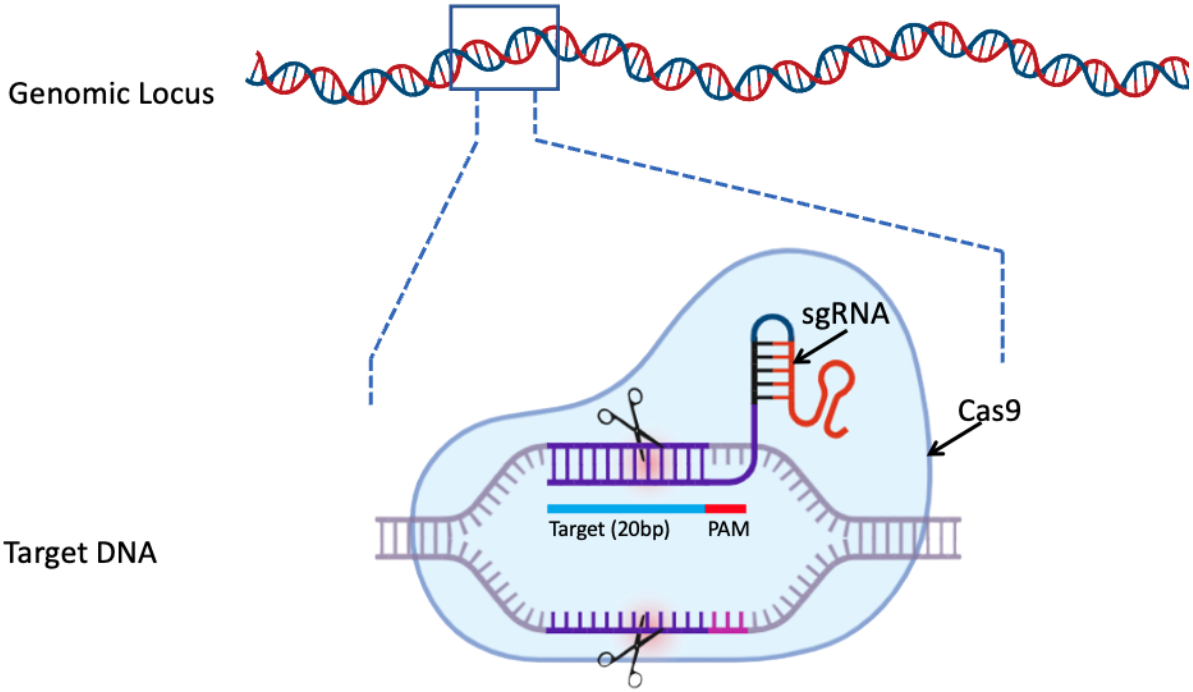


Figure 2.3. CRISPR/Cas9 mechanism to create site-specific mutations.
Cas9 protein from *S. pyogenes* (yellow) binds to the gRNA scaffold (red) and the sgRNA-Cas9 complex binds to the target sequence (blue). Cas9 creates double-strand breaks (red triangles) 3bp upstream of the 5'NGG PAM sequence. Image adapted from Ran *et al.*, 2013.

2.4.2. Identifying potential target sites for CRISPR gene editing

To identify potential target sites for CRISPR knockout, CRISPR Direct (http://crispr.dbcls.jp/) was used to identify targets with information on sequence specificity to minimise the chances of off-target mutations. Highly specific gRNA was selected based upon 20 base pair target sequence and also 12mer+PAM site which bind 12 bases adjacent to PAM sequence. The whole cDNA sequence of *smyhc1* was screened for potential target sites (Fig 2.4). Oligonucleotides were ordered for the insertion into expression vector pDR274 (Table 2.4).

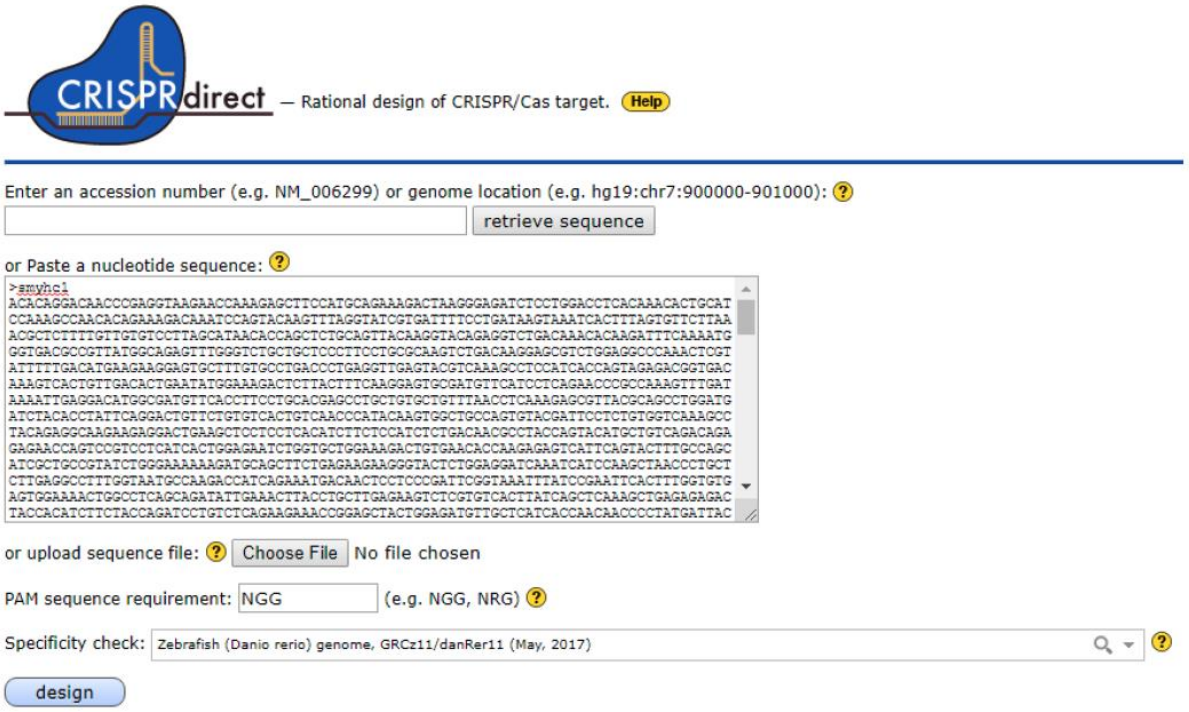


Figure 2.4. Identifying specific target sites in *smyhc1* using CRISPR Direct.
Interface of CrisprDirect (http://crispr.dbcls.jp/). The DNA sequence of *smyhc1* was entered into the text box. Zebrafish (*Danio rerio*) genome, GRCz11/DanRer11 was selected for the specificity check so possible off-targets can be analysed to ensure sequence comparison was using the most updated zebrafish gene database available.

Table 2.4. *smyhc1* CRISPR/Cas9 target sites

gRNA name	*Gene*	*Exon*	*Target Sequence +PAM (NGG or NG using Cas9BE3)*	*Oligonucleotides for insertion into pDR274*
gRNA smyhc1 KO1	smyhc1	2	CATGTCAAAAATACGAGTTTGGG	Oligo 1: 5'- TAGGTGTCAAAAATACGAGTTT -3' Oligo 2: 3'- ACAGTTTTTATGCTCAAACAAA -5'
gRNA smyhc1 KO2	smyhc1	4	ACCACAGAGGAATCGTACACTGG	Oligo 1: 5'- TAGGCACAGAGGAATCGTACAC -3' Oligo 2: 3'- GTGTCTCCTTAGCATGTGCAAA -5'
gRNA E1508del	smyhc1	29	CAGAGGAAATCTCTGACCTTACT	Oligo 1: 5'- TAGGTAAGGTCAGAGATTTCCT -3' Oligo 2: 3'- ATTCCAGTCTCTAAAGGACAAA -5
gRNA K1617del	smyhc1	30	TCTCAGACTGAAGAAGAAGATGG	Oligo 1: 5'- TAGGTCAGACTGAAGAAGAAGA -3' Oligo 2: 3'- AGTCTGACTTCTTCTTCTCAAA -5
gRNA K1729del	smyhc1	32	GCTGAATCAGAAGAAGAAGCTGG	Oligo 1: 5'- TAGGTGAATCAGAAGAAGAAGC -3' Oligo 2: 3'- ACTTAGTCTTCTTCTTCGCAAA -5
gRNA E1856K	smyhc1	34	CTGAAGAAGACCGTAAGAATCTG	Oligo 1: 5'- TAGGGAGGAAGACCGTAAGAAT -3' Oligo 2: 3'- CTCCTTCTGGCATTCTTACAAA -5

2.4.3. gRNA synthesis for smyhc1 KO

To synthesise gRNA plasmid DR274 was used, this plasmid contains a T7 promotor for gRNA synthesis and a gRNA scaffold next to the target sequence (Fig 2.5). To use this plasmid, pDR274 is digested with BsaI restriction enzyme (37°C, 24h; reaction mix: 1 µg of pDR247 vector, 5 µL of BsaI-HF enzyme, 5 µL of 10X NEB 4 buffer, ddH2O to 50 µl) and purified using Qiagen's QIAquick PCR Purification kit (Qiagen, #28106) and can be stored at -20 °C. BsaI cuts the pDR274 vector creating 'sticky ends' for the insertion of an oligonucleotide (oligo) sequence. The target sequence is flanked with sticky ends 5'- TAGG on oligo 1 and 5' – CAAA on oligo 2 to enable insertion into BsaI digested pDR274. Once the target sequence is ligated into the vector, the plasmid may be used to synthesise gRNA using T7 RiboMAX Large Scale RNA Production kit. Oligonucleotides of the target sequence were ordered from IDT Technologies (Table 2.4). The separate oligos were annealed together using 10x annealing buffer (0.4 M Tris-HCl, 0.2 M $MgCl_2$, 0.1 M NaCl, 10 mM EDTA). Oligonucleotide annealing reaction mix (5 uL Annealing buffer 10x, 1 uL 100 µM CRISPR oligo 1, 1 uL 100 µM CRISPR oligo 2, 43 uL H_2O) is heated to 95 °C for 5 min, then cools -1 °C each 30 s down to 25 °C and then incubated at 4 °C. Annealed oligos can be stored at -20 °C. Annealed oligos are then ligated into digested pDR274 (2h, RT; Reaction mix: 1 µL Digested pDR274 [5 µg], 3 µL annealed CRISPR oligos, 1 µL T4 DNA ligase, 5 µL 2X ligase buffer).

After ligation, 5 µL of ligation mix was used to transform 50 µL of NEB 5-α competent *E. coli.* Firstly, NEB 5-alpha competent *E. coli* were taken from -80 °C and thawed on ice. 5 µL of plasmid (pDR274+oligo ligated) was added to 50 µL *E. coli* and gently agitated. Cells were placed on ice for 30 min, then heat shocked at 42°C for 45 s and then replaced on ice for 5 min. 500 µL of SOC was added to cells aseptically and incubated at 37°C in an orbital shaker for 1 h. 100 µL and 200 µL of cells were aseptically plated to agar plates containing 30 µg/µL kanamycin and incubated at 37 °C overnight. 4 single colonies were inoculated in 10 mL of Luria broth (LB) containing 30 µg/µL kanamycin (4 colonies per gRNA were inoculated) and incubated overnight at 37°C in an orbital shaker at 225 rpm. Cells were pelleted using the centrifuge at 13, 000 *x g* for 10 min. The supernatant was removed, and the plasmid was isolated and purified using Qiagen Miniprep Kit. Minipreps were stored at -20°C. Insertions of oligos were checked by sequencing using M13 (-21) Forward Primer `5'–TGTAAAACGACGGCCAGT–3'`.

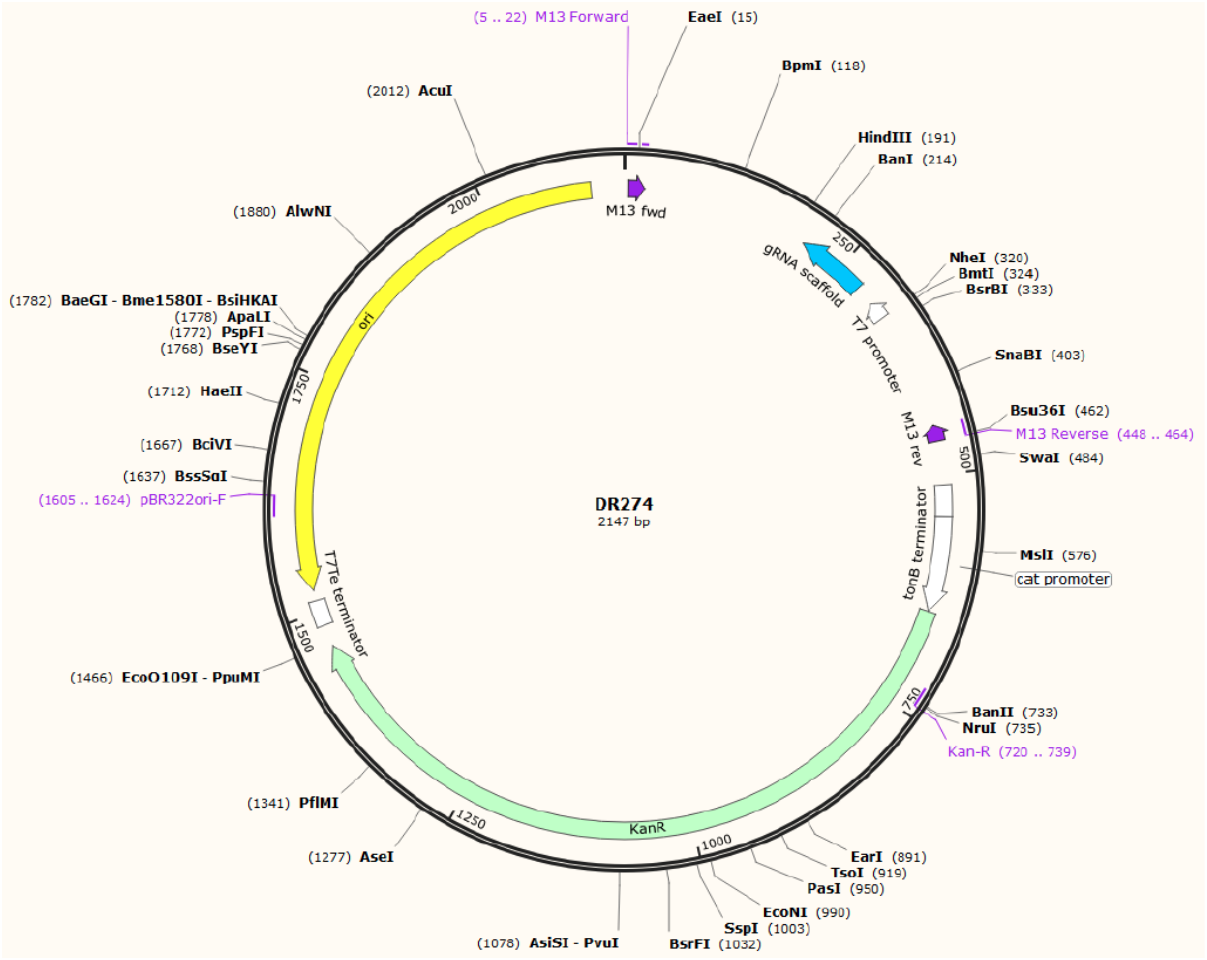


Figure 2.5. Plasmid DR274 to synthesise single guide RNA for CRISPR/Cas9 systems.
A) pDR274 contains a T7 promoter for *in vitro* RNA production and a kanamycin resistance gene for the selection of positive clones. B) pDR274 was digested with BsaI (shown in red) for insertion of the target sequence into the plasmid. Before sgRNA synthesis, the plasmid was cut with DraI to enable 285bp sgRNA to be synthesised at the T7 promotor and end after the gRNA scaffold sequence. The plasmid was obtained from Addgene.

pDR274 containing target sequence was digested using restriction enzyme DraI (Reaction mix: 5 µg pDR274+target sequence, 1 µL DraI RE, 2 µL 10x Tango Buffer and add ddH_2O to a final volume of 20 µL) Reaction mix was incubated at 37°C for 24 hrs. Linearized plasmid was then isolated and purified using a Qiagen purification kit and eluted at 25 µL. Concentration was measured using nanodrop.

Linearized pDR274+target sequence was used as the template DNA to produce gRNA with T7 RiboMAX Large Scale RNA Production kit (Promega). All reagents were thawed on ice and RNA synthesis was performed as followed: 0.5 µg template DNA, 8 µL 5X T7 transcription buffer, 12 µL 25 mM rNTPs mix, 4 µL T7 enzyme mix and ddH_2O up to 40 µL. The reaction mix was incubated at 37°C overnight. DNA template was removed by digesting with 1 µL QR DNase for 30 min. 2 µL sample of reaction was taken to run on 1% agarose gel.

RNA was purified using the phenol/chloroform/IAA method. The reaction was scaled up using water to at least 100 µL before starting. 100 µL phenol/chloroform/IAA was added to the reaction and vortexed followed by centrifugation at 13, 000 *x g* for 2 min. The upper phase is taken into a new tube. 10 µL 0.1 Volume of 3M Sodium acetate pH 5.2 and 1 volume of Isopropanol was added to the mix and placed on ice for 5 min. The reaction was centrifuged at 13, 000 *x g* for 10 min. The supernatant was poured off, leaving the pellet in the tube. Pellet was washed with 1 mL of 70% ethanol and then left to air dry for 5 min, then the pellet was resuspended in nuclease-free water (25 µL) and placed at 55°C for 5 min. 2 µL sample was taken and run on 1% agarose gel next to the 2 µL taken at the end of the sgRNA synthesis reaction. RNA concentration was measured using Qbit. Guide RNA was stored at -80°C in 5 µL aliquots.

2.4.4. CRISPR injections – smyhc1

Injection of genome editing tools at single-cell embryos increases the likelihood of injection mixture entering every cell. Injection needles were made using heat-treated capillary tubes that were pulled to form fine needles. Injection needles were attached to a microinjector to inject 1 nL of CRISPR injection mixture into each embryo. An injection mixture was prepared for each mutation type in Table 2.5. Embryos were provided from paired fish with known genotypes for each target gene. Male and female fish were paired the evening before and placed in 1 L breeding tanks with a separating barrier between the paired fish. In the morning of injections, upon light stimulation barriers were removed to enable breeding behaviour. Embryos were collected using a tea strainer and injected. Embryos were screened for 3 h and 24 h after injection to check the success of injections by analysing the fluorescence of rhodamine dextran (Fig 2.6).

Table 2.5. Injection mixtures prepared for smyhc1 KO or HR

Mutation	**Reagents**	**Volume (µL)**	**Final concentration**
Knock Out 1 (KO1)	gRNA KO1 (87.7ng/µL) Cas9 Protein (3220ng/µL) 5% Rhodamine dextran ddH20	4 0.5 0.5 0	~80pg per embryo ~300pg per embryo
Knock Out 2 (KO2)	gRNA KO1 (87.7ng/µL) Cas9 Protein (3220ng/µL) 5% Rhodamine dextran ddH20	2 0.5 0.5 2	~40pg per embryo ~300pg per embryo
K1617del	gRNA K1617del (87.8 ng/µL) Cas9 Protein (3220 ng/µL) ssoligo K1617del (20 µM) 5% Rhodamine dextran ddH20	1 0.5 0.5 0.5 2.5	~30pg per embryo ~300pg per embryo
K1729del	gRNA K1617del (53 ng/µL) Cas9 Protein (3220 ng/µL) ssoligo K1729del (20 µM) 5% Rhodamine dextran ddH20	0.5 0.5 0.5 0.5 3	~12.5pg per embryo ~300pg per embryo
E1856K	gRNA K1617del (132 ng/µL) Cas9 Protein (3220 ng/µL) ssoligo E1856K (20 µM) 5% Rhodamine dextran ddH20	2 0.5 0.5 0.5 1.5	~80pg per embryo ~300pg per embryo

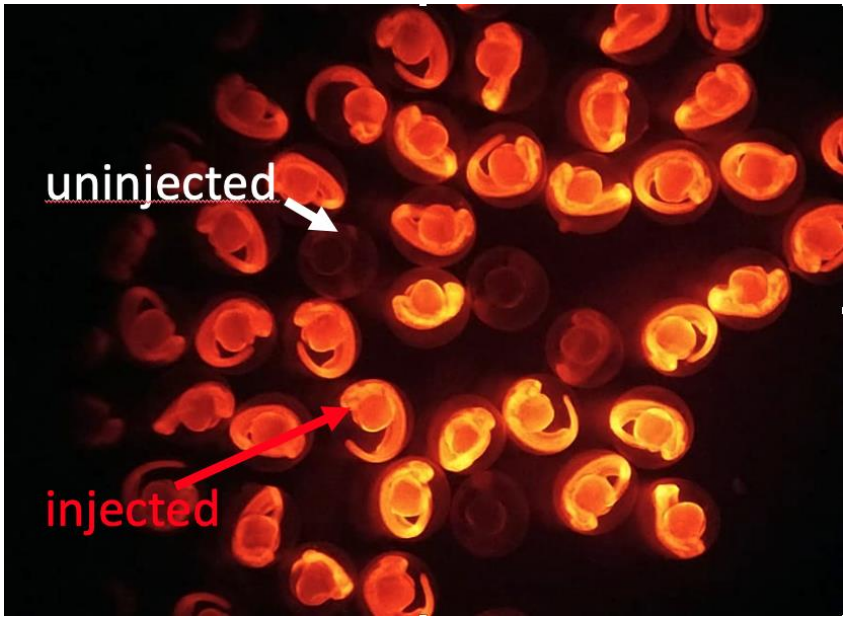


Figure 2.6. Analysis of injected embryos at 24 hours post-injection.
Red fluorescence indicates successful injection whilst lack of fluorescence in embryos shows unsuccessful injections.

2.5. Generating smyhc2-5 KO mutants using Alt-R CRISPR Cas9 system

2.5.1. smyhc2-5 deletion gRNA design

The whole cDNA sequence of *smyhc2 and smyhc5* were screened for potential target sites using CRISPRdirect. Target sequences for *smyhc2* and *smyhc5* Table 2.6 were used to generate the deletion of *smyhc2-5.* Alt-R CRISPR-Cas9 crRNA with specific target sequences in Table 2.7 were ordered from Integrated DNA Technologies (https://eu.idtdna.com/pages/products/crispr-genome-editing/alt-r-crispr-cas9-system). Alt-R crRNAs and tracrRNA were dissolved to 100 µL with ddH_20. crRNA:tracrRNA duplex were made (95°C, 5 min, Reaction mix: 1 µL crRNA, 1 µL tracrRNA and 3 µL Duplex buffer) to make a final concentration of 20 µM. Following heat treatment, crRNA:tracrRNA duplex was cooled to room temperature for a further 5 min. AltR CRISPR-Cas9 mix was assembled as described in Table 2.7 and heated to 37 °C for 10 min and cooled to room temperature for 5 min before injection. Injections of 1 nL were made at the one-cell stage and were reviewed at 24 hpf for successful injections with the presence of red fluorescence with rhodamine dextran as shown in Figure 2.6.

Table 2.6. smyhc2-5 CRISPR/Cas9 target sites

crRNA description	*Gene*	*Exon*	*Target Sequence +PAM*	*Strand*
smyhc2 KO1	Smyhc2	3	Acaatattgaacgcttattcagg	+
smyhc2 KO2	Smyhc2	5	GAGGTGGTCGTTGCCTACAGAGG	+
smyhc5 KO1	Smyhc5	1	GTATCTCAGGAAGTCGGACCGGG	+
smyhc5 KO2	Smyhc5	36	GCAGCTTACGGAACTTGGTCAGG	-

Table 2.7. Injection mixtures prepared for *smyhc2-5* KO
[NEB Cas9 EnGen® Spy Cas9 NLS , 20 µM is equal to 3.22 mg/ml (3220ng/ul).

Mutation	*Reagents*	*Volume (µL)*
CRISPR A – smyhc2 ex5 to smyhc5 ex1	crRNA+tracrRNA mix – smyhc2 KO2 (20 µM)	1
	crRNA+tracrRNA mix – smyhc5 KO1 (20 µM)	1
	EnGen-Cas9 buffer x10	0.5
	EnGen Cas9 protein (3220ng/µl)	0.5
	5% Rhodamine Dextran (Invitrogen, #D1816)	0.5
	dH20	0.5
		Total: 5
CRISPR A – smyhc2 ex5 to smyhc5 ex36	crRNA+tracrRNA mix – smyhc2 exon 5 KO2 (20 µM)	1
	crRNA+tracrRNA mix – smyhc5 exon 36 KO2 (20 µM)	1
	EnGen-Cas9 bufferx10	0.5
	EnGen Cas9 protein (3220ng/µl)	0.5
	5% Rhodamine Dextran	0.5
	dH20	0.5
		Total: 5

2.6. Genotyping

2.6.1. DNA Extraction using alkaline lysis method

DNA can be extracted from embryos or fin clips from adult fish using an alkaline lysis method. Single or pooled embryos or fin-clip 30 µL alkaline lysis buffer (25 mM NaOH, 0.2 mM EDTA) and heated to 95°C for 1 h. The reaction was stopped by adding 30 µl of neutralisation buffer (55 mM Tris-HCl, pH 8). Samples were spun using microfuge and 1-2 µL were used for PCR or HRM. DNA solutions are stored at 4 °C.

2.6.2. Primer design

Primers were designed for both HRM and sequencing using Primer3 Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) where the size of fragment and annealing temperature can be adjusted. Primers were designed to ideally be 20-22 nucleotides long and with an annealing temperature of less than 60 °C. known single nucleotide polymorphisms were considered and avoided to prevent primers from not annealing and to ensure HRM and PCR results are generated. Primers were designed for CRISPR mutations were targeting (Table 2.8). HRM PCR amplification of 100 bp fragment containing CRISPR target site. Sequencing primers were designed to give approximately 500 bp fragments that included the target site and both HRM primers. All primers are mapped onto *smyhc1* in Appendix 2.1.

Table 2.8. Primer list for HRM and sequencing

Smyhc1 – exon 2	**Sequence**
Forward HRM (amplicon size 107 bp)	5'-CGCAAGTCTGACAAGGAGC-3'
Reverse HRM	5'-GTGATGGAGGCTTTGACGTAC-3'
Forward Sequencing (amplicon size 603 bp)	5'-CCTGTGCTGTTCCTTTTCTCA-3'
Reverse Sequencing	5'-CCATGAGACTGTGTTGGCTG-3'
Smyhc1 – exon 4	**Sequence**
Forward HRM (amplicon size 115 bp)	5'-TCTGTGTCACTGTCAACCCA-3'
Reverse HRM	5'-AGTTCTCACCTGACAGCAT-3'
Forward Sequencing (amplicon size 280 bp)	5'-TGAGTGATGAACGTTGAGCC-3'
Reverse Sequencing	5'-AAATGAGGGAAGTTTTGTGCAT-3'
Smyhc1 – exon 30	**Sequence**
Forward HRM (amplicon size 106 bp)	5'-GAATCAGAGACTCGCAGCAG-3'
Reverse HRM	5'-ATGCCTGCCTGTTAGCCTG-3'
Forward Sequencing (amplicon size 801 bp)	5'-GCAGAGATCCAGACAGCCTT-3'
Reverse Sequencing	5'-ACATGGACAGTGTTGACATTCA-3'
Smyhc1 – exon 32	**Sequence**
Forward HRM (amplicon size 115 bp)	5'-TGAATGTCAACACTGTCCATGT-3'
Reverse HRM	5'-GCCTCCTCAACCTCAGTCTG-3'
Forward Sequencing (amplicon size 280 bp)	5'-TGACACACCTGTATTAGTAAACT-3'
Reverse Sequencing	5'-TTTCAGTAGCTTACCCTGGC-3'
Smyhc1 – exon 34	**Sequence**
Forward HRM (amplicon size 115 bp)	5'-ACACATACAGAAAACGATGAAGT-3'
Reverse HRM	5'-TTCAGCTGCAGTTTGTCCAC-3'
Forward Sequencing (amplicon size 450 bp)	5'-TCAGGCATTTTCTCTTCACACA-3'
Reverse Sequencing	5'-ACACAGGGACAAACAAAACATCA-3'
Smyhc2 – exon 5	
Forward HRM (amplicon size 173 bp)	5'-ACAATCAGGAGGTGGTCGTT-3'
Reverse HRM	5'-tgacgtgcccacaaaatcaa-3'
Forward Sequencing (amplicon size 800 bp)	5'-tcgtcatctcttccgcagAT-3'
Reverse Sequencing	5'-ttgacgtgcccacaaaatca-3'

Smyhc5 – exon 1	
Forward HRM (amplicon size 149 bp)	5'-ATGGAGGAGTTTGGAGCTGC-3'
Reverse HRM	5'-TCCCGACTGATAATGCTTCCT-3'
Forward Sequencing (amplicon size 493 bp)	5'-cgctttccttgtctggtgtg-3'
Reverse Sequencing	5'-atattccagagagccgtgca-3'
Smyhc5 – exon 36	
Forward HRM (amplicon size 168 bp)	5'-GGCTgtgagtgcttttctgg-3'
Reverse HRM	5'-AGCAATATCAGCCCTCTCCTC-3'
Forward Sequencing (amplicon size 414 bp)	5'-GTGGACAAACTGCAGCTGAA-3'
Reverse Sequencing	5'-ggagccttaacacttgcacc-3'

2.6.3. High-Resolution Melt Analysis

HRM PCR analysis is a method in which mutations, polymorphisms, and epigenetic changes can be detected in double-stranded DNA. The Vii™ 7 Real-Time PCR System was used in analysing MicroAmpR Optical 348- well plates and Applied Biosystems Melt Dr™ HRM Master Mix. PCR fragments (from DNA extracted from embryos or fin clips) amplified in this analysis were around 100 bp at the target site where the predicted mutation takes place. At the PCR step, DNA sequences are intercalated with a fluorophore in Melt Doctor master mix during the melt and anneal phase of PCR. This fluorophore can then be detected and measured during HRM analysis. As amplified DNA sequences anneal, they anneal to one another according to their proportional abundance. In wild type +/+ DNA sample, both + DNA strands have the same sequence, thus creating a perfect +/+ homoduplex. In heterozygous +/- DNA samples, there will be 50% + DNA strands and 50% - DNA strands. When annealing occurs, there will be 25% homoduplex +/+ wild type, 25% homoduplex -/- mutant and 50% +/- heteroduplex mutant. After the annealing of the DNA, the HRM process begins. Samples were slowly heated from 50 to 95 °C leading to the separation of DNA strands at their melting point. Fluorophores highly fluoresce when intercalated between bases of dsDNA and fluoresce much less when bound to ssDNA. Double-stranded DNA melt at different temperatures due to difference in melting points of duplexes. In wild type +/+ DNA samples, there will be 100% +/+ homoduplexes and have the same melting temperature. Thus, the melt curve will have one step. In heterozygous +/- DNA samples, there will be three steps to the melt curve due to the presence of 3 different duplexes. Heteroduplexes are the least stable and melt first, then the less-stable mutant -/- homoduplex, followed by wild type +/+ homoduplex. Between the two homoduplexes, there will be a small shift in melting temperature as the stability between the two are very similar.

HRM Mix for each well:

Melt Dr Master Mix (2X)	5 µL
HRM Forward Primer (Table 2.8)	0.4 µL
HRM Reverse Primer (Table 2.8)	0.4 µL
ddH_2O (double distilled water)	3.72 µL
Extracted genomic DNA	1 µL
	----------
	10 µL

2.6.4. Sanger Sequencing

The target site for gene sequencing was amplified by PCR to send for sequence analysis. Primers for *smyhc1* exons were designed and found in Table 2.8.

PCR reaction mix was made as below:

Polymerase Buffer (5x)	4 µL
Forward Primer (10µM)	0.4 µL
Reverse Primer (10µM)	0.4 µL
dNTPs (10µM)	0.5 µL
Polymerase (GoTaq, Phusion or Q5)	0.2 µL
ddH_2O	12.7 µL
Extracted genomic DNA	1 µL
	----------
	20 µL

Depending on which polymerase, the PCR cycling step will be specific to each enzyme. Here are the three cycling steps I used depending on the enzyme I used.

2.5.4.1. GoTaq® DNA Polymerase (M300) on thermocycler for PCR DNA amplification:

Step 1	95 °C	2 min
Step 2	95 °C	30 sec
Step 3	Tm-5 °C	30 sec (Tm found in Table 2.8)
Step 4	72 °C	1 min/kb (Repeat step 2-4 35x)
Step 5	72 °C	7 min
Step 6	4 °C	∞

2.5.4.2. Phusion ® High-Fidelity DNA Polymerase (NEB, M0530) on thermocycler for PCR DNA amplification:

Step 1	98 °C	30 sec
Step 2	98 °C	10 sec
Step 3	Tm+3 °C	30 sec (Tm found in Table 2.8)
Step 4	72 °C	30sec/kb (Repeat step 2-4 35x)
Step 5	72 °C	7 min
Step 6	4 °C	∞

2.5.4.3. Q5® High-Fidelity DNA Polymerase (M0491) on thermocycler for PCR DNA amplification:

Step 1	98 °C	30 sec
Step 2	98 °C	10 sec
Step 3	Tm+3 °C	30 sec (Tm found in Table 2.8)
Step 4	72 °C	30sec/kb (Repeat step 2-4 35x)
Step 5	72 °C	2 min
Step 6	4 °C	∞

PCR product was screened for presence or absence of amplified DNA. 5 µL of PCR product and reference 6 µL 100 bp ladder was loaded onto 2% agarose gel (containing SafeView, NBS Biologicals, NBS-SV1). Gel electrophoresis was set at 100V for 30 min. Gel analysed using GelDoc gel imager under UV lamp. The remaining 15 µL of PCR product was purified using Qiagen PCR Purification Kit (Qiagen, #28106) or Exo-CIP™ Rapid PCR Cleanup (NEB, #E1050) following the manufacturer's instructions. Purified DNA samples to be sent for sequencing were added with 2 µL of forward sequencing primers according to the *smyhc1* exon amplification. Sequences were sent to Genewiz or Eurofins for sequence analysis on Snapgene viewer (ver.3.1.2, GSL, Biotech).

2.7. Visualising Gene Expression

2.7.1. RNA extraction

Pools of 10-15 fish larvae were placed in a microfuge tube with 100 µl Tri-reagent (Sigma Aldrich, #T9424) and manually homogenised by physical abrasion with a tissue grinder (Thermofisher Scientific, #12-141-363). Probes were cleaned using 70% EtOH between each sample. RNA was separated using Phenol:Chloroform:Isoamyl (Merck, P3803) and vortexed for 5 sec and incubated for 10 min at room temperature. Samples were then centrifuged 13, 000 *g* at for 10 min, 4°C. RNA is

present in the top aqueous phase and transferred into a new microfuge tube and purified using a RNeasy mini kit (Qiagen, #74104).

2.7.2. cDNA synthesis

Template DNA for anti-sense RNA probe synthesis were made from the zebrafish cDNA library. cDNA synthesis using the oligo dT first-strand method with the SuperScript III Reverse Transcriptase kit (Invitrogen, #12087539).

In a nuclease free tube, the following were mixed and incubated for 65°C for 5 min, then placed in ice for 1 min. Mix was then centrifuged at 13, 000 *x g* for 1 min:

Oligo(dT)15 (50 ng/ µL)	0.5 µL
Total RNA (1 µg)	X µL
ddH_2O	11.5- X µL
dNTP mix (10 mM)	1 µL
----------	
13 µL	

the following were then added to the mix:

5x First Strand buffer (Superscript III)	4 µL
0.1 M DTT	1 µL
RNase inhibitor	1 µL
Superscript III RT enzyme	1 µL

Mix was pipetted up and down and incubated at 50°C for 60 min. The enzyme was then inactivated at 70°C for 15 min.

cDNA clone for *smyhc1* was located with the gene name *smyhc1* on the zebrafish genome database 'www.zfin.org' and PCR primers were designed using Primer3 output. Primers chosen were designed to produce the only target 5'UTR of the gene. T3 sequence was added to the start of the reverse primer for the anti-sense probe synthesis and the T7 sequence was added to the start of the forward primer for sense probe synthesis.

Antisense probe

T3 sequence at the start of the REV primer:

5’ GGATCCATTAACCCTCACTAAAGGGAAgcactgcacaaaggctcata

Sense probe

T7 sequence before the FWD primer:

5' TAATACGACTCACTATAGGGAGAtgtcctcacccggttttact

PCR reaction was performed with the reaction mix below:

cDNA	2.5 µL
smyhc1 T7 Forward Primer	1 µL
smyhc1 T3 Reverse Primer	1 µL
5x Phusion Buffer	10 µL
Phusion Polymerase	0.5 µL
dNTP mix (10 µM)	1.25 µL
ddH_2O	33.75 µL
	----------
	50 µL

PCR Program on the thermocycler was set for the following for amplification of *smyhc1* template DNA:

Step 1	95 °C	2 min
Step 2	95 °C	30 sec
Step 3	53 °C	30 sec
Step 4	72 °C	1 min 30 sec (Repeat step 2-4 40x)
Step 5	72 °C	7 min
Step 6	12 °C	∞

The amplification mix was then purified using Qiagen Purification Kit. Then used for probe synthesis.

2.7.3. RNA Probe synthesis

cDNA from zebrafish embryos was used to produce antisense RNA probes for *smyhc1*. The key ingredient for *in situ* hybridisation is the steroid digoxygenin, which binds to the anti-DIG antibody. I used an NTP mix containing Digoxigenin-Uridine Triphosphate (DIG-UTP) which labels all uridine nucleotides with DIG. Anti-Dig antibody is conjugated with enzyme and binds to DIG during *in situ* hybridisation. With the addition of NBT/BCIP which detects the presence of alkaline phosphatase, produces a detectable colour at the site of probe-target RNA binding. This protocol was performed as described by Thisse and Thisse, 2008.

The following were mixed in order at room temperature:

Lineraised template DNA (200 ng)	x µL
dH_2O	(13-x) µL
DIG-UTP NTP mix	2 µL
10x transcription buffer	2 µL
RNase inhibitor	0.5 µL
RNA polymerase	2 µL
0.1 M DTT	1 µL

	20.5 µL

Mix was incubated at 37°C for 2h. Following this incubation, 1 µL of DNAse was added and incubated at 37 °C for 15 min to degrade template cDNA. 1 µL of 0.5M EDTA was added to stop DNAse activity. The probe mix was purified using G-50 columns (Illustra, #27533001) and adjusted to 100 µL. The probe was aliquoted (20 µL) and stored at -80°C.

2.7.4. Embryo fixation – for *in situ* hybridisation

Embryos were selected developmental stages as described by Thisse and Thisse, 2008 and placed in a microfuge tube containing 500 µL 4% PFA. Fixation took place on a gentle rocker at 4°C overnight. If embryos were older than 24 hpf, chorions were removed manually using forceps 30 min before fixation to enable larvae tails to linearise. Embryos before 24 hpf were fixed with chorions intact and dechorionated after fixation. Following fixation, embryos were washed 2x 5 min in PBS and dehydrated with a series of methanol washes: 1x 5 min 50% MeOH 50% PBS, 2x 5 min 100% MeOH. Embryos were then stored at -20°C in 100% MeOH.

2.7.5. Whole-mount in situ hybridisation (WISH)

WISH is a technique used to label the presence of mRNA in the zebrafish embryo. WISH reveals the location and density of mRNA. DIG-oxygenin labelled RNA probes enter the fixed embryos and bind to target DNA by complementary binding. The excess probe was washed out as described below and immunohistochemistry is used to detect the probe using antibody and DIG-oxygenin. This protocol is followed and described by Thisse and Thisse, 2008.

Day 1

Embryos that were stored in MeOH from 2.6.4. where taken out from -20°C and acclimatised to room temperature. Dehydrated embryos were then rehydrated using a series of washes containing 0.1% Tween20 in PBS (PBTween). Rehydration steps were 1x 5 min 50% MeOH, 50% PBTween, followed by 2x 5min in PBTween. Then embryos were digested using proteinase K for a specific time and concentration depending on the age of the embryos according to Table 2.9, this step enables the probe to access deeper into the embryo/larvae for more accurate detection of RNA localisation. Proteinase K digestion was stopped with 2x 5 min washes using glycine (2 mg/mL in PBTween). Samples were then fixed again using 4% PFA for 20 min, gently rocking at room temperature. Then samples were washed with PBTween 2x 5min, room temperature.

Table 2.9. Proteinase K treatment according to the embryonic stage for WISH

Embryonic stage (hpf)	*Concentration of Prot K (µg/ml)*	*Time (minutes)*
24	10	10
26	30	6
28	30	8
30	30	10
33	30	13
36	30	16
37	30	17
40	50	12
45	50	13
47	50	17
50	50	19
56	50	22
60	50	26
74	50	36

To hybridise the embryos, a series of wash steps were made to prehybridise using hybe buffer: wash 5 min 50% hybe 50% PBTween at room temperature, and then prehybridised in hybe (containing yeast RNA and heparin) at 65°C for 1 h to reduce nonspecific binding. Pre-hybe was removed and replaced

with hybe containing 1:200 of probe (made in 2.7.3.) and incubated at 65°C overnight. Probe was then removed and can be reused if stored at -20°C for future use. Samples were washed on a 65 °C heat block and gently agitated between washes: 1x 10min 100% hybe, 1x 10 min 50% hybe 50% 2xSSC, 1x 10 min 2xSCC and then 4x 15 min 0.2x SSC. Samples were then moved to room temperature and were gently rocked on the rocking table between the next washes: 1x 5 min 50% 0.2x SSC 50% MABTween and 1x 5 min MABTween. Samples were blocked with MAB Block (2% Boehringer Blocking ReagentTM (BBR) in MAB) for 1 h at room temperature, on the rocking table, this prevents any non-specific binding of the antibody. MAB block was replaced with an anti-DIG antibody conjugated to alkaline phosphatase enzyme diluted in BBR in a 1:5000 dilution and was on a rocking table at 4 °C overnight. Embryos were then placed onto the rocking table for 1h at room temperature. A series of MABTween washes were made: 4x 15 min MABTween at room temperature on a rocking table. Samples were transferred onto a 24 well plate and MABTween was replaced with BCL Buffer III (0.1 M Tris-HCl, 0.1 M NaCl, 50 mM MgCl2, and 0.1% Tween20) for 10 min. BCL Buffer was then replaced with BCL buffer containing 20 µL/mL NBT + X-phos mixture (Roche, #11681451001) and was incubated at room temperature in the dark. Samples were incubated until colour development (15 min - 2h). The reaction is stopped temporarily by replacing the developing buffer with PBTween + 20 mM EDTA. Permanent stop of reaction by fixation with 4% PFA for 20 min and washed with PBS 2x 5min. Samples were stored at 4°C in PBS containing 0.02% azide.

To image samples, embryos/larvae were immersed in 100% glycerol on petri dishes and observed under a Leica MZ16F fluorescence stereomicroscope attached to iDS camera (#UI-3080CP-C-HQ R2) camera and lighting controlled with an LED ring light attachment.

2.8. Visualising Sarcomere Proteins Using Immunostaining

2.8.1. Embryo Fixation – for immunostaining

Embryos were selected developmental stages as described by Thisse and Thisse, 2008 and placed in a microfuge tube containing 500 µL 4% PFA for embryos less than 3 dpf and 2% PFA used for larvae older than 3 dpf. Fixation took place on the gentle rocker for 1h at room temperature. If embryos were older than 24 hpf, chorions were removed manually using forceps 30 min before fixation to enable larvae tails to linearise. Embryos before 24 hpf were fixed with chorions intact and dechorionated after fixation. Following fixation, embryos were washed 2x 5 min in PBS and washed with 2x 5min PBTx (1x PBS with 0.5% Triton x-100). Embryos were then stored at -4°C in PBS-azide. (0.02% azide). If zebrafish larvae were older than 1 dpf they were either treated with 1-phenyl 2-thiourea (PTU) at 24hpf in fish water before initial fixation or bleached using a bleaching reagent (3.3 mL H_2O_2, 5.95 mL

H_2O, 0.5 mL Formamide, 0.25 mL 20XSSC) to remove pigment. To stop bleaching, embryos washed in 2X 5 min with PDT (0.5 mL DMSO, 2 mL 20% Triton-x100, 47.5 mL H_2O) rocking at room temperature. Embryos were blocked with 5% goat serum diluted in PBTx for 1 h at room temperature followed by incubation with the 1st antibody 2% goat serum in PBTx at 4°C 1 night, if 2 dpf or younger or 2 nights if 2-5 dpf. Embryos were then incubated with the 2nd antibody in 2% goat serum and incubated at 4°C 1 night if 2 dpf or younger or 2 nights if 2-5 dpf. Embryos were then washed with PBTx 4x 15 min. Embryos were then stored at -4°C in PBS-azide (0.02% azide).

2.8.2. Immunostaining on sections

Fish larvae were washed 20% sucrose (in PBS) 2x 5 min on the rocking table, at room temperature. Fish larvae were then incubated with the 20% sucrose overnight at 4°C on the rocking table. The following day, larvae were positioned laterally with dorsal on top in embedding chambers containing OCT medium (Tissue-Tek, #16-004004). Embedded embryos were then snap frozen using liquid nitrogen and stored at -80°C. To section samples, embedded embryos were acclimatised to -22°C in the cryostat for 30 minutes before sectioning. To section samples, embedded embryos are loaded onto a cryostat chuck with dH_2O. Sections of 15 µm were cut and thaw-mounted onto poly-Lysine glass slides. Sections were air-dried on the slides at room temperature overnight. PAP pen was used to draw around samples to keep liquid staining to be enclosed within the sample. Before antibody staining, samples were rehydrated in with the addition of PBS for 5 min. Samples were incubated with primary antibody (Table 2.2) for 3 h, room temperature and then washed with PBS 2x 5 min. Secondary antibody incubation was made (Table 2.3) for 3 h, at room temperature and then further washed with PBS. Samples were mounted with 100 µL of mounting medium (Fluoromount, SouthernBiotech, Cat. 0100-01) with a coverslip on top and set overnight in the dark.

2.9. Zebrafish swimming velocity assay

To assess swimming velocity in response to mechanostimulation, siblings and mutants from 2dpf+ were randomly chosen and stimulated with forceps until a reaction was observed or until no movement would occur for 30s. Embryos at age 2 dpf and 5 dpf were recorded using MZ16 Light microscope (Leica, Watzlar, Germany), and larvae at age 15 dpf, 20 dpf and 30 dpf were recorded using Sony IMX586 OnePlus 7TPro Camera (OnePlus, Shenzhen, China). Larvae swimming velocity was measured using Tracker (https://physlets.org/tracker/), scale was set using photographs of graticule or presence of ruler from 15 dpf+ larvae. Larvae were treated with 100 uM BTS for 1 h at room temperature and the fish stimulation assay was repeated.

Chapter 3

Characterising primary biophysical defects in the presence of *MYH7* mutations

3.1. Introduction

Sarcomeres in striated muscle are made up of four main elements: bipolar myosin thick filaments, polar actin filaments, z-disks (to enable polar actin filaments to assemble into a bipolar structure) and titin (to connect thick filaments to z-disks). How mutations in *MYH7* lead to a mechanistic defect in sarcomere assembly and/or defective muscle contraction remain in question (Squire, 1973). There are two main mechanisms potentially affected by the presence of defective slow MyHC I molecules. Firstly, the ability for myosin to pack together into a myosin filament during sarcomere assembly (Sohn *et al.*, 1997; Cripps, 1999; Thompson *et al.*, 2012). Second, the functional positioning of myosin head during relaxed state or during contraction within the assembled sarcomere (Adhikari *et al.*, 2019; Sarkar *et al.*, 2020).

As described in my introduction chapter, MyHC I plays a key role in sarcomere assembly. The light meromyosin (LMM) structure is key for its intricate coiled coil structure which enables myosin monomers to dimerise and subsequently intertwine into a larger thick filament structure (Squire, 1973; Rahmani *et al.*, 2021). The overall structure of the LMM between vertebrates and invertebrates are very similar with some differences around skip residues (Sodek *et al.*, 1972; Hu *et al.*, 2016). The conserved structure of the LMM between vertebrates and invertebrates emphasise the importance of the amino acid arrangement for myosin molecules packing together (Squire, 1973; Rahmani *et al.*, 2021). Mutations in the LMM may lead to the improper formation and organisation of myosin filaments which may lead to an alteration of their length. Change in myosin filament length has been shown because of its elastic and structural properties (Wilson *et al.*, 2014; Irving, 2017). Myosin filament lengths have been measured during active state and relaxed state; myosin filaments appear 1% longer during active state than in relaxed state (Haselgrove, 1975; Ma *et al.*, 2018).

Such malformed filament backbones formed with defective slow MyHC I molecules may modify myosin head orientation and motor function. In preparation for muscle contraction, the myosin head projects in close proximity to Actin, whereby myosin is in a state what is termed disordered relaxed state (DRX). In the DRX state, the myosin head is ready to bind to Actin, hydrolyse ATP, and subsequently generate force enabling the sarcomere to contract (Stewart *et al.*, 2010; Cooke, 2011; Fusi, Huang and Irving, 2015). When in dormant state, myosin heads fold into what is termed super relaxed state (SRX) whereby myosin heads interact with each other and the thick filament backbone

to position the head to block actin and ATP binding sites (Hooijman, Stewart and Cooke, 2011; Alamo *et al.*, 2016). During the SRX state, myosin heads are unavailable to bind to actin and catalyse ATP to generate force (Huxley and Brown, 1967; Woodhead *et al.*, 2005; Alamo *et al.*, 2008). The ratio between DRX and SRX in different muscle fibre types differ and are determined by the functional demand of muscle type (Hooijman, Stewart and Cooke, 2011; Spudich, 2015; Trivedi *et al.*, 2018). The stabilisation of the SRX state is partially controlled by MyBP-C, involving the two MyBP-C binding sites on *MYH7* (Alamo *et al.*, 2017; Robert-Paganin, Auguin and Houdusse, 2018; Spudich, 2019).

Many hypertrophic cardiomyopathy (HCM) mutations in either the MyBP-C domain in *MYH7* or in MyBP-C itself have been shown to destabilise SRX state with increased proportion myosin heads in DRX state (Adhikari *et al.*, 2019; Sarkar *et al.*, 2020). A link between HCM mutations affecting the converter and C-terminal MyBP-C binding site have led to destabilise myosin in SRX state and thus leading to a predominance of myosin heads in the DRX state and subsequently leading to hypercontraction of the cardiac muscle (Alamo *et al.*, 2017; Toepfer *et al.*, 2020). MyBP-C connects to myosin at two sites, the N-terminal MyBP-C domain connects to myosin head region and C-terminal MyBP-C connects to myosin LMM (Luther *et al.*, 2008; Spudich, 2015). MyBP-C knockout studies have shown increased shift of myosin heads in SRX to DRX state and thus leading to hypercontractility and slowed relaxation (Stelzer, Fitzsimons and Moss, 2006; Moss, Fitzsimons and Ralphe, 2015; McNamara *et al.*, 2016; Christopher N. Toepfer *et al.*, 2019) but arrangement of myosin heads in thick filament are not severely disturbed (Luther *et al.*, 2008; Zoghbi *et al.*, 2008; McNamara *et al.*, 2016). Mutations in the LMM affecting slow skeletal muscle, particularly in the second MyBP-C binding domain which can also be bound by myomesin, may show similar shift of myosin heads from SRX to DRX state seen in patients with mutations in MyBP-C binding domain in the head region.

Mutations affecting the MyBP-C binding domain may also affect the giant molecular spring within the sarcomere, known as Titin. Titin has been shown to play a role in passive tension after the active tension from myosin and actin filaments in the active cross-bridge cycle (Cazorla *et al.*, 2001; Fukuda *et al.*, 2005). Titin is known to have extensible spring-like features to provide the passive tension after the sarcomere have been overstretched (Labeit and Kolmerer, 1995; Freiburg *et al.*, 2000). Passive tension from Titin have been shown to change thick filament length whereby M-lines within the A-band appear further apart (Irving *et al.*, 2011). Titin connects with myosin indirectly through MyBP-C at the crossbridge region (Tonino *et al.*, 2019). Myosin in the conventional “J” motif resemble myosin in a relaxed state whereby myosin heads interact with each other to form a “interacting heads motif” (IHM) and both myosin heads interact with S2 region (Alamo *et al.*, 2017; Woodhead and Craig, 2020).

Mutations in MyBP-C the myosin binding site may show weakened passive tension through poor interaction between myosin and titin through MyBP-C and thus, may present as muscle in exercises involving stretching the muscle.

The myomesin binding site is the alternative major domain in the myosin LMM region aside from the MyBP-C binding site mentioned earlier. There are 3 myomesin isoforms in humans: myomesin-1 is expressed in all skeletal and cardiac muscles, myomesin-2 is expressed in adult heart and fast skeletal muscle (Agarkova *et al.*, 2004), and myomesin-3 is expressed in slow skeletal muscle (Schoenauer *et al.*, 2008). A case of mutation in myomesin have been associated with HCM (Siegert *et al.*, 2011). Patients with mutations in EH-myomesin, a splice variant of myomesin-1 show DCM (Schoenauer *et al.*, 2011; Bollen *et al.*, 2017). Lack of myomesin-1 in human cell lines show sarcomere disassembly and regulation of muscle contraction (Hang *et al.*, 2021). Myomesin-3 knockout studies in zebrafish show no effect on sarcomere organisation suggesting the role of myomesin-1 show predominant involvement in sarcomere organisation than myomesin-3 (Xu *et al.*, 2012). However, the mechanism for sarcomere disassembly from lack of myomesin-1 remain unclear.

In this chapter, I study the primary defects in the presence of *MYH7* mutations muscle fibres by analysing muscle fibres extracted from the *vastus lateralis* of healthy controls and patients with mutations in *MYH7* (Table. 2.1). To identify whether mutations in *MYH7* lead to alteration in myosin packing and subsequent myosin filament length, a comparison between myosin filament length from muscle fibres from healthy controls and patients with *MYH7* mutations were made. Here we test for changes in myosin filament length using fluorescence microscopy, staining for slow myosin using antibody A4.951 (Webster *et al.*, 1988; Cho, Webster and Blau, 1993; Blagden *et al.*, 1997), followed by measuring the length of A-band (Methods 2.2.1). A similar technique has been used to measure change in actin filament length whereby aged mice show decreased actin filament length (Gokhin *et al.*, 2014). I first show through immunofluorescence that there is no observable change to thick filament length in the presence of *MYH7* mutations. To determine whether there were no changes in filament length, or the method of detecting changes was not sensitive enough to identify more subtle changes between active and relaxed states, I looked at the proportions of myosin in SRX and DRX states in the muscle fibres. Here, I could observe an increased proportion of myosin molecules in DRX state in muscle fibres from patients with LMM mutations. Mutations at the Myomesin and MyBP-C site show most percentage difference in proportion of myosin in DRX state. The extent of these alterations may vary from one mutation to another inducing muscle phenotype variability. However, degree of variability between LDM and MSM patients were not distinguishable through measuring

proportions of myosin in DRX and SRX states. It is concluded that mutations in LMM at the myomesin and MyBP-C site show increased DRX myosin head positioning by destabilising the SRX state.

3.2. Results

3.2.1. Mutation in *MYH7* show no change in myosin filament lengths

To assess whether *MYH7* mutations alter the length of myosin filaments, fibres were extracted from the *vastus lateralis* of healthy controls and patients with *MYH7* mutations (Table 2.1, Fig 2.1) and subsequently stained with two different antibodies. Firstly, I stained with MF20 to visualise all skeletal myosin filaments to identify whether there were overall changes in myosin filament length (Shimizu *et al.*, 1985). ImageJ plugin DDecon was used to deconvolute fluorescence microscopy images and subsequently measured for myosin filament lengths through their imaged fluorescence intensity peaks (Fig 3.1A). The variability of the measurements was 10.45% between controls (Fig 3.1). In control individuals, the overall mean myosin filament lengths were 1.75 µm (SD = 0.05, Fig 3.1A). Patients with mutations in *MYH7* show no difference in myosin filament length compared to healthy controls (One Way ANOVA $p>0.05$) suggesting that mutations in *MYH7* do not alter myofilament length. However, our findings may also suggest that mutations in *MYH7* are very subtle, and our analysis may only affect slow fibres exclusively.

To address whether slow specific myosin filament lengths change in the presence of *MYH7* mutations, muscle fibres were stained with A4.951, a slow type I myosin specific antibody (Fig 3.1B) (Webster *et al.*, 1988; Cho, Webster and Blau, 1993; Blagden *et al.*, 1997). Variability of the measurements between controls was 13.3%. In control individuals, the overall mean myosin filament length in type I fibres was 1.72 µm (SD = 0.28). There were observed no significant difference between controls and patient samples (Fig 3.1B) suggesting that mutations in *MYH7* do not alter myofilament length in slow fibres. Despite no observable change in thick filament length in slow fibres, this does not rule out the possibility the current analysis is not sensitive enough to detect a 1% change generated by a change between myosin in an active or relaxed state (Haselgrove, 1975; Ma *et al.*, 2018).

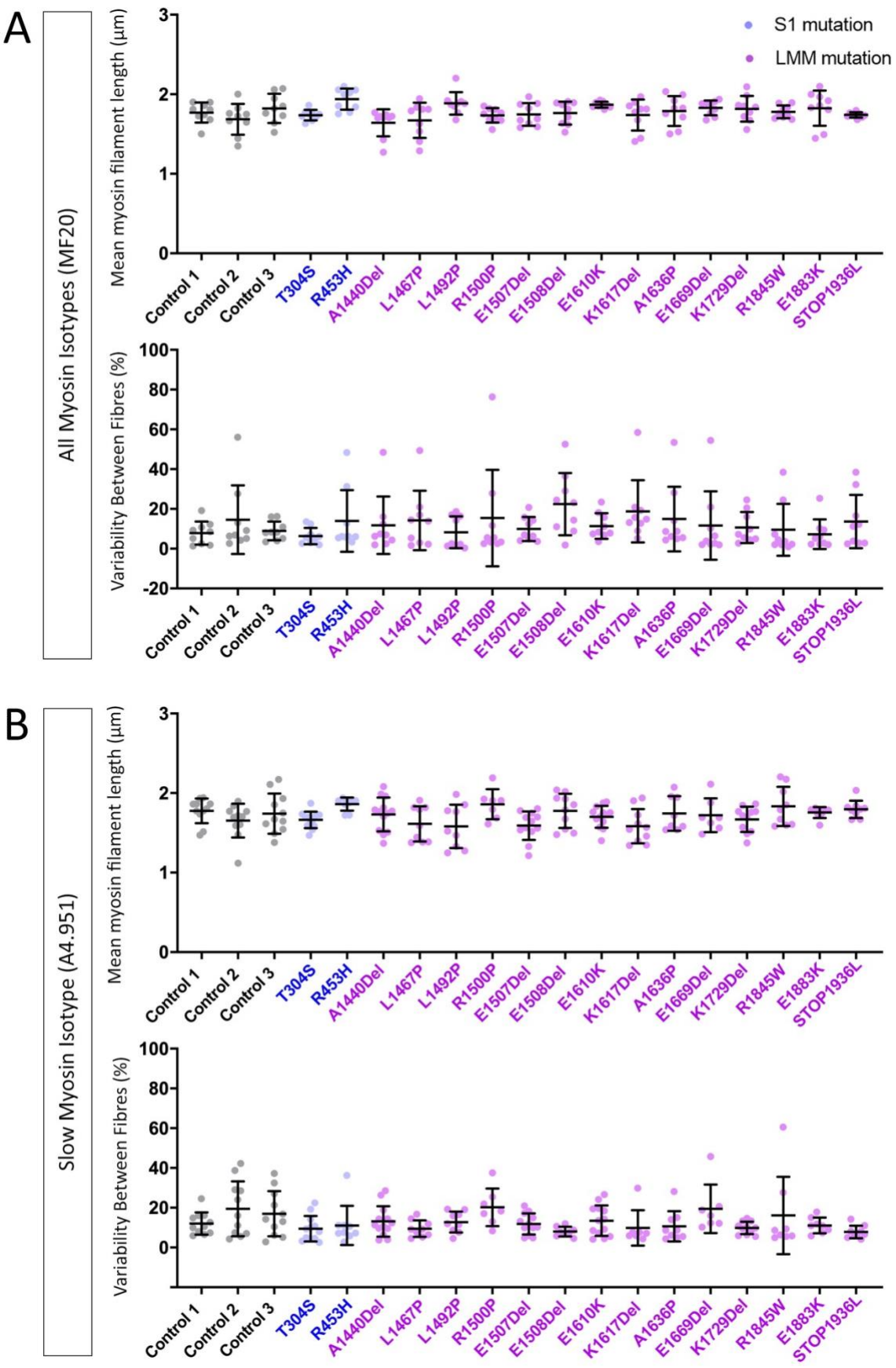


Figure 3.1. Myosin filament measurements (slow fibre types) of controls and patients with MYH7 mutations.

A) Thick filament length obtained by immunostaining with MF20 targeting against slow myosin. Compared to measurements from 3 healthy controls, data shows no change in filament length in patients. Variability between filament measurements across each fibre was analysed (standard deviation/mean) show no significant variation within each fibre measurement per sample. **B)** Thick filament length obtained by immunostaining with A4.951 targeting against slow myosin. No observable difference between controls and patients. Variability between filament measurements across each fibre was analysed below. Statistical analysis using one-way ANOVA between the mean measurements of each fibre. The colours of each plot indicate the location of mutation – No mutation (grey), S1 (blue) and LMM (purple).

3.2.2. Mutation in *MYH7* shifts myosin molecules in DRX state in patient fibres

Since there was no observable difference in myosin filament length from the previous fluorescence study, to detect subtle changes in myosin filaments, an investigation for more subtle structural changes in myosin head positioning were made. During the relaxed state, in the absence of Ca^{2+}, myosin molecules are present in two main states, the SRX and DRX (Fig 1.10). ATP turnover rate from myosin in a DRX state is 5 times faster than in SRX (McNamara *et al.*, 2015). To measure ATP turnover rate, fibres were incubated with fluorescently labelled ATP (Mant-ATP) and when flushed with non-fluorescently labelled ATP, all fibres initially show a rapid decrease in the fluorescence followed by a slower decay in fluorescence intensity (Fig 3.2A). Proportion of the two states in each fibre were calculated by fitting ATP turnover rate into a two-state exponential curve. Proportions of P1 showing rapid decay phase represent the DRX state and P2 showing slow decay phase represent the SRX state (Fig 3.2A).

Initial comparison of single traces from two different patients, one mutation in the S1 region and one mutation from the LMM (Fig 3.2B). Patient fibre with S1 mutation T304S show similar decay in fluorescence intensity to healthy control fibre, while fibres from LMM mutation K1617del showed faster decay compared with healthy control. As we tested more fibres from each patient compared with controls, we plotted the calculated percentage of DRX myosin molecules in each fibre from individual patients (Fig 3.2C). Since mutations primarily affect slow skeletal muscle, fibres were stained with A4.951 after measuring fluorescence decay to identify which fibres were slow fibres and measurements were isolated to create graph (Fig 3.2C). Remaining measurements from fast fibres show no difference between controls and patients (Appendix 3.1.2). Proportion of fast and slow fibres in each sample were counted and patients with LDM show higher proportion of slow fibres than controls, consistent with clinical data from muscle biopsies (Appendix 3.1.3). However, patients with HCM and MSM also show higher proportion of slow fibres compared to controls thus, the predominance of slow fibres alone may not be an accurate diagnostic for LDM from muscle biopsies from the *vastus lateralis* (Appendix 3.1.3). Cross-sectional biopsies to determine type I fibre predominance reveal more consistent results in defining LDM phenotype (Fig 1.5). Mean percentage

DRX myosin in healthy controls were 39.63% where variability of the measurements was 13.42% between controls. Patients with mutations in S1 region show similar proportion of DRX as seen in healthy controls. This suggests that mutations in my sample set directly affecting head region does not alter myosin head positioning in the relaxed state. However, fibres from patients with mutations in the LMM region show an increased proportion of DRX myosin compared to healthy controls, with an exception from mutation A1883E (Fig 3.2C). The mean proportion of DRX in the presence of LMM mutations were significantly higher at 55.45% compared to the 39.63% in healthy controls.

Mutations showing significantly higher DRX levels were plotted on the myosin protein map to identify possible affected binding sites (Fig 3.3). Position of these mutations clustered along the myomesin and MyBP-C binding site suggesting mutations in myomesin, and MyBP-C destabilize myosin in SRX state in slow skeletal muscle fibres. When comparing degree of variability of DRX and SRX proportions between LDM and MSM (Fig 3.2D) no distinguishment could be made from this data set, suggesting the mechanism of pathology between the two diseases remain unknown. Overall, our data indicate that mutations in the LMM influence remodelling of myosin filament length that cannot be detected through fluorescence microscopy in 3.2.1. and the head positioning during the relaxed states are possibly affected by the presence of mutations on and near the myomesin and MyBP-C binding site that destabilise myosin head positioning in the SRX state.

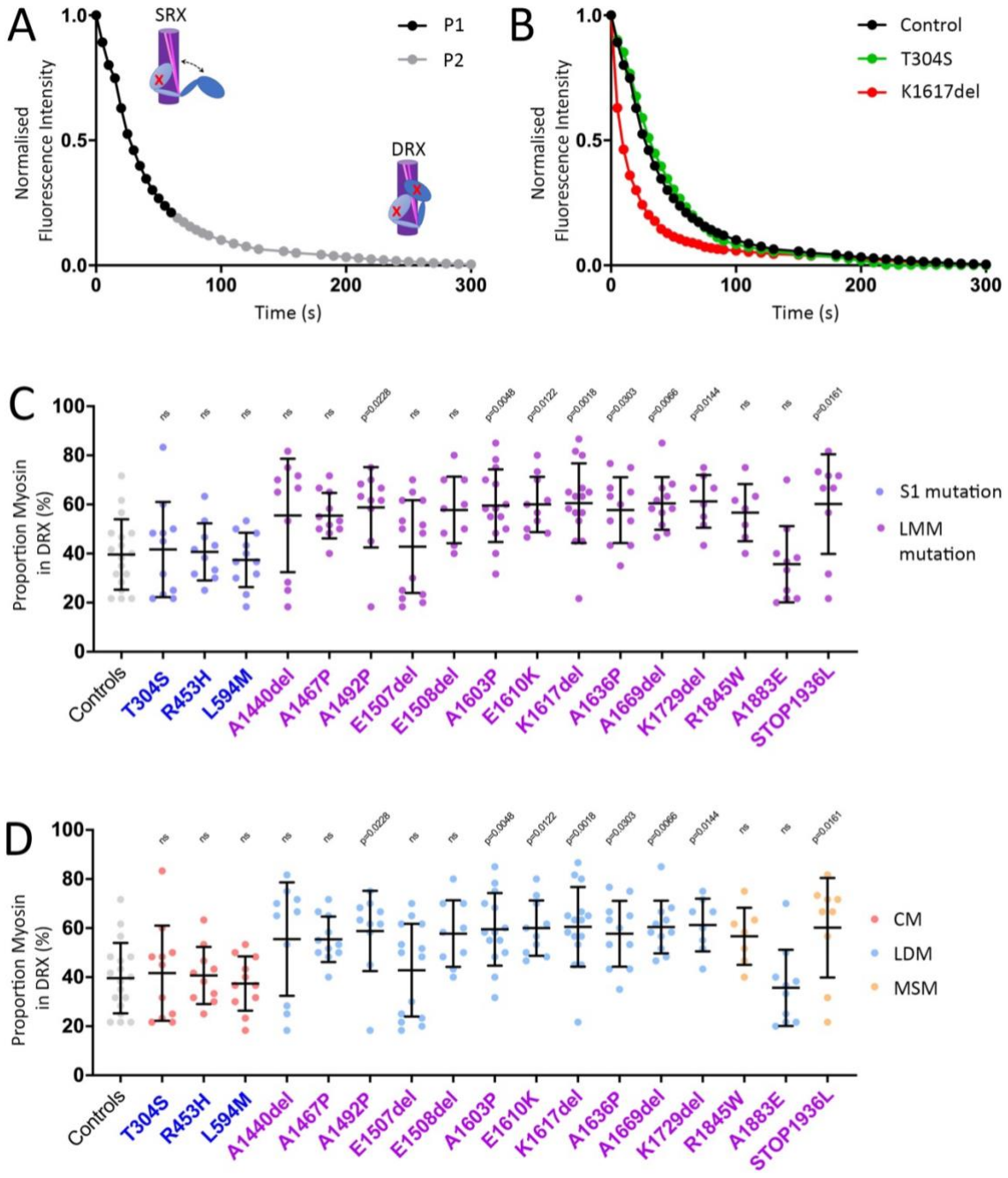


Figure 3.2. The proportion of DRX increases in patients with LMM mutations.
A) Single trace of fluorescence decay from one muscle fibre. Fluorescence decay is plotted on a two-state exponential decay curve. Rapid decay of fluorescence P1 represents the DRX state where ATP turnover is fast. The slow decay of fluorescence P2 represents the SRX state where ATP turnover is 5 times slower than DRX. Using a two-state exponential decay equation in Methods 2.2.2 percentage proportions of each state can be calculated. **B)** representative single comparison of fluorescence decay of three conditions from Controls (black), p.T304S (green) and p.K1617del (red). All fibres were incubated with 125 uM mATP and chased with 4mM ATP. The experiment was recorded from t=0s as mATP was flushed into the flow chamber and images were taken every 5s until t=180s and every 10s until t=300s. **C)** Slow fibres were selected for fibre type by immunostaining with A4.951 against slow myosin, data from positive staining fibres were plotted. Mutations in the LMM region increase DRX proportion in slow muscle fibres. The colours of each plot indicate the location of mutation – No mutation (grey), S1 (blue) and LMM (purple). **D)** Data presented with a mutation in *MYH7* according to the pathology of the disease. The proportion of DRX between DM and MSM patients is indistinguishable. Statistical analysis using one-way ANOVA between the mean measurements of each fibre. Statistics using PRISM GraphPad – One way ANOVA (p=0.05).

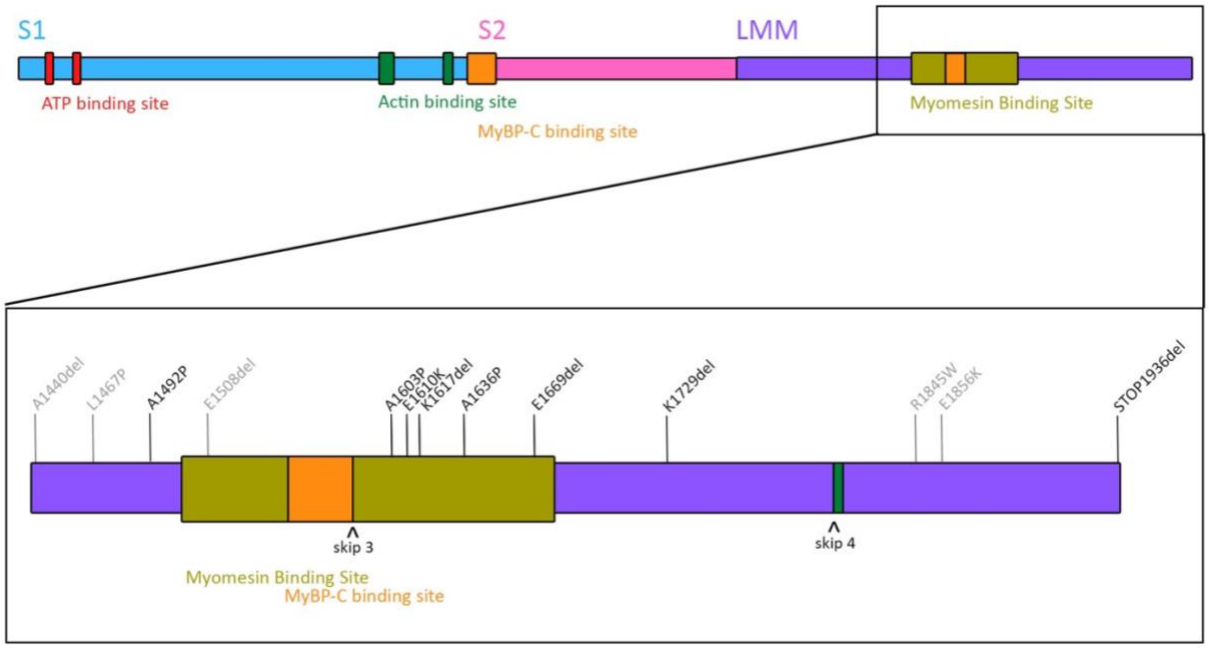


Figure 3.3. Fibres with mutations leading to DRX are primarily present in the myomesin binding site.
Map of MYH7 protein with binding domains labelled, close-up of C-terminal LMM region with labelled mutations from sampled patients. Patient mutations with significantly higher DRX proportions in Fig 3.3 are labelled in black and mutations with insignificant changes in DRX proportion are labelled in grey.

3.3. Discussion

In this chapter, I identify primary biophysical alterations in muscle fibres from patients with skeletal muscle disease causing mutations in *MYH7*. There are a few main findings. Firstly, myosin filament length is not observably altered in muscle fibres in the presence of *MYH7* mutations. Secondly, although our analysis using immunofluorescence was not sensitive enough to detect changes in myosin filament length, it is also not a suitable to detect more subtle functional changes in myosin elasticity when in active vs relaxed. Functional changes were assessed by identifying the proportion of myosin heads positioned in either the SRX or DRX state. Fibres isolated from patients with mutations near and on myomesin and MyBP-C site show increased proportion of myosin heads in DRX state than in healthy controls.

3.3.1. Sarcomere assembly remain intact in the presence of defective slow myosin molecules

The LMM structure have been described as an intricate coiled-coil structure which enables myosin monomers to dimerise and subsequently intertwine into a larger thick filament structure (Squire, 1973; Rahmani *et al.*, 2021). The conserved structure of the LMM between vertebrates and invertebrates emphasise the importance of the amino acid arrangement for myosin molecules packing together (Squire, 1973; Rahmani *et al.*, 2021). The LMM is essential for myosin filament formation whilst S1 and S2 region are dispensable (Sohn *et al.*, 1997; Cripps, 1999; Thompson *et al.*, 2012). As of

current, there have been no studies showing changes in myosin filament length. Despite such conserved intricate structure of the coiled-coil LMM, our results show a full formation of thick filaments into striations at regular intervals. Mutations in patients are dominant (heterozygous), typically missense or single amino acid deletions. Since patients with dominant mutations express both healthy and defective myosin, there is a degree of variability in the ratio of healthy to defective myosin molecules intermixed in the formation of thick filaments. Ratio of defective to healthy myosin molecules may be very small and thus, if there is defective organisation within the thick filament, they may be too subtle to detect using immunofluorescence to measure changes in thick filament measurement length.

3.3.2. Defective slow myosin MyBP-C binding domains destabilise myosin in SRX state

A link between hypertrophic cardiomyopathy (HCM) mutations affecting the converter and C-terminal MyBP-C binding site and MyBP-C itself have led to destabilise myosin in SRX state and thus leading to a predominance of myosin heads in the DRX state and subsequently leading to hypercontraction of the cardiac muscle (McNamara *et al.*, 2016; Alamo *et al.*, 2017; Christopher N. Toepfer *et al.*, 2019). Since MyBP-C connects to myosin at two sites and studies have shown that mutations affecting the N-terminal MyBP-C domain destabilises myosin in the SRX state (Luther *et al.*, 2008; Spudich, 2015). My main focus was to identify whether mutations in the C-terminal MyBP-C binding domain also show the same effect. Our findings from patients with mutations affecting the *MYH7* C-terminal MyBP-C binding domain show a shift in proportion of myosin heads from SRX state to predominantly in the DRX state. Shift of myosin head positioning towards the DRX state suggest the C-terminal MyBP-C binding domain show the same destabilising effect of the SRX state as the mutations found in the N-terminal MyBP-C domain in the slow myosin molecule and MyBP-C itself. Current data suggest the role of both MyBP-C sites and MyBP-C itself is to stabilise the SRX state through the interaction with slow myosin at both binding sites. Mutations affecting the interaction between slow MyHC, and MyBP-C have led to hypercontractile muscle in HCM patients and possibly hypercontractile and poor relaxing skeletal muscle. Since MyBP-C binding sites are affected, the role of titin in muscle contraction and relaxation may be affected. Titin connects with myosin indirectly through MyBP-C at the crossbridge region (Tonino *et al.*, 2019) and provides passive tension after the sarcomere has been overstretched (Labeit and Kolmerer, 1995; Freiburg *et al.*, 2000). Thus, mutations affecting the MyBP-C binding site reduces the ability for sarcomeres to return to relaxed state before muscle contraction and may lead to hypercontractile skeletal muscle.

3.3.3. Mutations in myomesin binding site dispensable for sarcomere organisation

Myomesin binding site overlap with the C-terminal MyBP-C binding site and brings the question whether myomesin interaction with slow myosin were affected in the presence of mutations at this site. The role of M-band protein myomesin have been described to regulate and stabilise the packing of myosin filaments into a hexagonal myosin filament lattice (Agarkova *et al.*, 2003; Hu, Ackermann and Kontrogianni-Konstantopoulos, 2015). The predominantly expressed myomesin gene is myomesin-1 and is expressed in all skeletal and cardiac muscles (Schoenauer *et al.*, 2008). Knockout of myomesin-1 in human cell lines show sarcomere disassembly and regulation of muscle contraction (Hang *et al.*, 2021). However, our data do not show sarcomere disassembly but rather myosin filaments organised into striations at regular intervals. Additionally, we have shown that mutations in the Myomesin/MyBP-C binding site overlap which may have destabilised myosin heads from SRX state into DRX state. Thus, mutations affecting the myomesin binding site is dispensable for sarcomere assembly in slow muscles. Mutations at the Myomesin/MyBP-C binding overlap site are more likely to affect the involvement of MyBP-C than Myomesin in regulating contractility and relaxation of slow muscles.

3.3.4. Conclusion

In conclusion, I demonstrate that there is no overall alteration in sarcomere organisation in the presence of defective myosin molecules. I provide evidence that in the presence of mutations affecting the *MYH7* MyBP-C binding site shift myosin heads from SRX state into predominantly DRX state. Shift of myosin head positioning may be due to destabilised SRX state. Although Myomesin and MyBP-C binding sites overlap in the LMM region, the likelihood of defects involving Myomesin appear unlikely. Despite current findings describing destabilising effects on slow myosin SRX state, there is no clear data to distinguish mechanistic defects between LDM and MSM patients. Since our current studies have only assessed the primary biophysical defects, how such distinct phenotypes are developed remain in question. To distinguish the mechanism of pathology leading up to the clinical phenotypes observed in LDM and MSM, studying the role of slow myosin during early developmental stages will aid in identifying the mechanistic defects associated with defective myosin molecules in the development of either LDM or MSM.

Chapter 4

Identify zebrafish equivalent gene to human *MYH7*

4.1. Introduction

In the previous chapter, I demonstrated primary biophysical defects in muscle fibres from patients with *MYH7* mutations. If and how these mutations affect the early stages of development remains unknown as all samples analysed so far were from adults. To study the developmental defects of slow myosin mutations affecting slow skeletal muscle, zebrafish disease models might prove advantageous. An essential first step in such a process would be to identify the zebrafish equivalent of the human *MYH7* gene that would be most likely to give a phenotype in a defined functional muscle. To identify a fish equivalent gene, I firstly look at the evolution of sarcomeric MyHC genes and whether zebrafish and humans have a common ancestor for slow MyHC. I next look at the expression of zebrafish MyHC genes and identify whether these genes are expressed in slow skeletal, and the heart ventricle as seen in humans.

In humans, there are a total of eleven sarcomeric MyHC genes. The oldest of these genes is *MYH16* which was ancestrally expressed for jaw muscles (Fig 4.1). A later duplication event led to the formation of *MYH15* and *MYH14 (MYH7B),* which were the ancestral skeletal and cardiac MyHC genes (Rossi *et al.*, 2010). The next duplication event formed two clusters, the *MYH6/7* cluster which is present in tandem on chromosome 14 (Yamauchi-Takihara *et al.*, 1989; Gulick *et al.*, 1991) and a fast skeletal MyHC cluster which is present in tandem on the human chromosome 17. The *MYH6* and *MYH7* cluster is known to have formed from a gene duplication event in mammals (Yamauchi-Takihara *et al.*, 1989; Gulick *et al.*, 1991). MYH6 and MYH7 have evolved to be different in their protein sequence and function, where *MYH6* is expressed in the atrium of cardiac muscle and *MYH7* is expressed in both ventricular cardiac muscle and slow type I fibres in slow skeletal muscle (Fig 4.1). This gene duplication event is can be seen in mammals amphibia, and lobe-finned fish but not easily identified in zebrafish or in avian (Desjardins *et al.*, 2002). Chicken has three MyHC genes, *MYH15* (formerly named *MYH6/VMHC/SM2*), *MYH7B* (formerly named *ssMYHC/SM1*) and *MYH7* (formally named *AMHC*) (Chen *et al.*, 1997) suggesting the use of ancestral MyHC genes for slow skeletal muscle. Tropical claw frogs have *myh6* and *gpc6* (*myosin-7*) genes present in tandem Ensembl Primary assembly 1:127,598,294-127,629,343 (Appendix 4.3). The coelacanth has *MYH6* and *MYH7* present in tandem on Scaffold JH126769.1: 686,924-731,742 (Appendix 4.3). In teleost fish, including zebrafish, there are a higher number of MyHC genes than there are in tetrapods (Watabe and Ikeda, 2006; Ikeda *et al.*, 2007), as teleost fish have undergone an additional round of genome duplication (Amores *et al.*, 1998;

Meyer and Schartl, 1999; Taylor *et al.*, 2001). Zebrafish have 5 slow MyHC genes (Stone Elworthy *et al.*, 2008): *smyhc1, smyhc2, smyhc3, smyhc4* and *smyhc5* (Stone Elworthy *et al.*, 2008), there are 3 zebrafish cardiac MyHC genes including *myh7, myh7l* and *myh6* (Zhang and Xu, 2009) and 6 fast MyHC genes: *myhcz2, myhc4, myha, myhz1.1, myhz1.2* and *myhz1.3* (Nord *et al.*, 2014). When looking at the evolution of slow MyHC in zebrafish in comparison to humans, it was unclear whether the divergence of *MYH6* and *MYH7* occurred before the separation of lobe-finned and ray-finned lineage or whether there was a separate divergence of the slow and cardiac cluster formed from an ancestral slow MyHC (Fig 4.2). To address this in my results, I look at the protein sequence of lobe-finned *MYH6* and *MYH7* to determine the characteristic amino acids to distinguish between the two proteins and identify whether these key amino acids can categorise zebrafish MyHC genes into an *MYH6* or *MYH7* group. This can also identify whether the ancestral MyHC that lead to the divergence of *MYH6* and *MYH7* is the same ancestral MyHC that diverged in teleost fish. Synteny analysis of these genes will describe whether *smyhc1-5, myh7, myh7l* and *myh6* were evolved from *MYH7* and describe which zebrafish genes arose from a teleost genome duplication event.

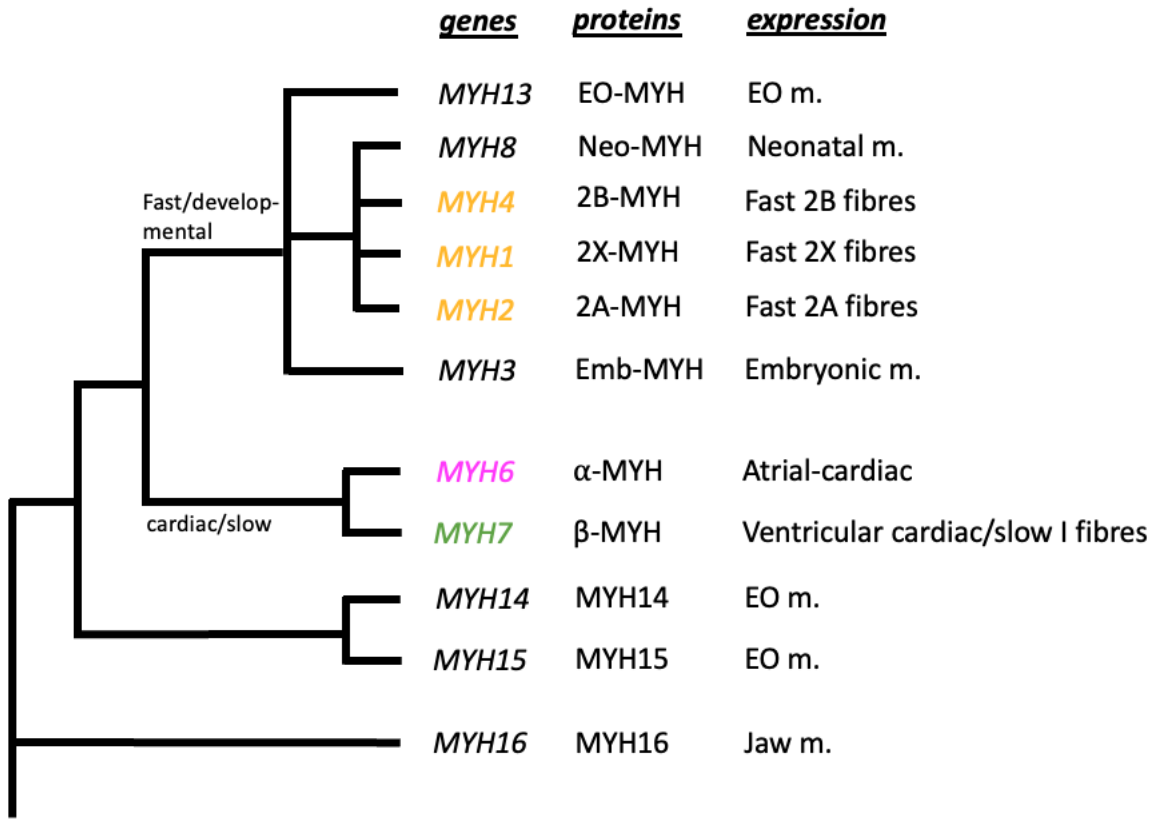


Figure 4.1. Schematic evolution gene tree describing mammalian MYH genes.
The phylogenetic tree on the left with gene name, protein name and location of expression in mammals. Branches are not to scale. Figure adapted from Rossi *et al.*, 2010. Abbreviations: EO-extraocular, Neo-neonatal, Emb-embryonic.

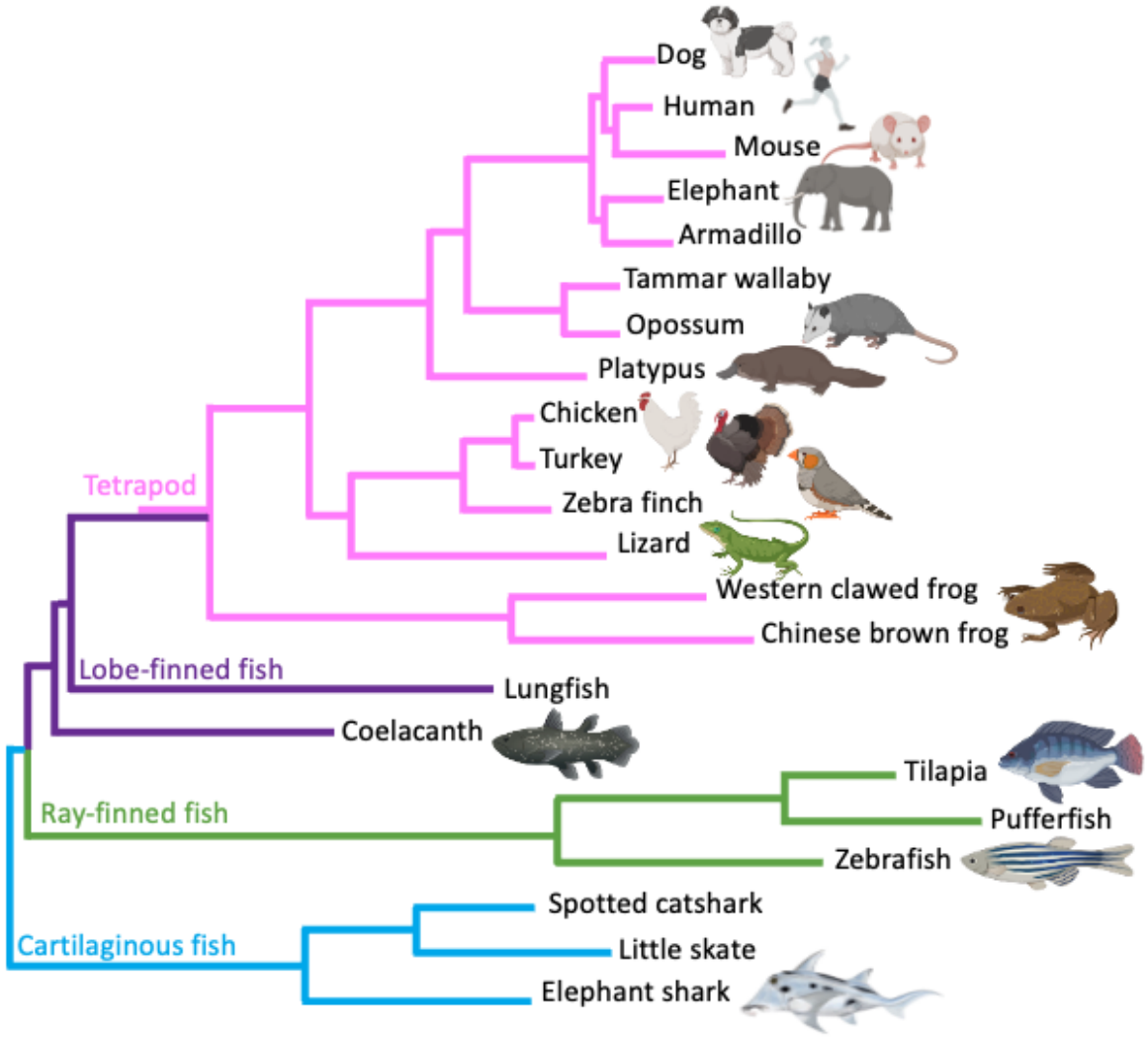


Figure 4.2. Phylogenetic tree of a range of tetrapod, lobe-finned fish, ray-finned fish, and cartilaginous fish. Phylogenetic tree using multiple sequence alignments of 251 genes comparing orthologs between a range of tetrapod, lobe-finned fish, ray-finned fish, and cartilaginous fish to describe the genetic relationship between human to zebrafish. Tree rooted with cartilaginous fish. Branches are not to scale. Figure adapted from Amemiya *et al.*, 2013.

Human *MYH7* is expressed in both the heart ventricle and slow skeletal muscle. In zebrafish, the expression of *smyhc1, smyhc2* and *smyhc3* genes are exclusively in slow muscle fibres (Stone Elworthy *et al.*, 2008) (Fig 4.3). In the early stages of development, *smyhc1* is predominantly expressed in slow fibres and in a small subset of slow muscles, *smyhc2* and *smyhc3* are expressed. *Smyhc2* shows localisation in the craniofacial muscles and a small subset of slow muscle, named supracarinalis anterior (sca), inferior obliquus (iob) and embryonic lateralis superficialis (els) and infracarinalis posterior (icp) (Fig 4.3). *Smyhc3* also shows weak localisation in craniofacial muscles and a subset of slow muscles named sca and els. At later stages, after 17 dpf to adulthood, secondary slow fibres, present at the horizontal myoseptum, *smyhc1* expression is replaced by the expression of *smyhc2* and *smyhc3* (Stone Elworthy *et al.*, 2008). Expression data for *smyhc4* and *smyhc5* were unknown as no *in*

situ hybridisation experiments were made for these genes. *Myh7* is expressed in the heart ventricle and not in the slow skeletal muscle (Fig 4.3)(Park *et al.*, 2009). *Myh7l* shows localisation in the heart ventricle and a weak signal in the tail (Fig 4.3)(Thisse and Thisse, 2004). However, no known studies for more specific probes to *myh7l* are published. The functional role of *smyhc1-3, myh7* and *myh7l* show similarity to human *MYH7*, where smyhc genes are expressed in slow skeletal muscle and myh7 genes are expressed in ventricular cardiac muscle. However, the functional roles of these zebrafish genes have been split across many genes in comparison to the single *MYH7* in humans. To identify how these genes arose and whether zebrafish slow MyHC genes are linked to human slow MyHC, I look at the gene synteny to first, identify whether these genes are linked to human *MYH7* or whether these genes derived from a common slow ancestral MyHC gene to *MYH6* and *MYH7*. Secondly, I will look at gene synteny between *smyhc1-5* and *myh7* and *myh7l* to determine whether these genes arose after a genome duplication event.

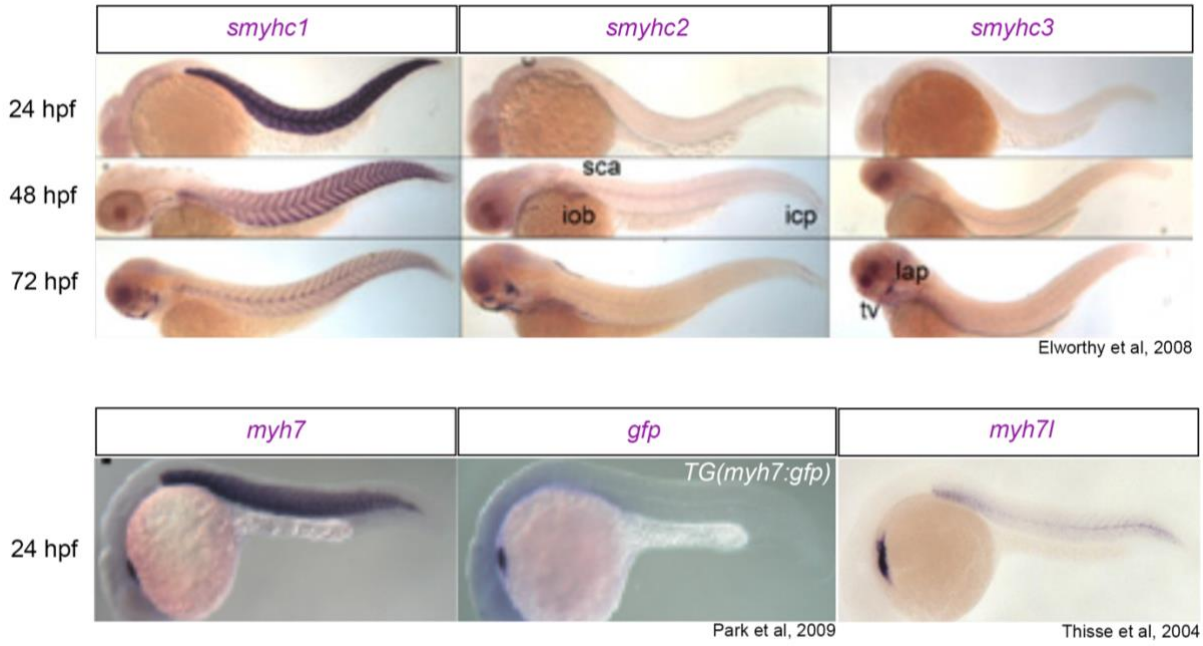


Figure 4.3. RNA localisation of *smyhc1-3* and *myh7/myh7l.*
ZFIN search of the whole-mount in situ hybridisation (WISH) for 5'UTR *smyhc1-3* sequences from 12-72hpf embryos (Elworthy et al., 2008). *smyhc1* show expression predominantly in slow skeletal muscle, *smyhc2* shows expression subset of slow muscle cells, in sca, iob and icp, and *smyhc3* also show expression in a subset of slow muscle, in the sca and els. *myh7* 24hpf. WISH targeting *myh7* mRNA at 24 hpf and targeting *gfp* mRNA in transgenic line *Tg(myh7:gfp)* (Park *et al.*, 2009). *Myh7* shows expression in the heart ventricle and appears in slow skeletal muscle, probe may cross hybridise with *smyhc1* which reveals expression in slow muscle. To prevent cross hybridisation, indirect detection of *myh7* expression using *Tg(myh7:gfp)* zebrafish and in situ against *gfp* reveal expression only present in the ventricle and no localisation in slow skeletal muscle. *Myh7l* shows expression in the heart ventricle and a slight appearance in slow skeletal muscle, which may also be due to cross hybridisation with *smyhc1*.WISH targeting *myh7l* mRNA at 24 hpf (Thisse et al., 2004). Abbreviations: supracarinalis anterior (sca), inferior obliquus (iob), embryonic lateralis superficialis (els) and infracarinalis posterior (icp). Figure permission granted from Stone Elworthy *et al.*, 2008b; Park *et al.*, 2009.

In this chapter, I first compare human *MYH7* with genes in the zebrafish genome using BLAST analysis. This gave me an initial list of candidate zebrafish genes *smyhc1-5, myh7, myh7l, myh6 and myh4* and *myhz2.* I then distinguished which of these candidate genes were true slow MyHCs and differentiate them from fast MyHC using their amino acid sequence. I identified synteny between zebrafish *smyhc1-5, myh7* and *myh7l* with human *MYH7* and zebrafish *myh6* show synteny to human *MYH6. Smyhc1-5* are syntenic to *myh7/myh7l* suggesting these genes arose from a teleost genome duplication event. Mapping human LDM and MSM mutations onto zebrafish *smyhc1-5, myh7* and *myh7l* protein sequences show mutations occurring at highly conserved amino acids. It is concluded that amongst data showing *smyhc1-5, myh7* and *myh7l* evolutionarily linked to human *MYH7*, the function of *smyhc1* showing broadest expression in slow skeletal muscle, *smyhc1* was chosen as the zebrafish equivalent gene for human *MYH7*.

4.2. Results

4.2.1. zebrafish smyhc1-5, myh7, myh7l and myh6 show similarities to human MYH6/7

The first step is to identify a list of zebrafish slow MyHC genes and distinguish these genes from fast MyHC genes. To achieve this, I performed a basic local alignment search tool (BLAST) analysis using human *MYH7* nucleotide and protein sequence against the zebrafish genome. Human MYH7 protein sequence was used for BLAST analysis and the top candidate proteins were firstly chosen based on at least 95% query and then further analysed for sequence identity. The query cover shows the percentage of amino acids in MYH7 aligned to sequences in the zebrafish database. Query covers that are less than 100% are due to shorter lengths of amino acid sequences in zebrafish genes compared to the length of human MYH7 sequences. A range of slow, fast and developmental myosin proteins was identified as possible candidates for the zebrafish equivalent to human MYH7 (Table 4.1). Identity scores for protein sequences were ranked for each candidate (Table 4.1) and smyhc1-5, myh7, myh7l and myh6 show the highest identity scores of 82-86% for protein sequences suggesting close amino acid sequence resemblance to human MYH7. CLUSTALO amino acid sequence alignment was used to compare human MYH proteins to all zebrafish candidate genes from BLAST analysis (Table 4.1, Appendix 4.2). Zebrafish smyhc1-5, myh7, myh7l and myh6 proteins cluster together with human MYH6 and MYH7 proteins suggesting the highest amino acid similarity to MYH6/7 (Fig 4.4). Myha, myhb, myhz1.1, myhz1.2, myhz1.3, myhz2 and myhc4 show amino acid identity scores of 76-78% however, these proteins were eliminated from candidate proteins as they show greater similarity to non-slow human proteins MYH13, MYH3, MYH8, MYH4, MYH1 and MYH2 (Table 4.1, Appendix 4.2). Proteins showing the lowest % identity scores were myh9a, myh9b, myh10, myh11a, myh11b and myh14. When comparing the amino acid sequence to human MYH proteins, myh9a, myh9b, myh10,

myh11a, myh11b and myh14 show greater similarity to non-slow human protein MYH14 (Fig 4.4, Appendix 4.2) and were therefore eliminated from the list of candidate proteins. Although smyhc1-5, myh7, myh7l and myh6 show the highest sequence identity to human MYH7 from BLAST analysis, sequence identity alone was not able to describe whether zebrafish genes are closely related to human MYH6 or MYH7. Whether zebrafish genes evolved from a pre-existing MYH6/7 before lobe-finned and teleost separation or whether zebrafish proteins derived from a single ancestral slow MyHC and diverged differently to mammals and amphibia.

Table 4.1. List of HsMYH7 candidate genes from protein BLAST analysis.

Gene Name	Protein Sequence Identity (%)
smyhc1	85.15
smyhc2	85.35
smyhc3	86.08
CU633479.4 (smyhc4)	85.8
CU633479.3 (smyhc5)	86.28
myh7	86.14
myh7l	86.23
myh6	82.45
myha	77.02
myhb	78.12
myhz1.1	76.98
myhz1.2	77.03
myhz1.3	77.35
myhz2	77.14
myhc4	77.03
myh7ba	75.27
myh7bb	72.07
myh9a	39.72
myh9b	42.41
myh10	40.89
myh11a	41.56
myh11b	38.17
myh14	39.31

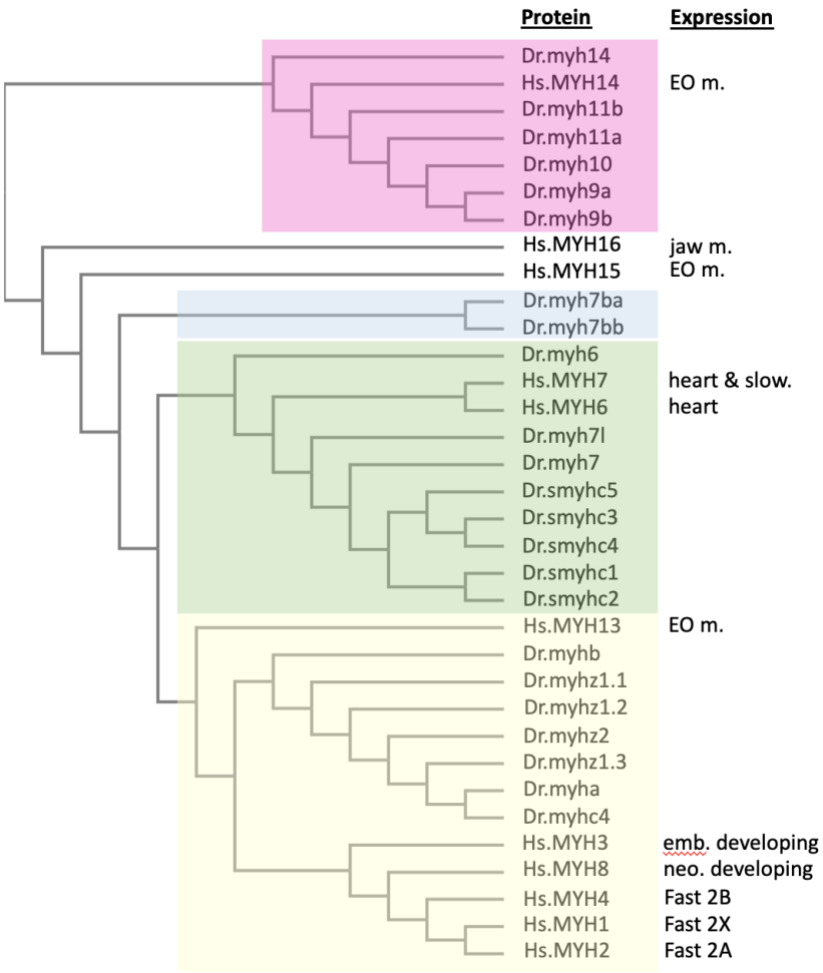


Figure 4.4. Phylogram showing MYH proteins and zebrafish myh proteins.
Cladogram showing CLUSTALO sequence alignment using full amino acid sequences aligned from human MYH proteins: MYH16, MYH15, MYH14, MYH13, MYH8, MYH3, MYH8, MYH4, MYH1, MYH2, MYH6 and MYH7 and zebrafish myh proteins: smyhc1-5, myh7, myh7l, myh6, myha, myhb, myhz1.1, myhz1.2, myhz1.3, myhz2, myhc4, myh9a, myh9b, myh10, myh11a, myh11b and myh14. Full sequence alignment in Appendix 4.2.

4.2.2. MYH6 and MYH7 diverged before lobe-finned and teleost separation

The presence of multiple zebrafish genes compared to two genes in lobe-finned lineage raises the question of how they arose during evolution. Whether there are more duplicates of the genes due to whole-genome duplication events and whether the common ancestor already had both MYH6 and MYH7 or a single slow MyHC. To identify whether the common ancestor of humans and zebrafish have both MYH6 and MYH7 or only a single ancestral slow MyHC, I performed a broader phylogenetic analysis incorporating information from across 76 MYH7-related proteins from a range of animal species (Fig 4.5). In mammals, the divergence of an ancestral slow MyHC gene formed just two MyHC branches, MYH6 and MYH7. In both xenopus and coelacanth, both show divergence of an ancestral slow MYH7 gene form two branches a myh6 and a myh7 branch. In ray-finned lineage (consisting of commonly known bony fish), the divergence of ancestral slow MyHC show more than two branches but form two main clusters, a first cluster consisting of smyhc1-5, myh7 and myh7l and a second

cluster consisting of myh6 (Fig 4.5). The formation of two clusters in ray-finned lineage suggests the ancestor had both myh6 and myh7 and one of these genes duplicated to form smyhc1-5, myh7 and myh7l. Since the myh6 cluster and the smyhc1-5, myh7 and myh7l clusters are derived from an ancestral MYH6/7 protein, it was unclear whether the myh6 cluster is closer related to mammalian MYH6 and smyhc1-5, myh7 and myh7l cluster to MYH7 or vice versa. Analysis of key amino acids that describe the differences between the MYH6/7 may indicate whether the nomenclature given to fish MyHC proteins resembles the nomenclature given to mammalian MyHC proteins. Although the formation of two main ray-finned lineage clusters can be observed, a myh6 cluster and a smyhc1-5, myh7 and myh7l cluster, the first cluster consisting of a smyhc1-5/myh7 branch and a myh7l branch appear to be closer related to MYH6/7 in mammals than the second myh6 cluster. In this first cluster alone, the smyhc1-5/myh7 branch and myh7l branch may resemble the divergence of MYH6 and MYH7 seen in mammals (Fig 4.5). This may suggest that the first cluster alone may be closely related to MYH6/7 and ray-finned myh6 may have evolved independently of ancestral MYH6/7. However, both smyhc1-5, myh7 and myh7l cluster and myh6 cluster do not cluster with fast MyHC from fish or mammals. Zebrafish myha, myhb, myhz1.1, myhz1.2, myhz1.3, myhz2 and myhc4 all cluster with mammalian fast MyHC proteins (Fig 4.5) suggesting that these genes are closely related to fast MyHC genes than they are to MYH6/7 and that there was an ancestral divergence of fast and slow. In conclusion, the common ancestor of humans and zebrafish have both an MYH6 and MYH7 and in zebrafish, both a myh6 cluster and a smyhc1-5, myh7 and myh7l cluster are closely related to human MYH6/7 but further analysis in protein sequence will be required to identify which of these clusters is closer related to MYH7.

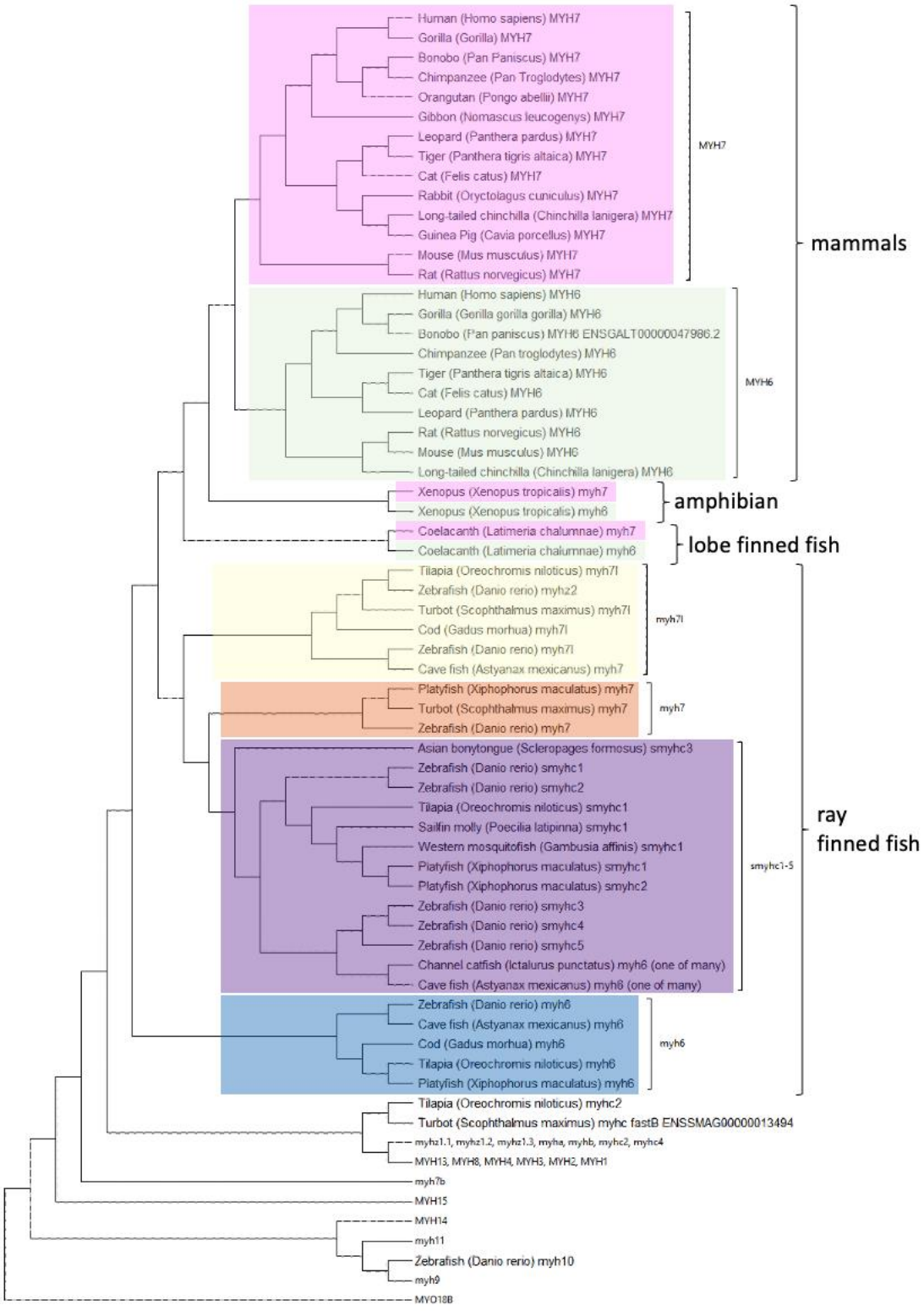


Figure 4.5. Phylogenetic neighbour-joining tree analysis of MYH6 and MYH7 related genes of 76 proteins across several vertebrates.

Phylogenetic neighbour-joining tree of MYH6 and MYH7 related genes. Mammalian amphibian and lobe-finned lineage show divergence in evolution to form MYH7 and MYH6 branches. Ray-finned lineage show more than one divergence to form myh6, myh7, myh7l and smyhc1-5. Phylogenetic neighbour-joining tree using protein sequence alignments made using MEGA-X Software (https://www.megasoftware.net/home).

4.2.3. Amino acid sequences unique to MYH7 are found in smyhc1-5, myh7 and myh7l

The nomenclature of zebrafish genes may not be correctly named to the corresponding human gene name. An analysis of amino acids unique to MYH6 vs MYH7 within mammalian proteins alone will first describe the divergence between mammalian MYH6/7. These amino acids can be used to compare MyHCs from ray-finned fish to describe whether zebrafish myh6 resemble human MYH6 and zebrafish smyhc1-5, myh7 and myh7l to human MYH7.

To identify whether ray-finned myh6 branch or smyhc1-5, myh7 and myh7l branches are more closely related to human MYH7, I investigated the amino acid sequences from lobe-finned MYH6 and MYH7 to find amino acids that define the separation between the two proteins. A variable amino acid describes changes in amino acid sequence between homologs. A signature amino acid for MYH6/7 describes a variation at one amino acid site to distinguish only between the two proteins. Protein sequences from MYH6 and MYH7 were aligned using CLUSTALW and initially, variable amino acids were selected that differentiated between the two MyHC proteins (Fig 4.6). There were 29 variable amino acids identified in the mammalian lineage (Fig 4.7A). Variable amino acids were initially determined as shown for amino acid 35 (Fig 4.6). The variable amino acid at this site is Threonine (T) for MYH6 and Lysine (K) for MYH7 across mammals (Fig 4.6). In the same position in the zebrafish smyhc1-5, myh7 and myh7l proteins, the aligned amino acid is K, as in lobe-finned MYH7. In ray-finned fish myh6, the aligned amino acid 35 is T, a sequence identical to mammalian MYH6 (Fig 4.6). Notably, the variable amino acid 35 is also a signature amino acid that can be used to distinguish between MYH6/7 in mammals and teleost fish.

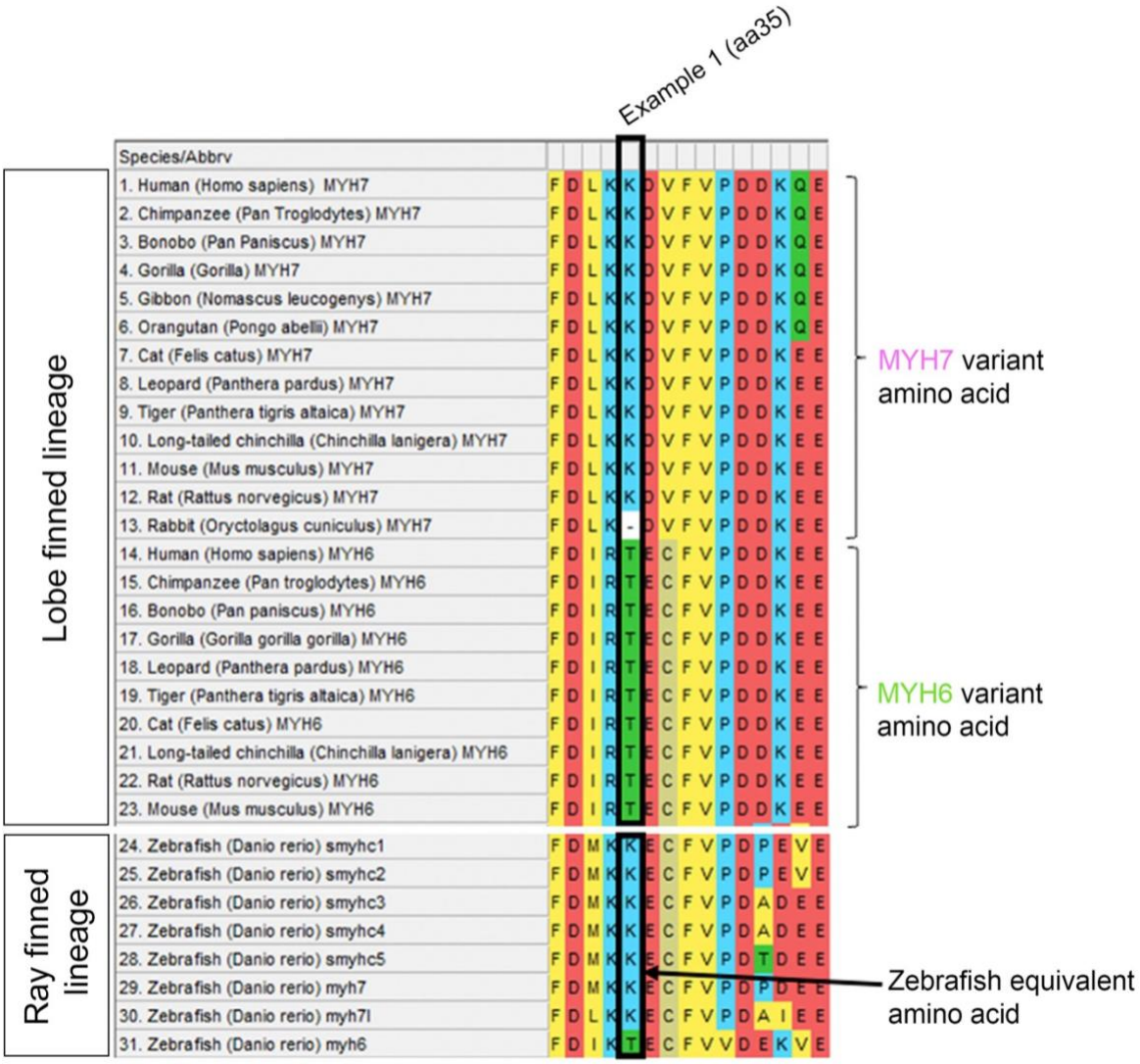


Figure 4.6. A mixture of lobe-finned MYH6 and MYH7 signatures was observed in ray-finned lineage.
A) Example of signature amino acid residue to distinguish between MYH6 and MYH7. Mammalian MYH7 protein sequence shows variant K35 and in mammalian MYH6, the variant T35. In Zebrafish at equivalent amino acid 35, smyhc1-5, myh7 and myh7l show K35 residue and zebrafish myh6 show T36 residue.

Amongst the 29 variable amino acids, I isolated amino acids that were able to distinguish between ray finned smyhc genes, myh7 and myh7l to myh6 (Fig 4.7B). Amongst ray finned genes in Fig 4.7B, %identity scores to either lobe finned MYH6 or MYH7 variable amino acids were determined (Fig 4.7B). The same variable/signature amino acid pattern observed in example aa 35 continues in zebrafish MyHC proteins at amino acids 282 and 318 in the S1 region and 1111 in the S2 region with amino acids D-T-L present in zebrafish smyhc1-5, myh7 and myh7l identical to mammalian MYH7 and amino acids N-V-N present in myh6 identical to mammalian MYH6 (Fig 4.7A). At these amino acids, 4 variable amino acids from zebrafish myh6 were identical to mammalian MYH6 and smyhc1-5, myh7 and myh7l to mammalian MYH7 (Fig 4.7A). Variable amino acids 35, 282, 318 and 1111 may be signatures to distinguish between MYH6 and MYH7 across other ray-finned fish however, further analysis from

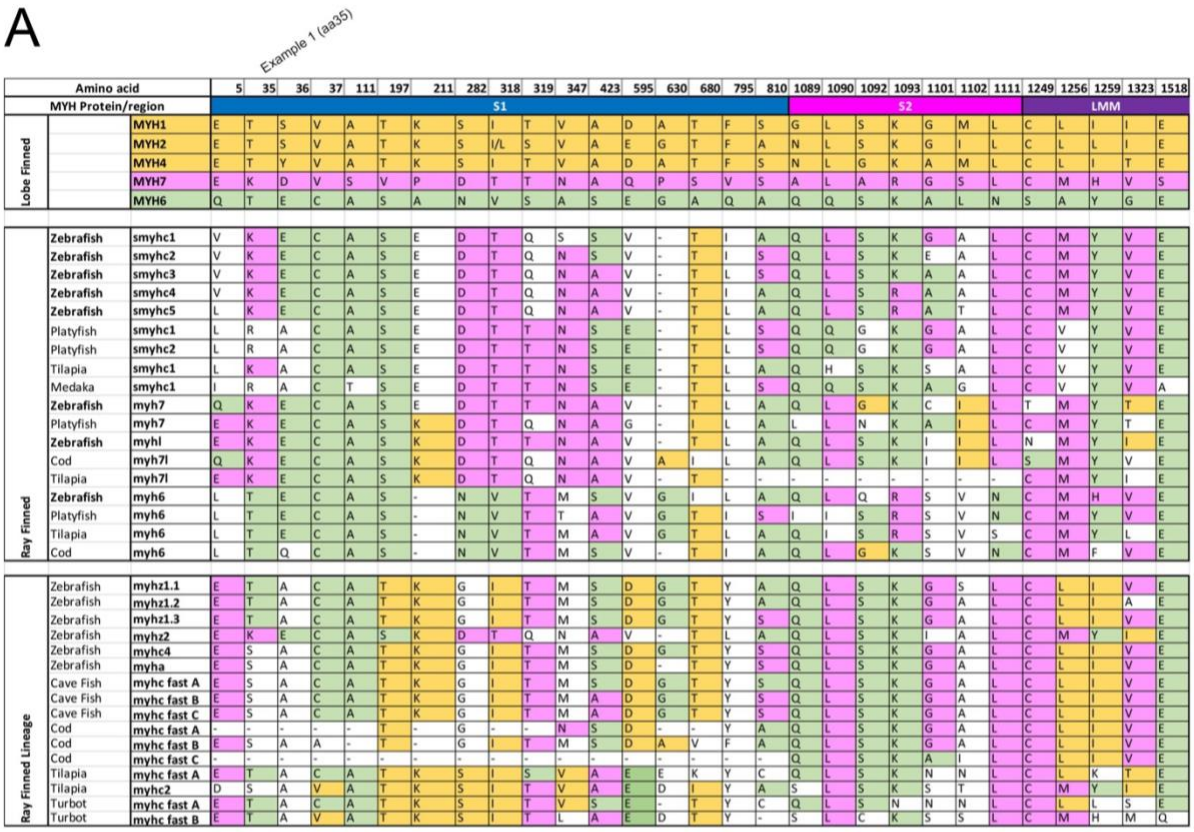
MyHCs from other ray-finned fish will describe which variable amino acids are signatures or ancestral MYH6/7 amino acids. A comparison of the 4 variable amino acids with other ray-finned fish MyHC genes calculated a percentage identity score to mammalian MYH6/7 (Fig 4.7B). Ray-finned MyHCs were categorised into two groups, the first group being smyhc1-5, myh7 and myh7l and the second group of myh6. Amino acids from the two groups were compared to variable amino acids from MYH7 and MYH6 to identify which variable amino acids are signatures to describe MYH6/7 in ray-finned fish and mammals. When comparing ray-finned smyhc1-5, myh7 and myh7l to mammalian MYH7 at amino acids 35, 282, 318 and 1111, the percentage identity for pattern K-D-T-L were 79%, 100%, 100% and 93% respectively (Fig 4.7B). In a comparison of ray-finned myh6 to mammalian MYH6, the percentage identity for pattern T-N-V-N were 100%, 100%, 100% and 75% respectively. Notably, at the same four variable amino acid sites, ray-finned smyhc1-5, myh7 and myh7l show 0% identity to mammalian MYH6 and ray-finned myh6 show 0% identity to mammalian MYH7 (Fig 4.7B). At these four amino acids, these are variable amino acids that can be described as signature amino acids as these amino acids can distinguish between MYH6/7 in both mammals and ray-finned fish. However, there is one counterexample to this pattern at amino acid 1093 in the S2 region where mammals, ray-finned smyhc1-5, myh7 and myh7l are 79% identical MYH6 variant R1093 and ray-finned myh6 is 75% identical mammalian MYH7 variant (Fig 4.7B). The remaining amino acids in Fig 4.7B show a combination of MYH6/7 variant amino acids where there was no clear distinction between smyhc1-5, myh7 and myh7l group or the myh6 group to mammalian MYH6/7. The overall percentage identity to either MYH6 or MYH7 from both ray-finned groups shows that smyhc1-5, myh7 and myh7l are 45% identical to mammalian MYH7 and 23% identical to mammalian MYH6. Ray-finned myh6 is 38% identical to mammalian MYH6 and 18% identical to mammalian MYH7 suggesting the group smyhc1-5, myh7 and myh7l are closely related to MYH7 and group myh6 are closely related to MYH6. Despite considering amino acids in Fig. 4.7B to determine whether ray-finned proteins are more identical to mammalian MYH6/7, the key consideration to distinguish between MYH6 and MYH7 are determined from signature amino acids 35, 282, 318 and 1111. Using these amino acids with *Xenopus, Coelacanth,* and old teleost fish, MYH7 signatures were identified with 100% identity to MYH7 signature K-D-T-L amino acid pattern and MYH6 signatures were identified with 29% identity to MYH6 signature T-N-V-N amino acid pattern (Appendix 4.1). Signature amino acids can distinguish MYH7 proteins from MYH6 proteins, however, MYH6 signature amino acids show divergence in amino acids in teleost, lobe-finned fish, and amphibians (Appendix 4.1).

Examination of further variable amino acid sites showing identical sequences to either only MYH6 or MYH7 can describe which amino acids are ancestral to ray-finned smyhc1-5, myh7, myh7l and myh6.

Amino acids from aa36 to aa197, aa1089, aa1092 and aa1518 show a high identity percentage (<50%) to lobe-finned MYH6 variant with amino acid sequence E-C-A-S-Q-S-E (Fig 4.7Ci). Amino acids aa319, aa1256 and in the LMM region, aa1323 with amino acid sequence T-M-V show high sequence similarity to MYH7 signature and 0% identity to mammalian MYH6 (Fig 4.7Ci). Further comparison of amino acids in Fig. 4.7Ci to fast MyHC (MYH1/2/4) proteins describe whether variant amino acids describe an ancestral to only slow MyHC or ancestral to fast, slow and cardiac MyHCs (Fig 4.7Cii). Amino acids with a high percentage identity for either MYH7, MYH6 or MYH1/2/4 were identified in aa37, aa111, aa319, aa1089 aa1092, aa1249, aa1323 and aa1518 suggesting amino acid sequence in ancestral for slow, fast and cardiac MyHC (Fig 4.7Cii). At amino acids 197 and 1256, ray-finned smyhc1-6 myh7, myh7l and myh6 show identical sequence to mammalian MYH6 and fast ray-finned MyHCs show high amino acid identity to fast MYH1/2/4, suggesting amino acid S197 and Y1256 are ancestral to MYH6/7 and amino acid T197 and L1256 are ancestral to fast MyHC (Fig 4.7Cii). Amino acids in Fig. 4.4C describe ancestral MyHC protein sequences but do not distinguish between MYH6/7 as a signature amino acid.

To summarise, there are 4 signature amino acids describing ray-finned smyhc1-5, myh7 and myh7l with identical amino acids to mammalian MYH7 and ray-finned myh6 to mammalian MYH6 (Fig 4.7B). There were 12 amino acids describing ancestral MyHCs where 10 of these variable amino acids were ancestral to slow, fast and cardiac MyHCs and 2 amino acids were ancestral to only MYH6/7 (Fig 4.7C). When excluding variant amino acids found in ancestral MyHCs in Fig. 4.7C, % identity of remaining variant amino acids in Fig. 4.7B show smyhc1-5, myh7 and myh7l show a higher percentage identity to mammalian MYH7 and less identity to mammalian MYH6 and ray-finned myh6 show higher percentage identity to mammalian MYH6 than to mammalian MYH7. The distinction between the two ray-finned MyHC groups into MYH6 or MYH7 groups indicates that the common ancestor of mammals and ray-finned fish had a distinguished MYH6 and MYH7 present. Since zebrafish smyhc1-5, myh7 and myh7l have a higher % identity to human MYH7 than to human MYH6, smyhc1-5, myh7 and myh7l remain as the zebrafish equivalent gene to human MYH7 and zebrafish myh6 is excluded as this protein show higher identity to mammalian MYH6. Although smyhc1-5, myh7 and myh7l show a high % identity to human MYH7, the phylogenetic tree describes the divergence of these proteins into three branches: smyhc1-5, myh7 and myh7l. To identify whether these genes are orthologous to human MYH7 and whether there are many paralogs in zebrafish due to teleost duplication events and subsequent zebrafish duplication events, synteny of the genes was examined in 4.2.4.

A



		Amino acid	5	35	36	37	111	197	211	282	318	319	347	423	595	630	680	795	810	1089	1090	1092	1093	1101	1102	1111	1249	1256	1259	1323	1518
		MYH Protein/region	S1																	S2									LMM		
Lobe Finned		MYH1	E	T	S	V	A	T	K	S	I	T	V	A	D	A	T	F	S	G	L	S	K	G	M	L	C	L	I	I	E
		MYH2	E	T	S	V	A	T	K	S	I/L	S	V	A	E	G	T	F	A	N	L	S	K	G	I	L	C	L	L	I	E
		MYH4	E	T	Y	V	A	T	K	S	I	T	V	A	D	A	T	F	S	N	L	G	K	A	M	L	C	L	I	T	E
		MYH7	E	K	D	V	S	V	P	D	T	T	N	A	Q	P	S	V	S	A	L	A	R	G	S	L	C	M	H	V	S
		MYH6	Q	T	E	C	A	S	A	N	V	S	A	S	E	G	A	Q	A	Q	Q	S	K	A	L	N	S	A	Y	G	E
Ray Finned	Zebrafish	smyhc1	V	K	E	C	A	S	E	D	T	Q	S	S	V	-	T	I	A	Q	L	S	K	G	A	L	C	M	Y	V	E
	Zebrafish	smyhc2	V	K	E	C	A	S	E	D	T	Q	N	S	V	-	T	I	S	Q	L	S	K	E	A	L	C	M	Y	V	E
	Zebrafish	smyhc3	V	K	E	C	A	S	E	D	T	Q	N	A	V	-	T	L	S	Q	L	S	K	A	A	L	C	M	Y	V	E
	Zebrafish	smyhc4	V	K	E	C	A	S	E	D	T	Q	N	A	V	-	T	I	A	Q	L	S	R	A	A	L	C	M	Y	V	E
	Zebrafish	smyhc5	L	K	E	C	A	S	E	D	T	Q	N	A	V	-	T	L	A	Q	L	S	R	A	T	L	C	M	Y	V	E
	Platyfish	smyhc1	L	R	A	C	A	S	E	D	T	T	N	S	E	-	T	L	S	Q	Q	G	K	G	A	L	C	V	Y	V	E
	Platyfish	smyhc2	L	R	A	C	A	S	E	D	T	T	N	S	E	-	T	L	S	Q	Q	G	K	G	A	L	C	V	Y	V	E
	Tilapia	smyhc1	L	K	A	C	A	S	E	D	T	T	N	S	E	-	T	L	A	Q	H	S	K	S	A	L	C	V	Y	V	E
	Medaka	smyhc1	I	R	A	C	T	S	E	D	T	T	N	S	E	-	T	L	S	Q	Q	S	K	A	G	L	C	V	Y	V	A
	Zebrafish	myh7	Q	K	E	C	A	S	E	D	T	T	N	A	V	-	T	L	A	Q	L	G	K	C	I	L	T	M	Y	T	E
	Platyfish	myh7	E	K	E	C	A	S	K	D	T	Q	N	A	G	-	I	L	A	L	L	N	K	A	I	L	C	M	Y	T	E
	Zebrafish	myhl	E	K	E	C	A	S	K	D	T	T	N	A	V	-	T	L	A	Q	L	S	K	I	I	L	N	M	Y	I	E
	Cod	myh7l	Q	K	E	C	A	S	K	D	T	Q	N	A	V	A	I	L	A	Q	L	S	K	I	I	L	S	M	Y	V	E
	Tilapia	myh7l	E	K	E	C	A	S	K	D	T	Q	N	A	V	-	T	-	-	-	-	-	-	-	-	-	C	M	Y	I	E
	Zebrafish	myh6	L	T	E	C	A	S	-	N	V	T	M	S	V	G	I	L	A	Q	L	Q	R	S	V	N	C	M	H	V	E
	Platyfish	myh6	L	T	E	C	A	S	-	N	V	T	T	A	V	G	T	I	S	I	I	S	R	S	V	N	C	M	Y	V	E
	Tilapia	myh6	L	T	E	C	A	S	-	N	V	T	M	A	V	G	T	L	A	Q	I	S	R	S	V	S	C	M	Y	L	E
	Cod	myh6	L	T	Q	C	A	S	-	N	V	T	M	S	V	-	T	I	A	Q	L	G	K	S	V	N	C	M	F	V	E
Ray Finned Lineage	Zebrafish	myhz1.1	E	T	A	C	A	T	K	G	I	T	M	S	D	G	T	Y	A	Q	L	S	K	G	S	L	C	L	I	V	E
	Zebrafish	myhz1.2	E	T	A	C	A	T	K	G	I	T	M	S	D	G	T	Y	A	Q	L	S	K	G	A	L	C	L	I	A	E
	Zebrafish	myhz1.3	E	T	A	C	A	T	K	G	I	T	M	S	D	G	T	Y	S	Q	L	S	K	G	A	L	C	L	I	V	E
	Zebrafish	myhz2	E	K	E	C	A	S	K	D	T	Q	N	A	V	-	T	L	A	Q	L	S	K	I	A	L	C	M	Y	I	E
	Zebrafish	myhc4	E	S	A	C	A	T	K	G	I	T	M	S	D	G	T	Y	S	Q	L	S	K	G	A	L	C	L	I	V	E
	Zebrafish	myha	E	S	A	C	A	T	K	G	I	T	M	S	D	-	T	Y	S	Q	L	S	K	G	A	L	C	L	I	V	E
	Cave Fish	myhc fast A	E	S	A	C	A	T	K	G	I	T	M	S	D	G	T	Y	S	Q	L	S	K	G	A	L	C	L	I	V	E
	Cave Fish	myhc fast B	E	S	A	C	A	T	K	G	I	T	M	A	D	G	T	Y	S	Q	L	S	K	G	A	L	C	L	I	V	E
	Cave Fish	myhc fast C	E	S	A	C	A	T	K	G	I	T	M	A	D	G	T	Y	S	Q	L	S	K	G	A	L	C	L	I	V	E
	Cod	myhc fast A	-	-	-	-	-	T	-	G	-	-	N	S	D	-	-	Y	A	Q	L	S	K	G	A	L	C	L	I	V	E
	Cod	myhc fast B	E	S	A	A	-	T	-	G	I	T	M	S	D	A	V	F	A	Q	L	S	K	G	A	L	C	L	I	V	E
	Cod	myhc fast C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Q	L	S	K	A	I	L	C	L	I	V	E
	Tilapia	myhc fast A	E	T	A	C	A	T	K	S	I	S	V	A	E	E	K	Y	C	Q	L	S	K	N	N	L	C	L	K	T	E
	Tilapia	myhc2	D	S	A	V	A	T	K	S	I	T	V	A	E	D	I	Y	A	S	L	S	K	S	T	L	C	M	Y	I	E
	Turbot	myhc fast A	E	T	A	C	A	T	K	S	I	T	V	S	E	-	T	Y	C	Q	L	S	N	N	N	L	C	L	L	S	E
	Turbot	myhc fast B	E	T	A	V	A	T	K	S	I	T	L	A	E	D	T	Y	-	S	L	C	K	S	S	L	C	M	H	M	Q

B

Amino acid	MYH Protein/region	Ray-finned MyHC: smyhc1-5, myh7, myh7l — %ID to MYH7	smyhc1-5, myh7, myh7l — %ID to MYH6	myh6 — %ID to MYH7	myh6 — %ID to MYH6
5	S1	21%	14%	0%	0%
35	S1	79%	0%	0%	100%
211	S1	0%	0%	0%	0%
282	S1	100%	0%	0%	100%
318	S1	100%	0%	0%	100%
347	S1	93%	0%	0%	0%
423	S1	57%	43%	50%	50%
595	S1	0%	29%	0%	0%
630	S1	0%	0%	0%	75%
810	S1	36%	57%	0%	75%
1090	S2	64%	21%	50%	0%
1093	S2	14%	79%	75%	25%
1101	S2	21%	36%	0%	0%
1102	S2	14%	0%	0%	0%
1111	S2	93%	0%	0%	75%
1249	LMM	79%	7%	100%	0%
1259	LMM	0%	100%	25%	50%
Average		45%	23%	18%	38%

Ci)

Amino acid	MYH Protein/region	Ray-finned MyHC: smyhc1-5, myh7, myh7l — %ID to MYH7	smyhc1-5, myh7, myh7l — %ID to MYH6	myh6 — %ID to MYH7	myh6 — %ID to MYH6
36	S1	0%	71%	0%	100%
37	S1	0%	100%	0%	100%
111	S1	0%	93%	0%	100%
197	S1	0%	100%	0%	100%
319	S1	43%	0%	100%	0%
680	S1	0%	0%	0%	0%
795	S1	0%	0%	0%	0%
1089	S2	0%	93%	0%	75%
1092	S2	0%	71%	0%	50%
1256	LMM	71%	0%	100%	0%
1323	LMM	64%	0%	75%	0%
1518	LMM	0%	93%	0%	100%

Cii)

Amino acid	MYH Protein/region	Ray-finned MyHC: %ID to MYH7	%ID to MYH6	%ID to MYH1/2/4
36	S1	42%	0%	0%
37	S1	0%	85%	6%
111	S1	0%	88%	0%
197	S1	0%	58%	39%
319	S1	67%	3%	0%
680	S1	0%	0%	82%
1089	S2	0%	82%	0%
1092	S2	0%	76%	6%
1249	S2	94%	3%	0%
1256	LMM	52%	0%	36%
1323	LMM	0%	67%	12%
1518	LMM	0%	94%	0%

Figure 4.7. The mixture of lobe-finned MYH6 and MYH7 signatures was observed in ray-finned lineage.
A) Species from lobe-finned lineage and ray-finned lineage in CLUSTALW and phylogenetic tree (neighbour joining) were used for the analysis of sequences. There are 29 variable amino acids identified through technique in Fig 4.6. Lobe finned lineage show MYH7 (pink) and MYH6 (green) signatures that distinguish between them as analysed. Residues in zebrafish show a mixture of MYH6 and MYH7 signatures. Amino acid locations are labelled with MYH protein regions S1(blue), S2(pink) and LMM (purple). CLUSTALW alignments were made using MEGA-X Software **B)** Variable amino acids that could categorise MYH6 and MYH7 within both lobe finned and ray finned MyHC genes. **C) i)** amino acids that were unable to categorise ray finned MyHC genes to either the MYH6 or the MYH7 cluster but rather describe the ancestral slow MyHC gene. **ii)** comparison of amino acid %identity between ancestral slow amino acid (MYH6/7) sequence and fast MyHC signatures (yellow).

4.2.4. Zebrafish *smyhc1-5, myh7* and *myh7l* syntenic to human *MYH7*

To identify whether zebrafish smyhc1-5, myh7 and myh7l are orthologous to human *MYH7*, synteny was examined between mammals including humans, gorillas and mice, lizard as an amphibian example and a range of ray-finned fish including zebrafish, mummichog, platyfish and goldfish. *MYH6* and *MYH7* are located next to each other on chromosome 14 with *IL25* and *CMTM5* downstream to *MYH7* and upstream to *MYH6* are *NGDN, ZFHX2, THTPA, AP1G2* and *JPH4,* and this is conserved across the mammals such as the mouse and gorilla (Fig 4.8). Common wall lizards have *NGDN, ZFHX2, THTPA* and *AP1G2* upstream to *MYH7* and *CMTM5* downstream. Almost all flanking genes of lobe-finned lineage are conserved. In ray-finned lineage, Mummichog *smyhc1* have no similar genes downstream but has *ngdn, pabpn, ZFHX4* and *thtpa* upstream to *smyhc1.* Platyfish have *pabpn, ZFHX4* and *thtpa* upstream *smyhc1* but no similar genes downstream. Zebrafish *smyhc1-5* and *myh7/myh7l,* Goldfish *smyhc1-4* and *myh7* and platyfish *myh7/myh7l* share no similar flanking genes to humans, Gorilla, Lizard and mice however there are some conserved flanking genes shared with platyfish *smyhc1* and mummichog *smyhc1.* Thus, upstream flanking genes of zebrafish *smyhc1-5* are shared by goldfish *smyhc1-4*, platyfish and mummichog *smyhc1* and upstream flanking genes of Mummichog *smyhc1* are shared by lobe-finned lineage MYH7.

Zebrafish have two clusters of human MYH7 equivalent genes: a *smyhc1-5* cluster and a *myh7/7l* cluster. To identify whether zebrafish *myh7* and *myh7l* exist from a teleost genome duplication event, synteny was examined between zebrafish *smyhc1-5* and *myh7*/*myh7l.* Zebrafish *smyhc1, smyhc2, smyhc3, smyhc4* and *smyhc5* are located next to each other on chromosome 24 with *KCNH2* downstream to *smyhc1* and *cebp1, wdr48a, scnlab* and *acvr2ba* upstream to *smyhc5*. Zebrafish *myh7* and *myh7l* are present next to each other on chromosome 2 with *kcnh2a* and *map4l* downstream to *myh7* and *wdr48b* and *acvr2bb* are upstream to *myh7l.* To identify whether *smyhc* genes and *myh7/myh7l* exist due to a teleost duplication event or a zebrafish specific duplication event, a synteny analysis between zebrafish *myh7/7l* with goldfish *myh7* and platyfish *myh7/myh7l.* Goldfish have *kcnh2a* and *map4l* downstream *myh7* and *trnau1abp, ano8a, plvapa, nr2f6* and *kcn1a* upstream *myh7.* Platyfish have *map4l* downstream to *myh7* and *trnau1abp* is upstream to *myh7*. Almost all flanking genes of *myh7/myh7l* are shared between zebrafish, goldfish and platyfish suggesting that *smyhc* genes and *myh7/7l* genes exist from a teleost duplication event.

Mammalian *MYH6/7* are found located next to each other but in teleost fish, *smyhc genes* and *myh7/7l* genes are separate from *myh6* genes. Zebrafish have *slc24a29, slc25a47a, dIO3B, ppp2r5cb* and *hsp90aa1.2* downstream *myh6* (Fig 4.9). Atlantic herring, channel catfish and goldfish *myh6* also share

similar flaking genes to zebrafish myh6 showing conservation in teleost fish (Fig 4.9). Humans have *SLC24A29, DIO3B, PPP2R5CB* and *HSP90AA1.2* 594+ genes upstream to *MYH6/7* suggesting there was a chromosome inversion near the MYH6/7 site in teleost fish where *MYH6* separated from *MYH7* where teleost *myh6* is orthologous to mammalian *MYH6*. In conclusion, there are two clusters of zebrafish genes orthologous to human MYH7 which are *smyhc1-5* and *myh7/myh7l* where both clusters of genes are paralogous to each other.

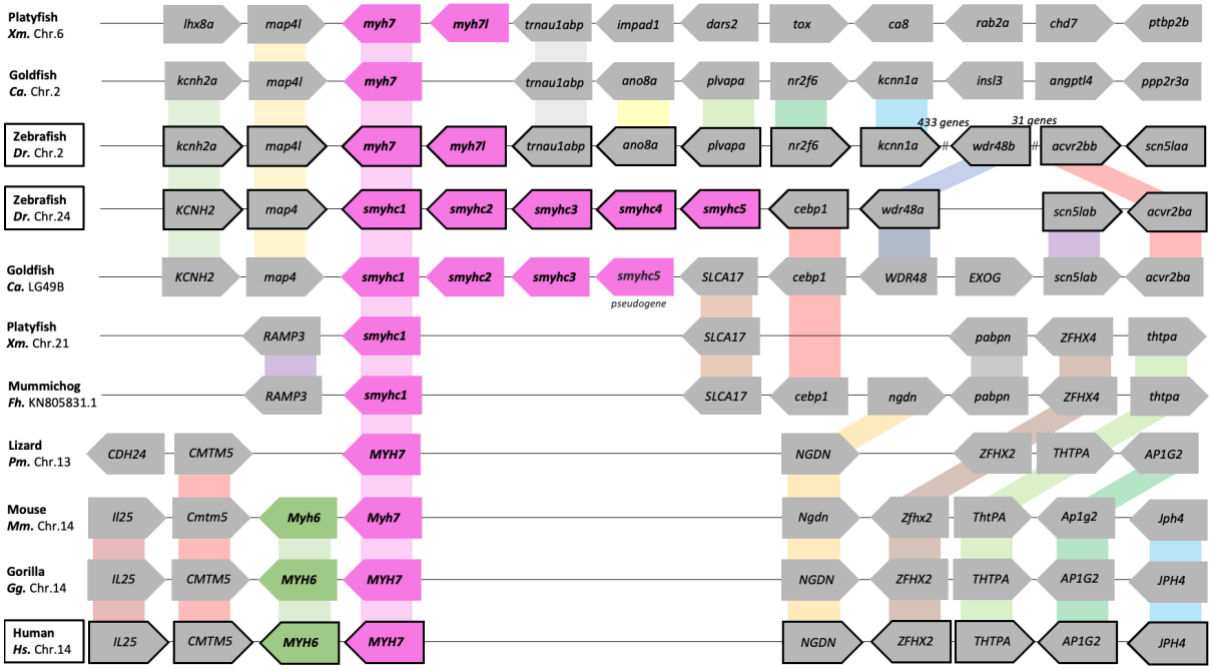


Figure 4.8. Synteny of flanking genes in lobe-finned MYH7 to ray-finned smyhc and myh7 genes.
Colours indicate homologs of genes and all genes present adjacent are directly neighbour genes unless // is present. Chr : chromosome.

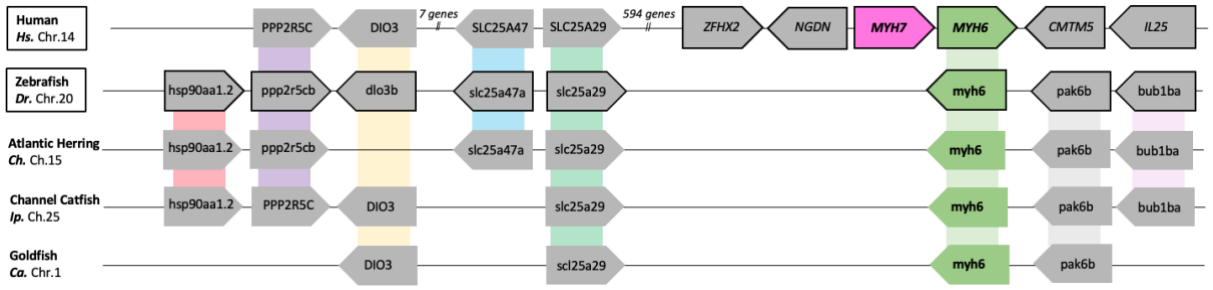


Figure 4.9. Synteny of flanking genes in human MYH6 to ray-finned myh6 genes.
Colours indicate homologs of genes and all genes present adjacent are directly neighbour genes unless // is present. Chr : chromosome.

4.2.5. LDM and MSM mutations affect conserved amino acids in MYH7

Mutations in MYH7 are relatively subtle, for example, one amino acid change or an amino acid deletion. Despite such subtle mutations, they have a huge impact on clinical phenotype and suggest mutations may occur in highly conserved amino acids in the slow myosin LMM region, thus affecting the head positioning of slow myosin shown in chapter 3.2.2. To identify whether LDM or MSM mutations in human MYH7 affect highly conserved amino acids, CLUSTALO protein sequence alignment of the LMM region using sequences from human *MYH7* and zebrafish *smyhc1-5, myh7* and *myh7l* (Fig 4.10). There are 31/41 patient mutations affecting conserved amino acids where 100% sequence identity is shared between human *MYH7* and zebrafish *smyhc1-5, myh7* and *myh7l*. There are 3/41 patient mutations affect highly conserved amino acids but not 100% sequence identity between human *MYH7* and zebrafish *smyhc1-5, myh7* and *myh7l*. Amongst the 3 amino acids affected by patient mutations, amino acid L1492 is present in smyhc1-5 and myh7 but not myh7l, L1646 is present in smyhc1-5 and not myh7 and myh7l, X1936 is present in smyhc1-5 and myh7l but not in myh7l. At amino acids L1492, L1646 and X1936, zebrafish *smyhc1-5* marginally show a higher number of conserved amino acids affected by LDM and MSM mutations than *myh7* and *myh7l* suggesting overall *smyhc1-5* share more conserved amino acids with human *MYH7.* 4/41 patient mutations affect variable amino acids where zebrafish amino acid variant diverged from human *MYH7* amino acid sequence. The majority of LDM and MSM mutations affect highly conserved amino acids in both humans and zebrafish thus, zebrafish *smyhc1-5* share the most similarity in key functional amino acids for myosin function with human *MYH7*.

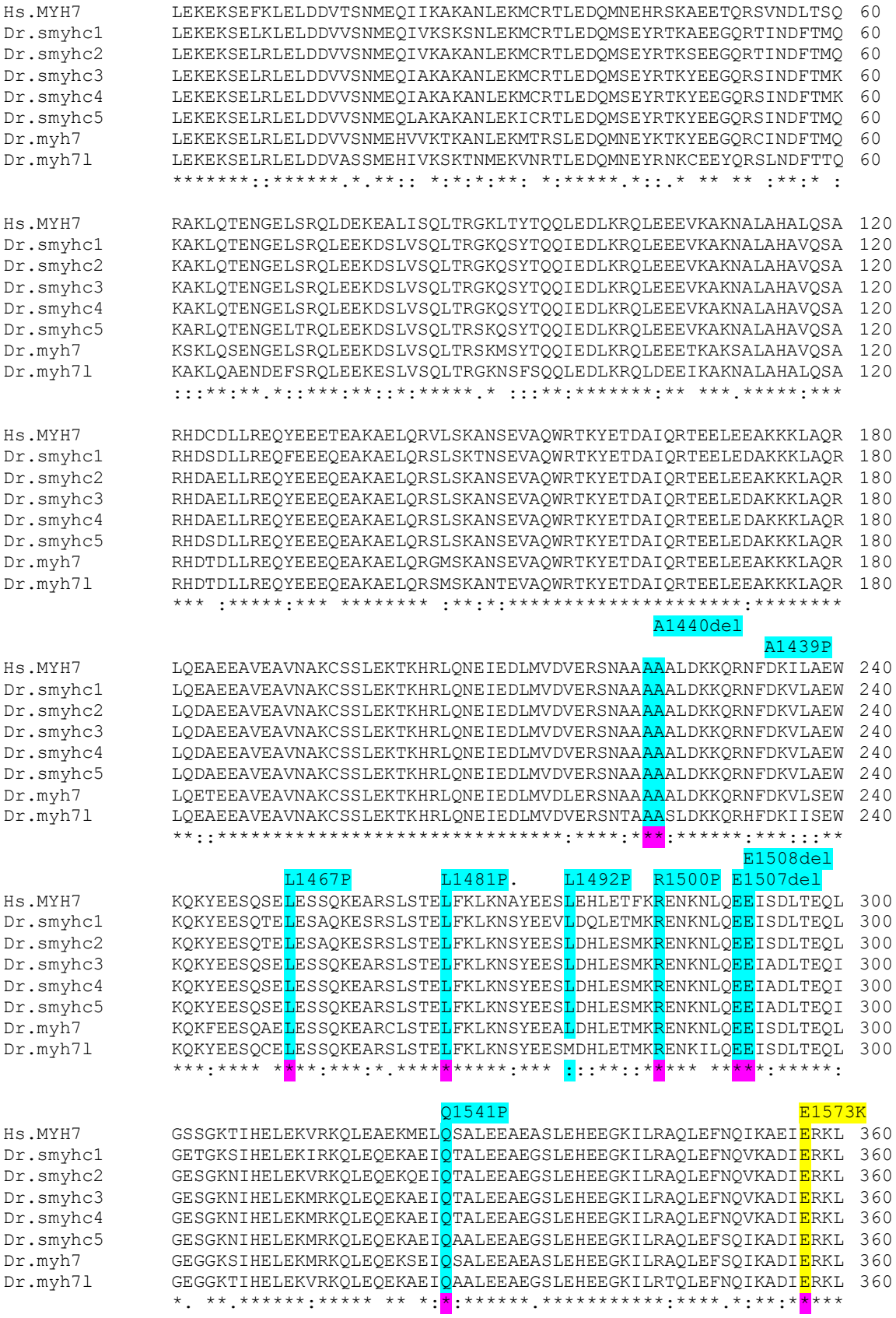


Figure 4.10. CLUSTALO protein sequence alignment of LMM regions from human MYH7 gene with zebrafish smyhc1.

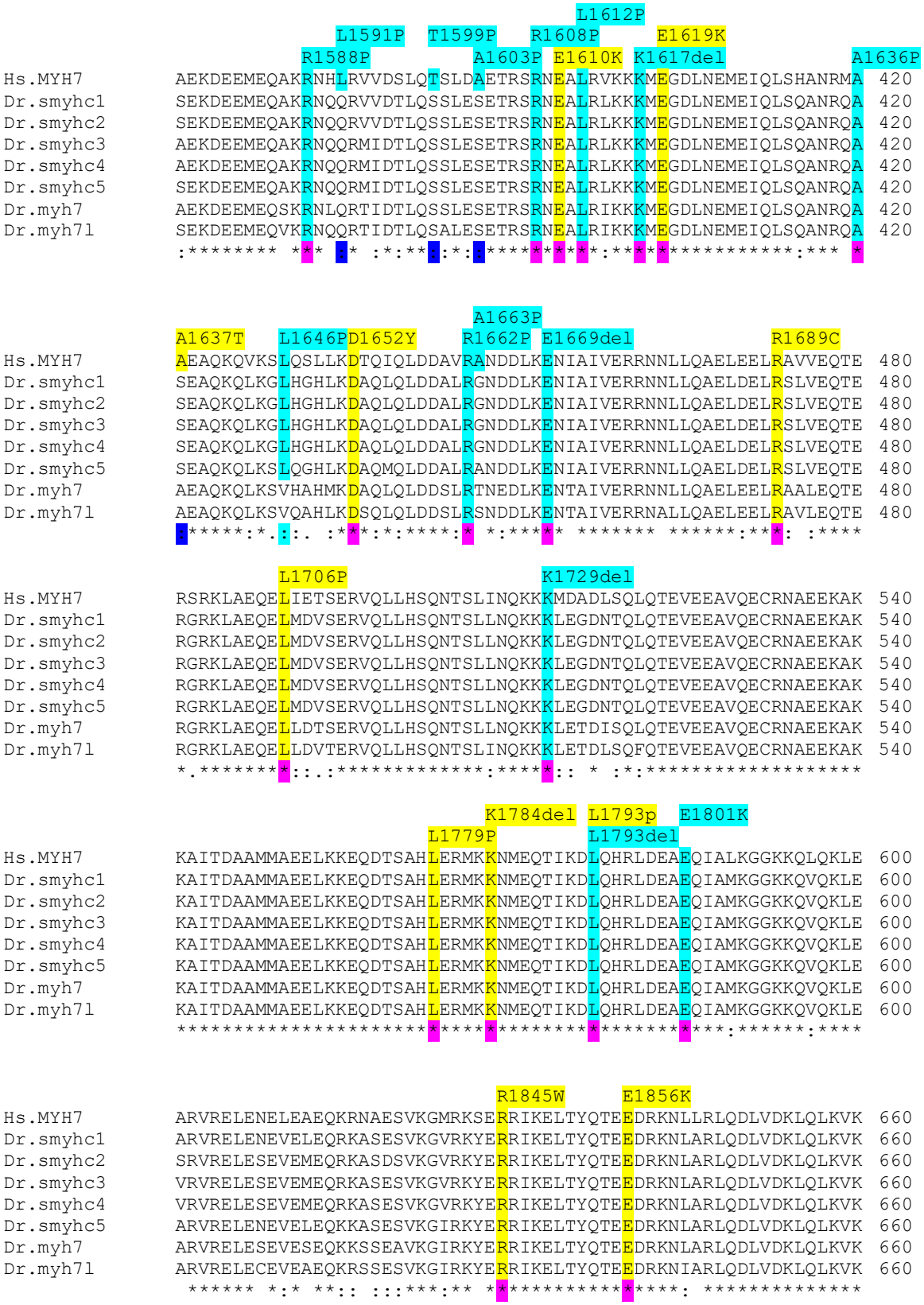


Figure 4.10. CLUSTALO protein sequence alignment of LMM regions from human MYH7 gene with zebrafish smyhc1.

Figure 4.10. CLUSTALO protein sequence alignment of LMM regions from human MYH7 gene with zebrafish smyhc1.
Human MYH7 protein sequence shows high levels of sequence identity in the LMM. Sequence alignment using CLUSTALO (https://www.ebi.ac.uk/Tools/msa/clustalo/). Abbreviations and colour codes: LDM – Laing distal myopathy, MSM – myosin storage myopathy, * - 100% sequence identity, : - 75%+ sequence identity, . - <75% sequence identity.

4.3. Discussion

In this chapter, I look to see which zebrafish gene to target to generate an accurate disease model of mutations found in human *MYH7*. There are several main findings. Firstly, many zebrafish equivalent genes from my initial BLAST search in 4.2.1, suggested candidates were *smyhc1-5, myh7, myh7l* and *myh6.* Secondly, in 4.2.2. protein sequence alignment and drawing the phylogenetic tree shows two main branches for mammalian MYH6 and MYH7 and two main branches for ray-finned fish smyhc1-5/myh7/myh7l and myh6. Thirdly, there are 4 signature amino acid sequences distinguishing MYH6 from MYH7. Ray-finned smyhc1-5, myh7 and myh7l show a higher resemblance to mammalian MYH7 and ray-finned myh6 to mammalian MYH6. Fourthly, I identified gene synteny between zebrafish *smyhc1-5, myh7* and *myh7l* to human *MYH7* and zebrafish *myh6* to human *MYH6*. Lastly, CLUSTALO amino acid alignment between human *MYH7* and zebrafish *smyhc1-5, myh7* and *myh7l* show the majority of LDM and MSM patient mutations affect highly conserved amino acids.

4.3.1. MYH6 and MYH7 existed in the common ancestor of human and zebrafish

Mammalian MYH6 and MYH7 are located next to each other on the same chromosome and exist from a duplication event (Yamauchi-Takihara *et al.*, 1989; Gulick *et al.*, 1991). Consistent with current data, synteny analysis between human, gorilla and mouse show *MYH6* and *MYH7* positioned in tandem with conserved flanking genes (Fig 4.8). However, it was unclear whether the presence of *MYH6* and *MYH7* seen in mammals were conserved in birds, fish and amphibians (Desjardins *et al.*, 2002). Contradictory to Desjardins *et al.* (2002), Ensembl search in tropical claw frogs and coelacanth, *Myh6* and *Myh7* orthologs exist in tandem with conserved flanking genes to mammals suggesting *MYH6/7* are

conserved in amphibian and lobe-finned fish (Appendix 4.1). However, in chickens, there are three slow MyHC genes *MYH15*, *MYH7B* and *MYH7* (Gonzalez-Sanchez and Bader, 1985; Yutzey, Rhee and Bader, 1994; Chen *et al.*, 1997; Machida *et al.*, 2002) where *MYH15* and *MYH7B* are orthologous to mammalian *MYH15* and *MYH7B* (Desjardins *et al.*, 2002) but it was unclear whether *MYH7* may be orthologous to mammalian *MYH6/7* where an *MYH6/7* may have existed ancestral avian lineage and may have been lost during avian evolution.

Zebrafish smyhc1-5 are orthologous to human MYH6/7 (McGuigan, Phillips and Postlethwait, 2004). However, it was difficult to identify whether MYH6 and MYH7 exist in the common ancestor of zebrafish in ray-finned linage or whether there was a unique radiation of ray-finned myh genes from a single ancestral MYH6/7 gene. CLUSTALO protein alignment between a range of teleost fish to mammalian MYH6/7 show teleost smyhc1-5, myh7, myh7l cluster together and teleost myh6 clustered together (Fig 4.5). Teleost smyhc1-5, myh7, and myh7l clusters are orthologous to mammalian MYH6/7 (Fig 4.5) supporting data from McGuigan. (2004) and additionally, the teleost myh6 cluster are orthologous to mammalian MYH6/7 suggesting MYH6/7 exist in the common ancestor of both mammals and teleost. Although teleost smyhc1-5, myh7, myh7l and myh6 cluster with mammalian MYH6/7, teleost smyhc1-5, myh7 and myh7l genes do not exist in tandem with myh6 genes. Synteny analysis show separation of *smyhc1-5, myh7* and *myh7l* from *myh6* genes in ray-finned lineage but not in mammals and amphibians (Fig 4.8, Fig 4.9) suggesting ray-finned lineage *smyhc1-5, myh7, myh7l* and *myh6* genes did not radiate from a single ancestral MYH gene but from an ancestor with pre-existing *MYH6/7*. The present finding rejects the hypothesis that zebrafish do not show conservation of *MYH6/7* in Desjardins. (2002) as our findings suggest the common ancestor of both humans and zebrafish have pre-existing *MYH6/7*. In conclusion, the presence of *MYH6* and *MYH7* in tandem is conserved across mammals, amphibians and lobe-finned fish and in contrast, ray-finned lineage show conservation of *MYH6/7* but a separation of gene location of *MYH6 (myh6)* from *MYH7 (smyhc1-5, myh7, myh7l)* genes.

4.3.2. Zebrafish *smyhc1-5, myh7* and *myh7l* are orthologous to human *MYH7*

Zebrafish *smyhc1-5, myh7, myh7l* and *myh6* are orthologous to human *MYH6/7* however there was no current data to suggest which of these zebrafish myh genes were orthologous to either *MYH6* or to *MYH7.* In mammals, *MYH6/7* have fewer exons than fast and developmental MYH genes. Mammalian *MYH6* do not have an intron 13 and 37 and *MYH7* only lack intron 37 (Liew *et al.*, 1990; Epp *et al.*, 1993; McGuigan, Phillips and Postlethwait, 2004), teleost do not show the same pattern of

missing intron seen in mammals to distinguish between *MYH6/7*. When determining differences between *MYH6/7* amino acids in 4.2.3. there are 4 signature amino acids to distinguish between *MYH6/7* in mammals and ray-finned fish (Fig 4.7). Utilising the 4 signature amino acids, zebrafish *smyhc1-5, myh7* and *myh7l* are shown to be orthologous to human MYH7 and zebrafish *myh6* to human *MYH6*. Present findings complement studies of the expression pattern of human *MYH7* and zebrafish orthologs *smyhc1-3* and *myh7*. Human *MYH7* is expressed both in slow skeletal muscle and in the heart ventricle and Zebrafish have separate myh orthologs expressing *smyhc1-3* only in slow skeletal muscle (Stone Elworthy *et al.*, 2008) and *myh7* is expressed in the heart ventricle (Park *et al.*, 2009). Human *MYH6* is predominantly expressed in the heart atrium and complimenting this expression pattern, zebrafish *myh6* is also expressed in the heart atrium (Huang *et al.*, 2005). Thus, I show zebrafish *smyhc1-5, myh7* and *myh7l* are orthologous to human *MYH7* and not to human *MYH6*.

4.3.3. *smyhc1-5, myh7* and *myh7l* exist from a teleost duplication event

In teleost fish, including zebrafish, there are many slow MyHC orthologs to the one *MYH7* in mammals (McGuigan, Phillips and Postlethwait, 2004; Watabe and Ikeda, 2006; Ikeda *et al.*, 2007) as teleost fish have undergone an additional round of genome duplication (Amores *et al.*, 1998; Meyer and Schartl, 1999; Taylor *et al.*, 2001). Zebrafish *smyhc1-5* share a syntenic relationship to *myh7* and *myh7l* (Fig 4.8) and may have arisen from a teleost duplication event of *smyhc* to *myh7* on another locus has also been observed in goldfish and platyfish (Fig 4.8). Both zebrafish *smyhc* and *myh7* appear to have undergone further tandem duplication to form *smyhc2-5* and *myh7l,* respectively. Platyfish only have one *smyhc* gene and two *myh7* genes, whereas goldfish have four *smyhc* genes and one *myh7* gene (Fig 4.8) suggesting either gene duplication occurred independently in zebrafish in comparison to new teleost fish or gene duplication may have occurred and subsequently lost in new teleost fish. Studies on *smyhc1-5* from McGuigan et al. (2004) show tandemly arrayed genes are either all skeletal myhc genes or all cardiac. *Smyhc1-5* are shown to be paralogs as they have high sequence similarity between *smyhc* genes with minimal gene conversion and intergenic region lengths similar to those in fast skeletal genes (Weiss *et al.*, 1999; McGuigan, Phillips and Postlethwait, 2004). Tandem duplication is a result of more recent gene conversion as there is a more varied number of tandem gene duplication events in platyfish, goldfish and zebrafish. The increased number of tandemly duplicated genes makes it difficult to isolate a single ortholog to human *MYH7* but rather a cluster of zebrafish genes reflects the function of a single gene in humans. Despite the difficulty in isolating one single gene as the zebrafish equivalent gene, zebrafish *smyhc1-3* are only expressed in slow skeletal muscle which may

prove advantageous in studying developmental defects associated with human *MYH7* mutations affecting slow muscle with no cardiomyopathy.

Smyhc1 is predominantly expressed at the early stages of development and subsequently replaced by *smyhc2* and *smyhc3* in adulthood (Stone Elworthy *et al.*, 2008), *smyhc1* may prove advantageous to target for zebrafish KO to study early developmental defects. Anti-sense morpholino (AMO) experiments to knock down *smyhc1* in *Danio rerio* revealed paralysed embryos with defective myosin filament organisation (Codina *et al.*, 2010), and defective M-line organisation (Xu *et al.*, 2012) suggesting a role in slow skeletal muscle at early stages of development. There have been knockdown experiments on zebrafish targeting *smyhc1-4* by co-injection of *smyhc1* AMO, which targeted the 5'UTR of *smyhc1*, and *smyhc2-4* AMO targeting highly conserved coding sequence (in *smyhc2-4*) in the first exon (Naganawa and Hirata, 2011). *Smyhc1-4* knockdown shows no motility following touch at 24 hpf but shows normal burst swimming at 48 hpf (Naganawa and Hirata, 2011). Lack of contraction at 24 hpf followed by contraction at 48 hpf suggest knockdown show an effect on slow muscle and not fast (Naganawa and Hirata, 2011). Knockdown of *smyhc1-4* shows a role in slow skeletal muscle as shown in Codina et al. (2010), however, the additional phenotype was not reported. Mutations in *myh7* producing early stop codons show defects in ventricle contractility respectively (Auman *et al.*, 2007). As no skeletal muscle phenotype was described in these mutants, they may have little role in skeletal muscle thus, *myh7* is not ideal for studying defects in slow muscle. No known phenotype was identified with *myh7l* and *smyhc5.* Overall, present data suggest knockdown of *smyhc1* shows a predominant slow skeletal muscle phenotype and thus, shows a high chance for generating an observable phenotype when creating disease mutations.

4.3.4. Conclusion

In conclusion, I demonstrate that the common ancestor of humans and zebrafish had a pre-existing *MYH6* and *MYH7* and zebrafish *smyhc1-5, myh7, myh7l* and *myh6* are orthologous to this clade. Signature amino acids were able to distinguish between *MYH6* and *MYH7* in mammals and teleost fish where *smyhc1-5, myh7* and *myh7l* are orthologous to mammalian *MYH7*. I provide evidence for a whole-genome duplication event and subsequent gene duplications in zebrafish *smyhc1-5, myh7* and *myh7l* using gene synteny analysis. *Smyhc1* show broad localisation of expression in the slow skeletal muscle at the early stages of development with evidence for early developmental defects in slow muscle (Codina *et al.*, 2010; Xu *et al.*, 2012). In chapter 5, I target *smyhc1* using CRISPR/Cas9 to create null mutations to identify phenotypes associated with loss of *smyhc1* function. By studying the

phenotype associated with *smyhc1* null mutants, it will aid in identifying possible phenotypes associated with more subtle LDM and MSM mutations.

Chapter 5

Studying sarcomere assembly in the absence of *smyhc1*

5.1. Introduction

Laing Distal Myopathy (LDM) and Myosin Storage Myopathy (MSM) mutations affecting early stages of developmental defects are unknown and clinical phenotypes have only been analysed in adults and children. In the previous chapter, I demonstrated that zebrafish *smyhc1* was orthologous to human *MYH7*, particularly in the role in sarcomere assembly during early development. In this chapter, I aimed to assess the role of *smyhc1* in early development using loss of function (LOF) experiments.

During development, myosin molecules self-assemble into thick filaments by interlocking at their C-terminal coiled-coil rod domain (Atkinson and Stewart, 1991; Sohn *et al.*, 1997; Ikebe *et al.*, 2001; Ojima *et al.*, 2015). In mammals, embryonic (MyHC-emb) and neonatal (MyHC-neo) myosin molecules are predominantly present in thick filaments of fast skeletal muscle during early stages of development (Whalen *et al.*, 1981) and are later expressed in a specialized subset of muscles such as the extraocular, masticatory, laryngeal muscles and muscle spindles (Schiaffino *et al.*, 2015). In mice and rats, MyHC-emb is expressed at E9.5, with expression peaking at E15 (Lyons *et al.*, 1990); MyHC-neo is then expressed at E10.5 and peak at 5 days post-birth (Lyons *et al.*, 1990; Lu *et al.*, 1999). These early myosin molecules are later replaced with adult myosin as the animal matures (Lowey, Waller and Bandman, 1991). Specifically, MyHC-2A and MyHC-2X are expressed during mammalian fetal and late fetal stages through to adulthood and MyHC-2B is expressed at postnatal stages through to adulthood (Schiaffino *et al.*, 2015). Thus, embryonic, and neonatal MyHC are predominantly expressed in fast muscle during early development and is replaced by fast skeletal MyHC in adulthood.

Slow MyHC expression shows a different expression pattern to fast MyHC. Mammalian slow myosin (*MYH7*) is expressed during embryonic and fetal developmental stages through to adulthood in slow skeletal muscle and heart ventricles (Narusawa *et al.*, 1987; Schiaffino *et al.*, 2015). Chick embryos express three embryonic MyHCs during early embryonic development and one neonatal MyHC expressed in neonatal developmental stages and both embryonic and neonatal MyHC genes continue to express in skeletal muscle in adulthood in contrast to only a subset of muscle seen in mammals (Bandman and Rosser, 2000). In slow muscle fibres, *MYH15 (SM1), MYH7b (SM2)* and *MYH7 (SM3)* are expressed in skeletal muscle during embryonic development and continue to express in skeletal

muscle through to adulthood (Tidyman, Moore and Bandman, 1997; Rushbrook *et al.*, 1998), alongside the single fast MyHC (Merrifield *et al.*, 1989; Tidyman, Moore and Bandman, 1997). Only pectoral muscles in birds have a complete switch from embryonic MyHC expression to exclusively fast MyHC (Bandman and Rosser, 2000). In zebrafish, there are six fast MyHC genes clustered on chromosome 5. *Fmyhc1.1, fmyhc1.2,* and *fmyhc1.3* are predominantly expressed only in fast skeletal muscles and *fmyhc2.1, fmyhc2.2* and *fmyhc2.3* are expressed in fast skeletal muscles and head muscles. In slow skeletal muscle, *smyhc1* is predominantly expressed during embryonic development (Stone Elworthy *et al.*, 2008). Zebrafish *smyhc2* and *smyhc3* are expressed in a subset of muscle at the early stages of development but their expression replaces *smyhc1* in slow skeletal muscle in adulthood (Stone Elworthy *et al.*, 2008). Knockdown of zebrafish *smyhc1* results in paralysis at 24 hours post fertilisation (hpf) and slow muscles that show defective thick and thin filament assembly (Codina *et al.*, 2010; Xu *et al.*, 2012). Zebrafish *smyhc1* morphants also show loss of myomesin-3 localisation in slow muscles (Xu *et al.*, 2012). However, myomesin-3 knockout (KO) show no effect on the sarcomere assembly of thick and thin filaments (Xu *et al.*, 2012). Knock out of zebrafish *smyhc1* also show paralysis at 24 hpf and sarcomeres in slow fibres show no thick filament and M-lines (Li *et al.*, 2020). Zebrafish *smyhc1* KO mutants show reduced food intake and reduced survival rate of incomplete penetrance. To reconcile the function of *smyhc1*, I investigated the phenotype associated with LOF experiments using CRISPR/Cas9 to knock out *smyhc1* in zebrafish.

Invertebrates show differing expression of MyHC genes compared to vertebrates. *C. elegans* have four MyHC isoforms encoded by five distinct genes for the formation of the muscle in the body wall (Waterston and Francis, 1985; Miller, Stockdale and Karn, 1986). Paramyosin, encoded by *unc-15* is expressed at the core of the A-band where Paramyosin is essential for the base of the thick filament formation (Waterston, Fishpool and Brenner, 1977). MyHC-B, encoded by *unc-54* make up most of the thick filament and positions on the outermost segment of the thick filament (Waterston and Francis, 1985). The middle segment of the thick filament is made up of MyHC-A, encoded by *myo-3*, in between Paramyosin and unc-54 layers (Waterston and Francis, 1985). *Myo-1* and *myo-2* encode MyHC-C and MyHC-D, respectively and are expressed exclusively in the pharyngeal muscle (Miller, Stockdale and Karn, 1986). In *D.melanogaster*, there are at least 14 MyHC isoforms during the embryonic, larval, pupal to adult stages (Bernstein *et al.*, 1983; George, Ober and Emerson, 1989; Hess *et al.*, 2007). Each MyHC isoform is distinct through alternative splicing events of multiple alternative exons at 5 positions of the gene from a single *Mhc* gene (George, Ober and Emerson, 1989; Zhang and Bernstein, 2001) and is controlled using transcriptional regulatory sequences (Arredondo *et al.*, 2001; Kelly, Meadows

and Cripps, 2002; Marín, Rodríguez and Ferrús, 2004; Mas, García-Zaragoza and Cervera, 2004). In early indirect flight muscle (IFM) myogenesis, there is one early MyHC isoform (MyHC-IFM19) containing all the alternate exons, except exon 18 MyHC-IFM18 (Orfanos and Sparrow, 2013). During late IFM myogenesis, MyHC-IFM19 expression declines, but remains at the core of the thick filament (Orfanos and Sparrow, 2013) and the MyHC-IFM18 isoform is predominantly expressed and continue the same expression pattern through to adulthood. MyHC-IFM18 make up the majority of the exterior myosin thick filament while MyHC-IFM19 remains at the core of the filament structure (Hastings and Emerson, 1991; Suggs *et al.*, 2017).

Early developmental phenotypes associated with gene knockout experiments can provide insight into how mutants change myosin function *in vivo*. CRISPR/Cas9 genome editing has been used as a tool to generate null alleles in the zebrafish genome (Chang *et al.*, 2013; Hruscha *et al.*, 2013; Hwang *et al.*, 2013). CRISPR/Cas9 genome editing is made by creating a double-strand break using Cas9 protein at the target gene site and activating DNA damage repair including non-homologous end joining (NHEJ) or homologous recombination (Ran *et al.*, 2013; Chang *et al.*, 2017). Homologous recombination utilises a template DNA for DNA repair and ensures DNA repair is made without mistakes (Ran *et al.*, 2013). NHEJ is an error-prone mechanism for DNA repair whereby insertion and/or deletion (INDEL) mutations are likely to occur (Chang *et al.*, 2017). INDEL mutations can cause frame-shift mutations with premature stop codons and thus produce a non-functional truncated protein or degradation of mRNA by triggering nonsense-mediated decay (NMD) (Lykke-andersen and Jensen, 2015; Hug, Longman and Cáceres, 2016). Using CRISPR/Cas9, the generation specific mutations leading to LDM or MSM in *smyhc1* gene using HR and the generation of *smyhc1* KO lines utilising the NHEJ pathway in the zebrafish genome is possible to investigate the associated phenotype.

Generating LDM and MSM models were tested but not included in this chapter as the methods and strategies involved were not optimal to generate specific mutations. The first limitation in targeting the LMM region is the high level of sequence identity between all *smyhc* genes and *myh7/myh7l* which may lead to further off target mutations but also lead to a mixture of HR and NHEJ between all targets. Prior to injections using short single oligonucleotides (ssoligo), initial tests for cutting using specific gRNA mostly positive (Appendix 5.1). All three gRNA show high levels of mutagenesis detected by HRM analysis when injected with gRNA K1617, K1729 and E1856 but not observed in E1508. A second method to HR to insert point mutations was to use base editing which is a new development of CRISPR-Cas9 was to retain the ability to target specific DNA loci and convert G-C base pairs to A-T base pairs

with the optimal site of base change around -17 to -13 upstream of the PAM site (Zhang *et al.*, 2017a). An expansion of target gene location, a Cas9 variant known as Cas9-VQR recognises 5' NGA as the PAM sequence. BE have been fused to VQR-Cas9 to have BE features and recognise 5' NGA. Studies using fusion BE/Cas9-VQR on zebrafish have demonstrated BE in target sequences (Zhang *et al.*, 2017b). Although utilising HR and base editing were available, screening for the presence of specific mutations were rare. Mutations identified from HR strategies were mainly INDEL mutations as a result from NHEJ and ssoligo insertion were never identified. When using base editing tools, no changes in bases were identified in any of injected embryos. Problems encountered in attempt to create a LDM and MSM model lead me to generate a *smyhc1* KO mutant to understand the role of *smyhc1* in relation to human *MYH7*, but *smyhc1* KO mutants can subsequently be inserted with the defective *smyhc1* or human *MYH7* using expression vectors.

In this chapter, I attempted to use CRISPR/Cas9 with homologous recombination to generate disease mutations for LDM and MSM in *smyhc1* but strategies used failed to generate such mutants. I subsequently generated null mutations in *smyhc1* using CRISPR/Cas9 to examine its role in zebrafish sarcomere assembly. Guide RNAs were targeted to *smyhc1* in exon 2 and exon 4, both alleles leading to frameshift mutations with an early stop codon. *Smyhc1* mutants were viable and fertile. Homozygous *smyhc1* mutants were paralysed from 24 hpf, the time at which pharyngula period begins (Kimmel *et al.*, 1995). Swimming activity resumes at 48 hpf, coinciding with the onset of fast fibre formation. *Smyhc1* mutants show paralysis from 2-20 days post fertilisation (dpf) when treated with N-benzyl-p-toluene sulphonamide (BTS), a drug to block fast myosin activity, revealing slow myosin remains inactive in mutants. Slow muscle-mediated swimming activity resumes at 30 dpf in *smyhc1* mutants. Immunohistochemistry revealed loss of slow myosin at early stages and is compensated at 30 dpf. Work to generate *smyhc1-5* KO mutants is ongoing to knock out all slow myosin function.

5.2. Results

5.2.1. Generation of *smyhc1* mutant alleles

To generate *smyhc1* mutants, CRISPR/Cas9 genome editing with single gRNAs was used to target the second (gRNA1 KO1) or fourth (gRNA2 KO2) coding exons of zebrafish *smyhc1* (Fig 5.1A). Wild type parents lacking any polymorphisms in the CRISPR target site, crossed together, and their embryos injected at one-cell stage with sgRNA and Cas9 protein. Survival of injected embryos up to 5 dpf was high – 72/79 (91%) for gRNAKO1 and 60/65 (92%) for gRNAKO2, compared to 46/50 (92%) and 49/50 (98%) of their respective un-injected siblings. Embryos injected with gRNA KO1 and gRNA KO2 were

35/72 (49%) and 60/60 (100%) positive for rhodamine dextran, indicating successful delivery of CRISPR/Cas9 reagents (Fig 5.1B). Ten embryos from each injected group were screened for mutations using high-resolution melt (HRM) analysis and PCR. Unlike un-injected controls, which displayed a single melt curve peak, the majority (9/10) of DNA samples extracted from gRNA KO1-injected embryos showed a shift in their melting curves, indicating mutagenesis presented as a shouldered or a double peak (Fig 5.1C). All DNA samples extracted from gRNA KO2 injected embryos show a shift in melting curve (10/10) compared to un-injected controls (5/5) (Fig 5.1C). Thus, HRM analysis revealed evidence of successful mutagenesis in embryos injected with either gRNA KO1 or gRNA KO2, a finding that was confirmed by sangar sequencing using primers flanking the *smyhc1* CRISPR target site (Fig 5.1D). The remaining F0 embryos injected with gRNA KO1 and gRNA KO2 were raised to adulthood (Fig 5.1.B).

F0 adults, mosaic for mutations in *smyhc1* were outcrossed to wild-type fish to screen for germline transmission of *smyhc1* mutations to generate F1 heterozygous mutants (Fig 5.2A). There were eight putative F1 lays from F0 adults injected with gRNAKO1 outcrosses and four putative F1 lays from F0 adults injected with gRNAKO2 outcrosses. F1 lays were screened (n=16 per lay) for mutagenesis using HRM analysis and subsequent sequencing (Fig 5.2A). There were 3/8 F1 lays from F0 gRNAKO1 injected parents found to show germline transmission of mutations in F1 progeny with a transmission frequency of 50% (Fig 5.1A). F1 progeny from parents injected with gRNAKO1 show 2 different mutations in exon 2 of *smyhc1*, one with a 4-bp deletion and a second allele with 10-bp deletion where both mutations lead to predicted translation into a truncated protein due to frameshift mutation leading to an early stop codon (Fig 5.2B). When screening for germline transmission of mutations in lays obtained from F0 gRNAKO2 injected parents show ¼ F1 progeny with germline transmission with a transmission frequency of 68.7% (Fig 5.1A). F1 progeny from parents injected with gRNAKO2 show 3 different mutations in exon 4 of *smyhc1*, one with a 3-bp deletion with 1-bp insertion, a second allele with 5-bp deletion where both care frameshift mutations leading to an early stop codon and are predicted to lead to truncated *smyhc1* protein. A third allele with a 3-bp deletion and 12-bp insertion shows an in-frame mutation where an early stop codon was not predicted to be present (Fig 5.2B). Next, both *smyhc1* mutant lines were outcrossed to wild-type fish to minimise possible background mutations in the F2 generation. F2 generation was viable and fertile in both *smyhc1*^{kg179} and *smyhc1*^{kg180} lines and was obtained at expected Mendelian ratios (Fig 5.2C). F2 heterozygous fish were bred to homozygosity at F3, mutants were viable and fertile and obtained at an expected Mendelian ratio of 1:2:1 of wild-type:heterozygous:mutant in both *smyhc1*^{kg179} and *smyhc1*^{kg180} lines (Fig 5.2C).

Overall, *smyhc1*^{kg179} and *smyhc1*^{kg180} were the chosen alleles and bred to homozygosity for LOF analysis.

5.2.2. *Smyhc1*$^{kg179/kg179}$ and *smyhc1*$^{kg180/180}$ mutants are functionally null

Smyhc1 was targeted at the earliest exons by CRISPR/Cas9 genome editing and *smyhc1*^{kg179} and *smyhc1*^{kg180} alleles were isolated. *Smyhc1*^{kg180} contains a 4 bp deletion in exon 2 leading to a frameshift at amino acid 28 and an early stop codon at amino acid 32 after a 5 amino acid nonsense tail thus, predicting a loss of function allele lacking all conserved domains. *Smyhc1*^{kg179} contains a 3 bp deletion and 1 bp insertion in exon 4 leading to a frameshift at amino acid 134 and an early stop codon at amino acid 148 after a 15 amino acid nonsense tail thus, predicting a loss of function allele lacking the majority of conserved motifs (Fig 5.3A). mRNA containing early stop codons are often regulated through a degradation pathway called nonsense-mediated decay (NMD) (Lykke-andersen and Jensen, 2015; Hug, Longman and Cáceres, 2016). NMD involves mRNA to screen non-functional mRNA transcripts by utilising an RNA binding complex called the exon junction complex (EJC). In the absence of an early stop codon, the ECJ is displaced by the ribosome during translation and protein is produced. In the presence of an early stop codon, however, EJCs present downstream of the early stop codon remain and thus trigger NMD for the degradation of mRNA (Hug, Longman and Cáceres, 2016). To screen for evidence that *smyhc1*^{kg179} and *smyhc1*^{kg180} alleles are null, homozygous mutants for *smyhc1* were screened for NMD.

To identify whether NMD of *smyhc1* occurs in *kg179* and *kg180* mutants, whole-mount *in situ* mRNA hybridisation (ISH) analysis was performed on mutants and their sibling controls from *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ in-crosses at 24 hpf. Wild-type and heterozygous siblings from both *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ in-cross show *smyhc1* ISH signal in slow skeletal muscle in the trunk, as expected (Stone Elworthy *et al.*, 2008). Reduction of *smyhc1* ISH signal was observed in both *smyhc1*$^{kg179/kg179}$ and *smyhc1*$^{kg180/180}$ siblings, indicating mRNA is degraded through NMD (Fig 5.3B). Although trace level of *smyhc1* ISH signal can be observed in both *smyhc1*$^{kg179/kg179}$ and *smyhc1*$^{kg180/180}$ mutants (Fig 5.3B) it was unclear whether smyhc1 protein levels were reduced in homozygous mutants compared to their siblings. Slow muscle fibres were analysed from embryos obtained from in-crosses of *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ using F59 to specifically label slow MyHC in embryonic slow fibres (Devoto *et al.*, 1996). Both *smyhc1*$^{kg179/kg179}$ and *smyhc1*$^{kg180/180}$ mutations lead to complete loss of slow fibres compared to wild type and heterozygous siblings (Fig 5.3C). From these data, I conclude that

smyhc1^{kg179} and *smyhc1*^{kg180} are null alleles and are likely to produce a strong loss of function phenotype.

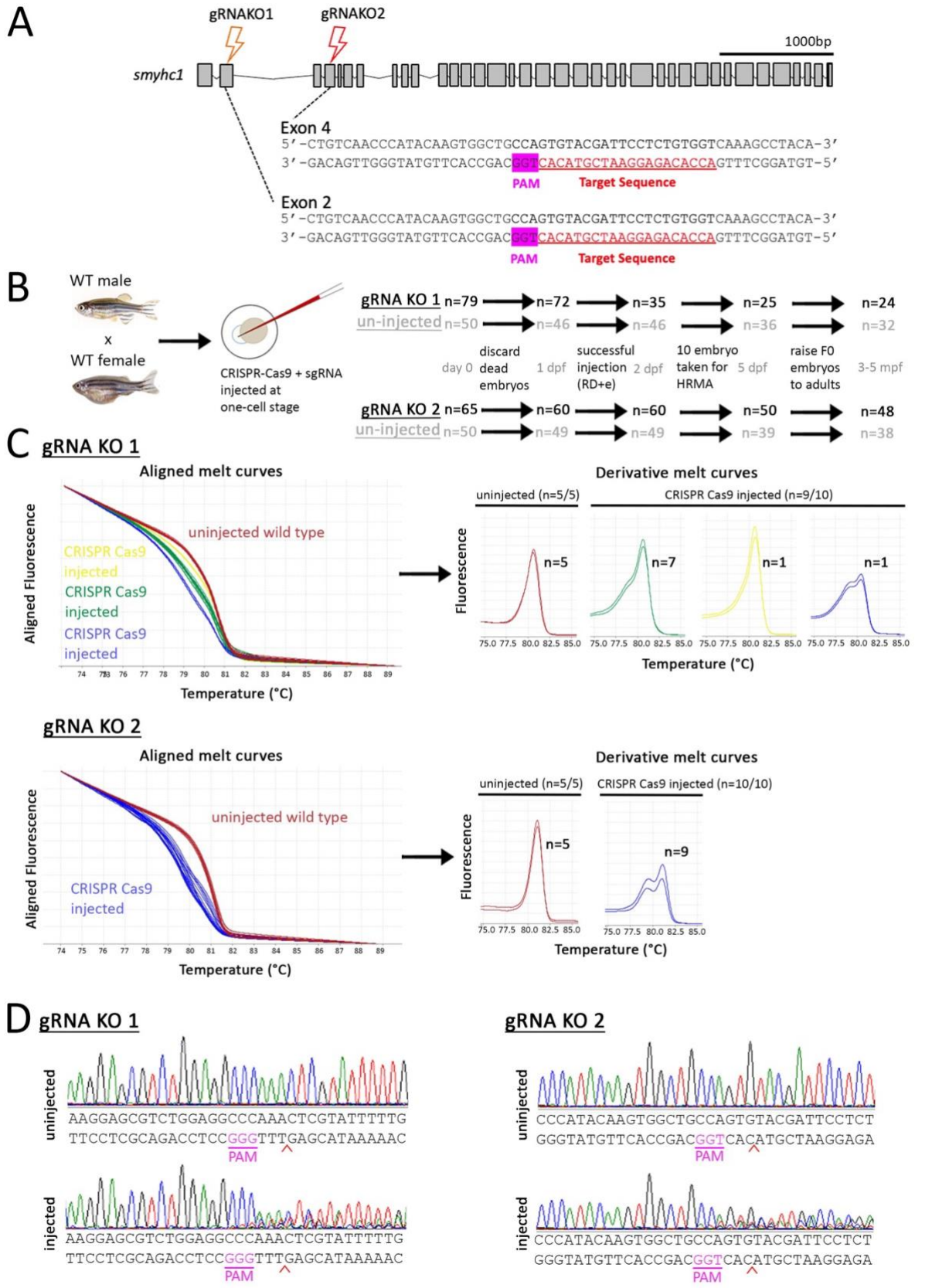


Figure 5.1. CRISPR/Cas9 knockout of *smyhc1*

A) Schematic diagram describing CRISPR/Cas9 design to target *smyhc1* exon 2 and exon 4. CRISPR/Cas9 genome editing consists of a Cas9 protein with a single guide RNA (sgRNA). sgRNA consists of a target sequence (red text and underlined) adjacent to a PAM site (pink highlight). Exons are shown in grey boxes with gRNAKO1 targeting exon 2 and gRNAKO2 targeting exon 4. **B)** Schematic workflow for CRISPR/Cas9 to generate *smyhc1* mutants at F0. In-cross of wild type to give embryos for injection of CRISPR/Cas9 reagents targeting *smyhc1*. Survival of injected and un-injected (controls) embryos described (black text) with the number of embryos from injection to 5 dpf. Un-injected controls (grey text) underneath CRISPR injected data. **C)** High-resolution melt (HRM) curves to screen for mutations in CRISPR/Cas9 injected embryos. Aligned melt curves enabled un-injected embryos (red curves) as a control to differentiate whether mutagenesis occurred in CRISPR/Cas9-injected embryos (yellow, green, or blue curves). Examples of derivative melt curves (right) from un-injected vs CRISPR/Cas9 injected individuals. Un-injected embryo shows melt at approximately 81 °C and injected embryos with gRNAKO1 show shouldered and double peaks at approximately 77-82 °C. Un-injected embryo shows melt at approximately 81 °C and injected embryos with gRNAKO1 show shouldered and double peaks at approximately 78-82 °C. Both gRNAKO1 and gRNAKO2 injected individuals show shouldered and double peaks indicating mutations in *smyhc1* led to the formation of heteroduplex amplicons in HRM analysis. **D)** Example sequencing traces from un-injected and CRISPR/Cas9-injected individuals. Sequence become unreadable 3 bp upstream of PAM site (pink and underlined) where Cas9 protein cuts (red arrow) which indicated random mutagenesis in mosaic F0s.

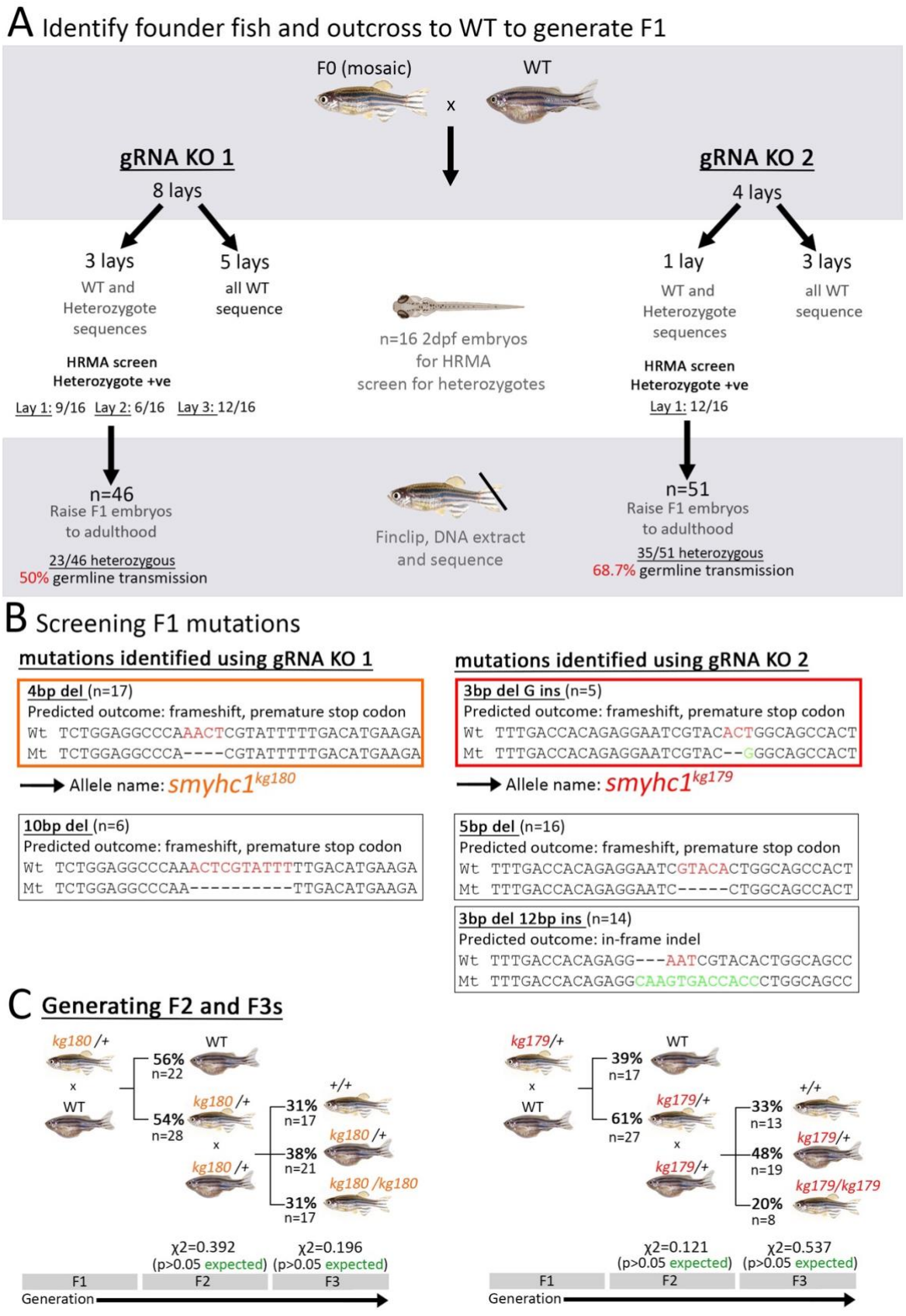


Figure 5.2. CRISPR/Cas9 mutagenesis in F1 embryos.

A) Founder fish identified by crossing mosaic mutant F0 adults to wild-types and screened for germline transmission of mutations using HRM and sequencing. **B)** DNA sequence of mutations identified in F1 adult fin clips. The sequence presented from 5' to 3' and predicted outcome described above sequence. Deletion sequences are shown in red text and insertion sequences are shown in green text. F1 fish with 4 bp deletion in exon 2 were chosen to generate a mutant line called *smyhc1*^{kg180} (orange box) and F1 fish with 3 bp deletion and 1 bp insertion in exon 4 were chosen to generate mutant line *smyhc*^{kg179} (red box). **C)** Heterozygous F1 adults crossed with wild-type to generate wild-type and heterozygous F2 generation to out-cross possible background mutations. F2 heterozygous adults were then in-crossed to generate F3 generation of wild-type, heterozygous and mutants. All crosses generated genotypes to Mendelian ratios when tested with the Chi-squared test ($p>0.05$).

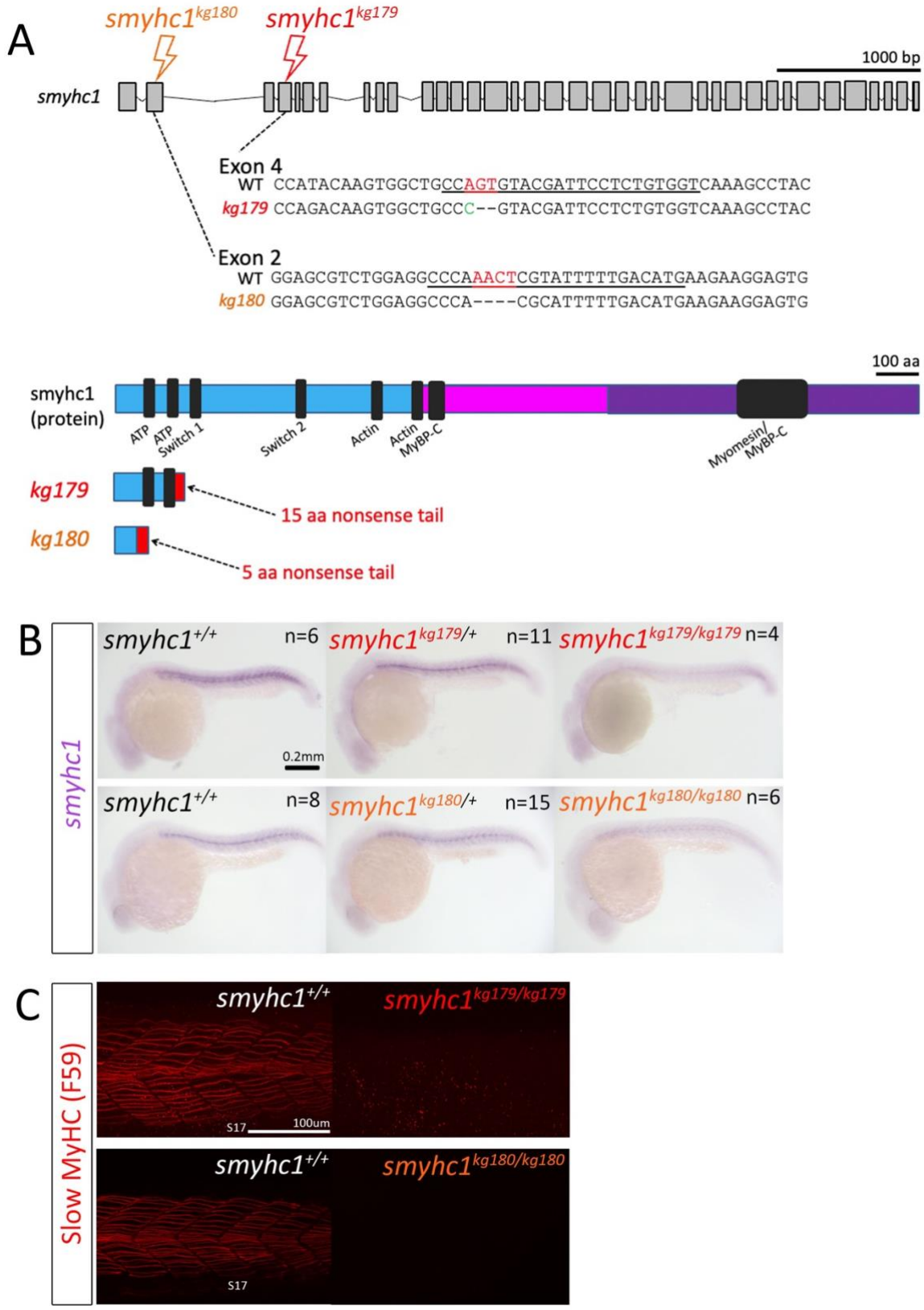


Figure 5.3. Genome editing generates likely null alleles of zebrafish *smyhc1.*

A) Schematic of smyhc1 genes and proteins showing the position of *smyhc1*^{kg179} and *smyhc1*^{kg180} mutant alleles. *Smyhc1*^{kg179} frameshift mutation produces a truncated protein with the first 144 amino acids followed by a 15 amino acid nonsense tail lacking the majority of conserved domains. The *Smyhc1*^{kg180} produces a truncated protein at amino acid 28 followed by a 5 amino acid nonsense tail. **B)** *in situ* RNA hybridisation for smyhc1 mutant and wild type siblings from a *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ in-cross reveals nonsense-mediated decay (NMD) of mutant *smyhc1*^{kg179} and *smyhc1*^{kg180} mRNA AT 24 hpf. In a randomly selected sample from a *smyhc1*$^{kg179/+}$ in-cross, 4/21 were shown to be mutant, 11/21 were heterozygous and 6/21 normal expressors were wild type upon sequence genotyping. From a *smyhc1*$^{kg180/+}$ in-cross, 6/29 embryos were shown to be mutant, 15/29 were heterozygous and 8/29 normal expressors were wild type upon sequence genotyping. Scale bar: 0.2 mm. **C)** Maximum intensity projections of S59 stained 24 hpf embryos showing somites with somite 17 centred and labelled. Wild-type siblings (left) show slow fibre staining and *smyhc1*$^{kg179/kg179}$ and *smyhc1*$^{kg180/kg180}$ mutant siblings show no slow fibre stain (right). Scale bar: 100μm.

5.2.3. *Smyhc1* mutants show no morphological defects and are viable and fertile

Lays from *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ in-crosses were first examined under a bright-field microscope at 1 to 5 dpf to identify any morphological and skeletal muscle defects. Mutant larvae were detected at the expected frequency suggesting null mutations in *smyhc1* are not embryonically lethal (Fig 5.4). In two separate lays of *smyhc1*$^{kg179/+}$ and one lay from *smyhc1*$^{kg180/+}$ in crosses, no change in head, somite, tail, yolk sac, fin, pigmentation, or body length was observed (Fig 5.4). However, consistent with the previous study on antisense morpholino targeting *smyhc1* (Codina *et al.*, 2010; Xu *et al.*, 2012) and in studies from *smyhc1* KO mutants (Li *et al.*, 2020) homozygous mutants for both *kg179* and *kg180* were immotile (Fig 5.5; 21/82 (26%) and 12/52 (23%) immotile embryos, respectively). Thus, lack of Smyhc1 in *kg179* and *kg180* mutants leads to fish that appear immotile but morphologically normal.

Since no obvious morphological defects were observed during the early stages of development, I looked at adult stages to determine whether lack of *smyhc1* affects survival beyond 5 dpf and into adulthood at 4 months. F3 embryos were generated from *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ in-crosses, 100 randomly selected embryos from each cross and were monitored for 4 months. Growth of all siblings from crossed fish was divided into tanks of 50 and mixed-sex and genotype to ensure competition. At 4 mpf, 82% and 94% survival were observed from *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ crossed fish. Any fish that died was fin-clipped and genotyped. Dead fish were a combination of wild type and heterozygous mutants suggesting death did not correlate with the lack of *smyhc1* (Appendix 5.1.4). Genotyping of 4 mpf fish revealed that both lays conformed to Mendelian ratios (Fig 5.6A) and thus, suggested that lack of *smyhc1* is not lethal for zebrafish development to adulthood. At 4 mpf adult fish were examined for morphological defects and males and females were categorised by gender. In both *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ in-crosses, no change in head, body shape, jaw shape, eye shape and colour, tail, fin, or pigmentation pattern were observed (Fig 5.6B). Length and weight

measurements were taken on all fish at 4 mpf; when comparing between siblings, no significant difference in length or weight was observed between sex-matched wild type, heterozygous or mutant siblings (Fig 5.6C). Homozygous *smyhc1*$^{kg179/179}$ and *smyhc1*$^{kg180/kg180}$ mutant males and females were observably fertile. We conclude that wild type *smyhc1* is a non-essential gene for life in an aquarium.

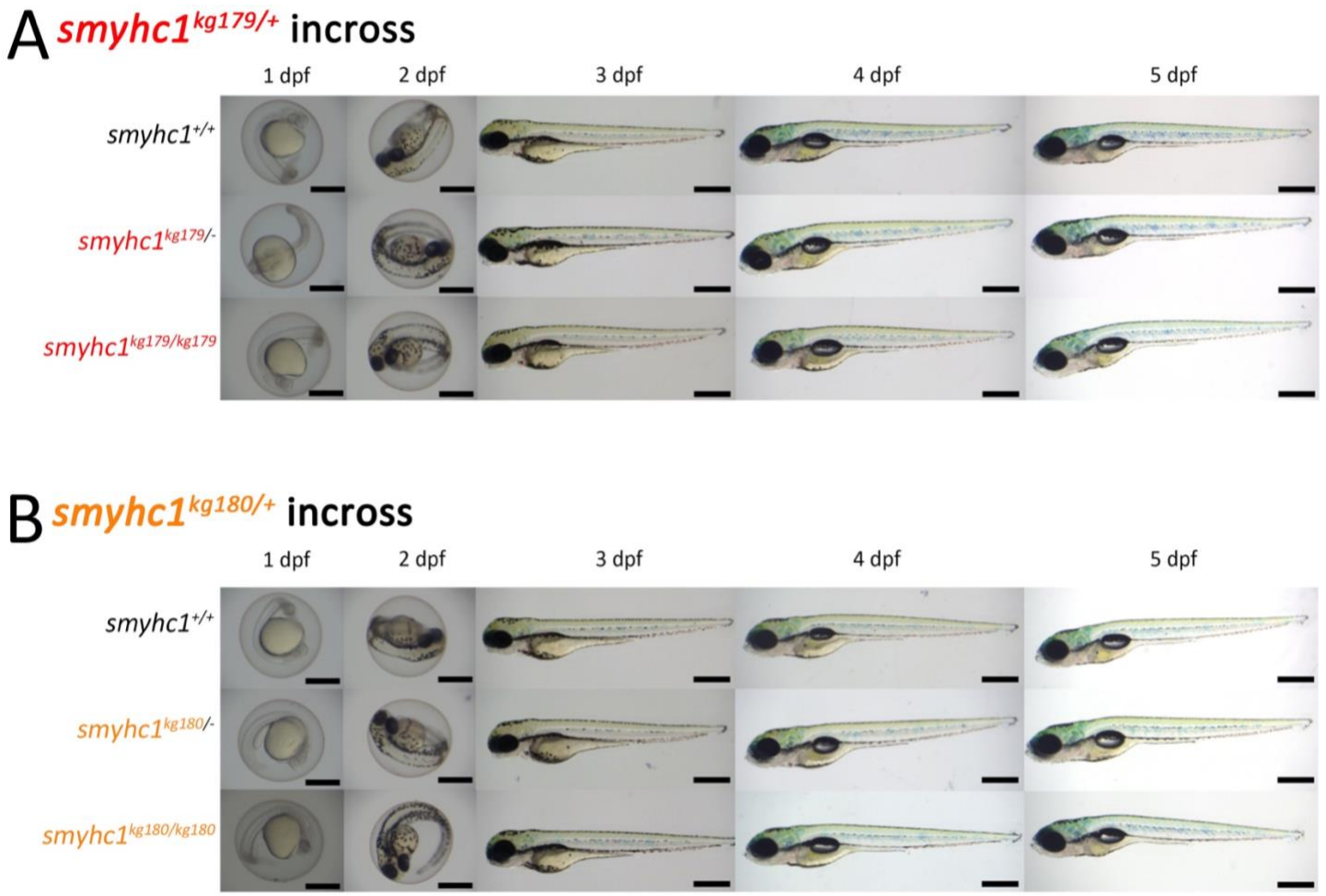


Figure 5.4. Zygotic *smyhc1* mutants show no morphological defects.
A) Bright-field images of 1,2,3,4 and 5 dpf larvae from *smyhc1*$^{kg179/+}$ heterozygous in crosses (wild type n=10, heterozygous n=12, mutant n=2) and **B)** *smyhc1*$^{kg180/+}$ heterozygous in crosses (wild type n=2, heterozygous n=18, mutant n=4). Fish are shown anterior towards the left and dorsal upwards with genotyped heterozygotes and mutants below their respective wild type siblings. Scale bars: 0.5 mm.

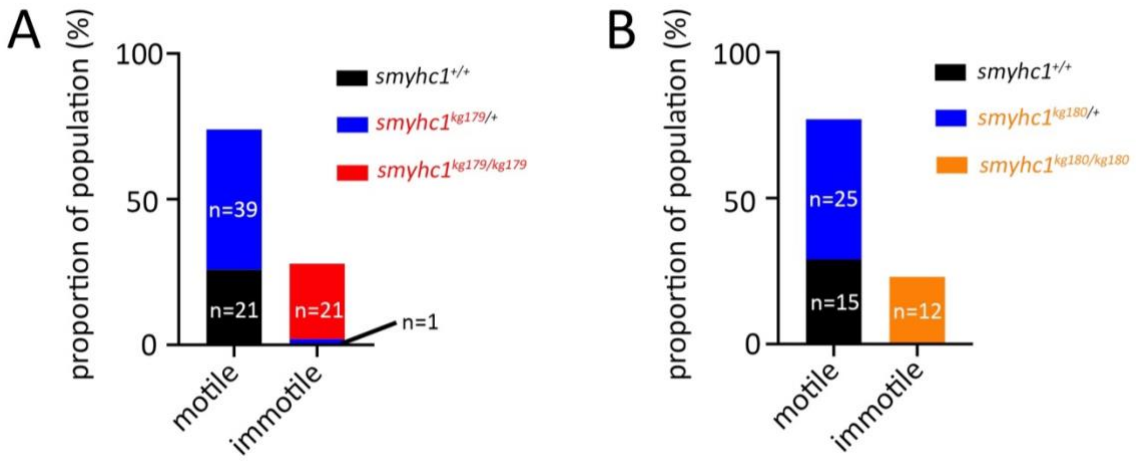


Figure 5.5. Mutation of smyhc1 reduces swimming velocity.
Randomly selected larvae were dechorionated at 24 hpf and examined for presence or absence of tail coiling movement **A)** in *smyhc1*$^{kg179/+}$ in-cross (n=82) and **B)** in *smyhc1*$^{kg180/+}$ in-cross (n=52). The genotype of the fish was revealed after the examination.

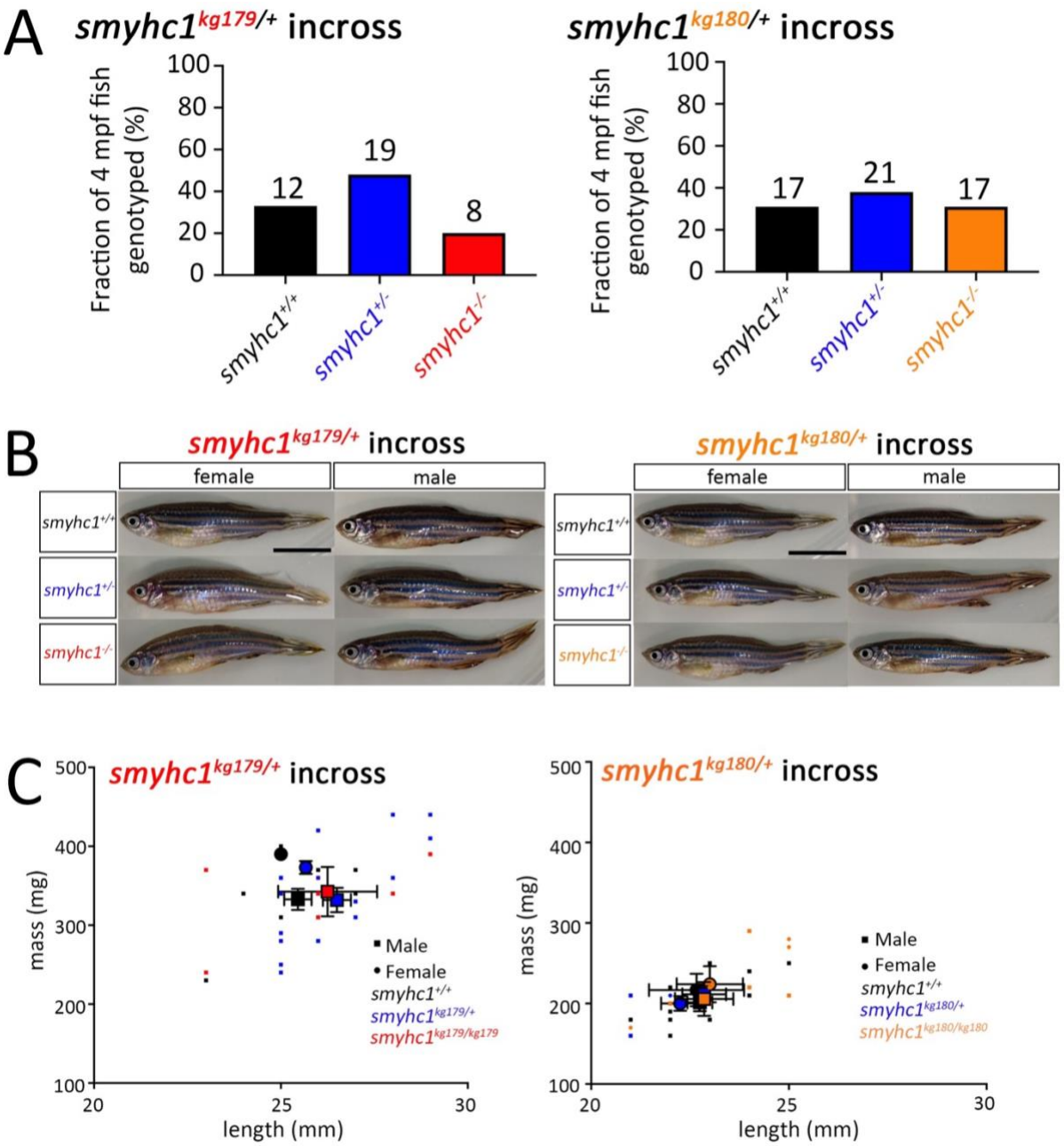


Figure 5.6. Zygotic *smyhc1* mutants survive to adulthood.
A) Adults derived from in crosses of *smyhc1*$^{kg179/+}$ (n=39) and *smyhc1*$^{kg180/+}$ (n=55) fish were genotyped at 4 mpf showing the expected Mendelian ratios. Fish numbers above each bar. **B)** Images of adults derived from in crosses of *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ at 12 mpf. Scale bars: 1 cm. **C)** Length and mass of genotyped siblings from *smyhc1*$^{kg179/+}$ and *smyhc1*$^{kg180/+}$ in crosses at 4 mpf show no significant difference between genotypes. S.E.M. error bars in C. Large symbol reflect means for each sex and genotype and individual data points plotted in small. Overall length and weight were less in adult fish from *smyhc1*$^{kg180/+}$ lay compared to adult fish from *smyhc1*$^{kg179/+}$ in-cross, a difference that may reflect an uncontrolled environmental or genetic background effect.

5.2.4. Movement defects persist in *smyhc1* mutant

We next determined whether the defective movement in *smyhc1* mutants persists beyond 24 hpf. Previous studies have shown immotility in zebrafish at 24 hpf that were AMO knockdown or CRISPR/Cas9 KO of *smyhc1* (Codina *et al.*, 2010; Xu *et al.*, 2012; Li *et al.*, 2020; Whittle *et al.*, 2020). *Smyhc1*$^{kg179/+}$ were in-crossed to generate wild type, heterozygous and homozygous mutant embryos. Chorions were removed at 24 hpf and tail-coiling movement was analysed to categorise motile and immotile fish. By 48 hpf, fast muscle fibres, which do not express *smyhc1*, have assembled striated myofibrils (Stickney, Barresi and Devoto, 2000). At 48 hpf, immotile mutants regained tail muscle motility and appeared to move similarly to wild-type and heterozygous siblings (Fig 5.7A). Nevertheless, to determine whether swimming was affected by the loss of Smyhc1, embryos were examined for swimming velocity upon touch stimulation. Homozygous *smyhc1*$^{kg179/kg179}$ mutants showed significantly reduced swimming velocity compared to their wild type and heterozygous siblings, with mean velocity reduced from 284.5 to 136.35 mm/s-1 (Fig 5.7B1). At 5 dpf, *smyhc1*$^{kg179/kg179}$ mutants continued to show significantly reduced swimming velocity compared to their wild-type and heterozygous siblings, with mean velocity 542.8 to 374.8 (Fig 5.7B2). From 17-30 dpf, there was not a statistically significant difference between *smyhc1*$^{kg179/kg179}$ and their siblings (Fig. 5.7B3-5). Thus, loss of Smyhc1 results in reduced swimming capacity in young larvae.

To examine motility driven by slow fibres, 48 hpf embryos were treated with 50 µM N-benzyl-p-toluene sulphonamide (BTS), an inhibitor for fast muscle myosin II (Cheung *et al.*, 2002; Li and Arner, 2015) and their swimming velocity was recorded (Fig 5.7A). All embryos showed strongly reduced swimming velocity after treatment with BTS (Fig. 5.7B). However, at 2, 5, 17 and 20 dpf homozygous *smyhc1*$^{kg179/kg179}$ mutants were more affected than their wild-type and heterozygous siblings, showing very little twitching or no movement (Fig 5.7B1-4). At 30 dpf, there was no significant difference between homozygous *smyhc1*$^{kg179/kg179}$ mutants and their siblings (Fig 5.7B5). Thus, slow fibre motility remains compromised in young mutant larvae.

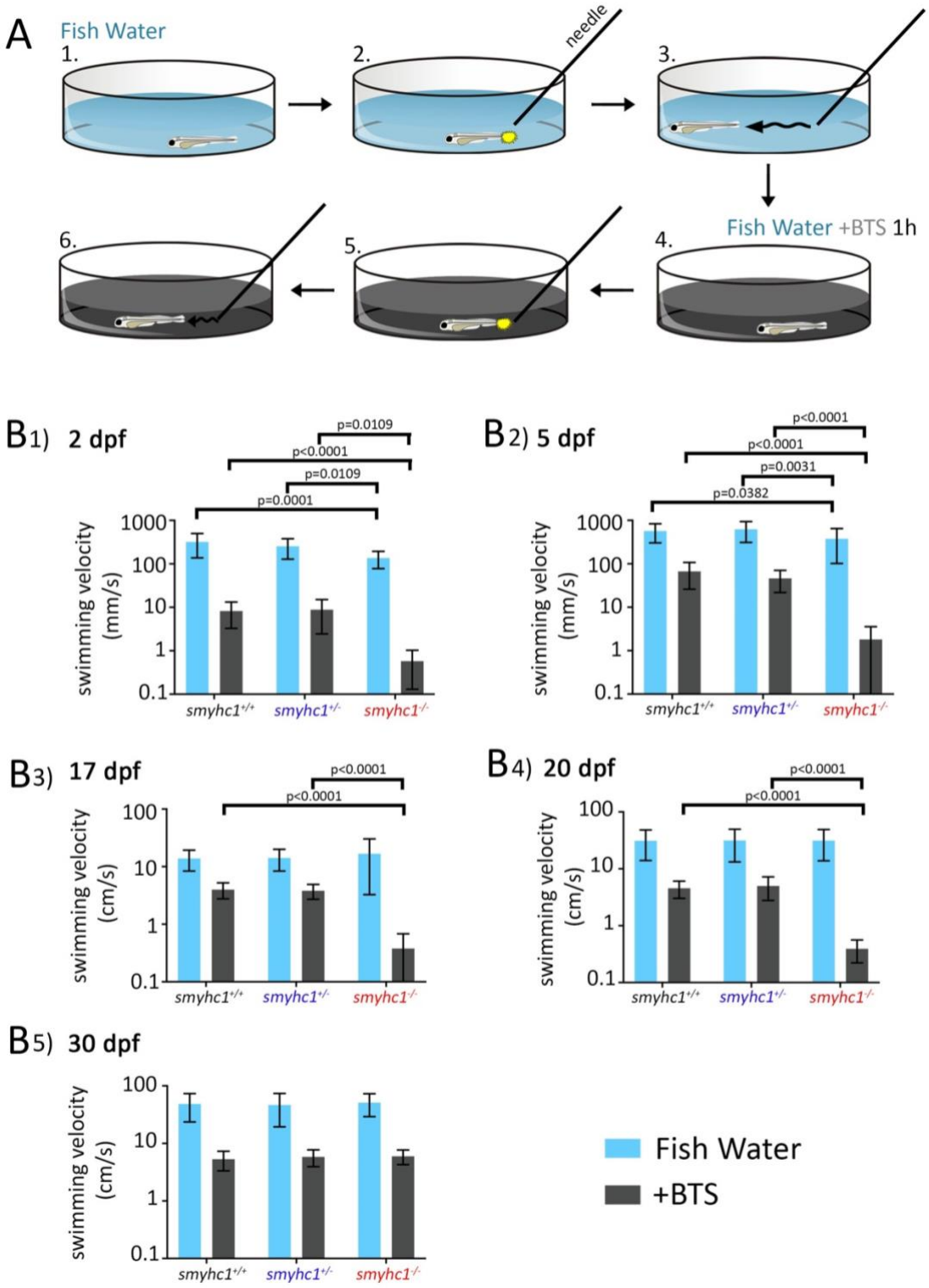


Figure 5.7. Loss of *smyhc1* reduces swimming velocity.
A) Schematic describing workflow to test for swimming velocity in fish water and subsequently fish water with *N*-benzyl-*p*-toluene sulphonamide (BTS). **B)** Zebrafish larvae from *smyhc1*$^{kg179+}$ in-crosses were recorded for swimming activity upon touch stimulation using a needle in fish water (blue bar). Fish were treated with 50 μM BTS for 10 minutes and recorded for swimming activity again (grey bar). Zebrafish larvae tested at **B1)** 2 dpf, **B2)** 5 dpf, **B3)** 17 dpf, **B4)** 20 dpf, **B5)** 30 dpf. Plots obtained using at least 3 separate lays from heterozygous *smyhc1*$^{kg179+}$ in-crosses average numbers in Appendix 5.2. Log10 scale bar on the Y-axis. Error bars ± SD. Statistics using one way ANOVA on GraphPad PRISM.

5.2.5. Defective sarcomere organisation observed in slow fibres

To examine the defects caused by a mutation in *smyhc1*, we used F59 and S58 antibodies which are known to detect MyHC in zebrafish slow muscle fibres at 48 hpf (Crow and Stockdale, 1986; Devoto *et al.*, 1996). At 48 hpf, staining was absent for both antibodies in mutant compared to wild type confirming their reaction with Smyhc1 (Fig 5.8A). Subsequently, however, mutant larvae regained some slow MyHC immunoreactivity (Fig 5.8A-B). At 72 hpf, wild-type larvae have S58 immunoreactivity in the head and trunk muscles (Fig 5.8B). In *smyhc1*^{kg179} mutants, S58 positive slow myofibers continued to be undetectable in either slow or fast somitic trunk and tail muscle, whereas they were readily detected in superficial slow fibres and slow muscle pioneer fibres of siblings (Fig 5.8B). In contrast, S58 immunoreactivity was detected in the head and cardiac muscle of mutants at a level similar to that observed in wild-type siblings (Fig 5.8B). Muscle fibres in three somitic regions continued to show slow MyHC immunoreactivity in mutants, Firstly, low levels of S58 immunoreactivity were presented in thin muscle fibres at the dorsal and ventral somitic extremes. Secondly, thin fibres with weak S58 stains were sometimes present at the horizontal myoseptum near the muscle pioneer fibres (Fig 5.8A). Moreover, specific subsets of muscle fibres thought to be generated from somatically-derived muscle precursor cells (mpcs) also showed S58 immunoreactivity in mutants. The sternohyoid (sh), posterior hypaxial and supracarinalis anterior (sca), inferior obliquus (iob), supracarinalis posterior (scp) and infracarinalis posterior (icp) muscles all stained well (Fig 5.8B). Thus, mutation of smyhc1 prevented slow MyHC accumulation in most somitic muscle fibres, but not in locations where new fibres are produced from matrix metalloproteinases (MMPs).

To examine the defects in other sarcomeric structures caused by a mutation in *smyhc1*, we used phalloidin to detect actin and a-actinin antibody known to detect Z-line structures in zebrafish muscle fibres at 24 hpf. The lack of Smyhc1 in slow fibres allowed us to examine the formation of myofibrils in these cells in the absence of this MyHC. *Smyhc1*$^{kg179/+}$ in-cross lays were examined for actin structure at 24 hpf with phalloidin-Alexa488. In the absence of Smyhc1 protein, we observed that actin filament organisation was severely defective (Fig 5.8B). In wild type siblings, F-actin was organised into sarcomeric thin filament units arrayed at regular intervals along the slow muscle fibre length into myofibrils. In mutant embryos, by contrast, the overall F-actin signal was reduced and disrupted thin filament organisation was observed. Filamentous actin was thin and wavy with actin accumulating at somite borders. Bundles of actin were also observed dotted across the surface of the myotome (Fig 5.8B). Nevertheless, there were a few fibres that showed some regions with organised F-actin filament along the horizontal myoseptum (Fig 5.8B). *Smyhc1*$^{kg179/+}$ in-cross lays were examined for actin

structure at 24 hpf with an a-actinin antibody. In the absence of Smyhc1 protein, Z-disk structures were disorganised (Fig 5.8B). In wild type siblings, z-disks were organised in thin arrays at regular intervals along slow muscle fibres. However, in mutants, the α-actinin signal was overall reduced with accumulation at somite borders. Thin wispy a-actinin elongated from somite borders but is very faint in signal. Thus, lack of Smyhc1 resulted in disorganised sarcomeres in slow fibres.

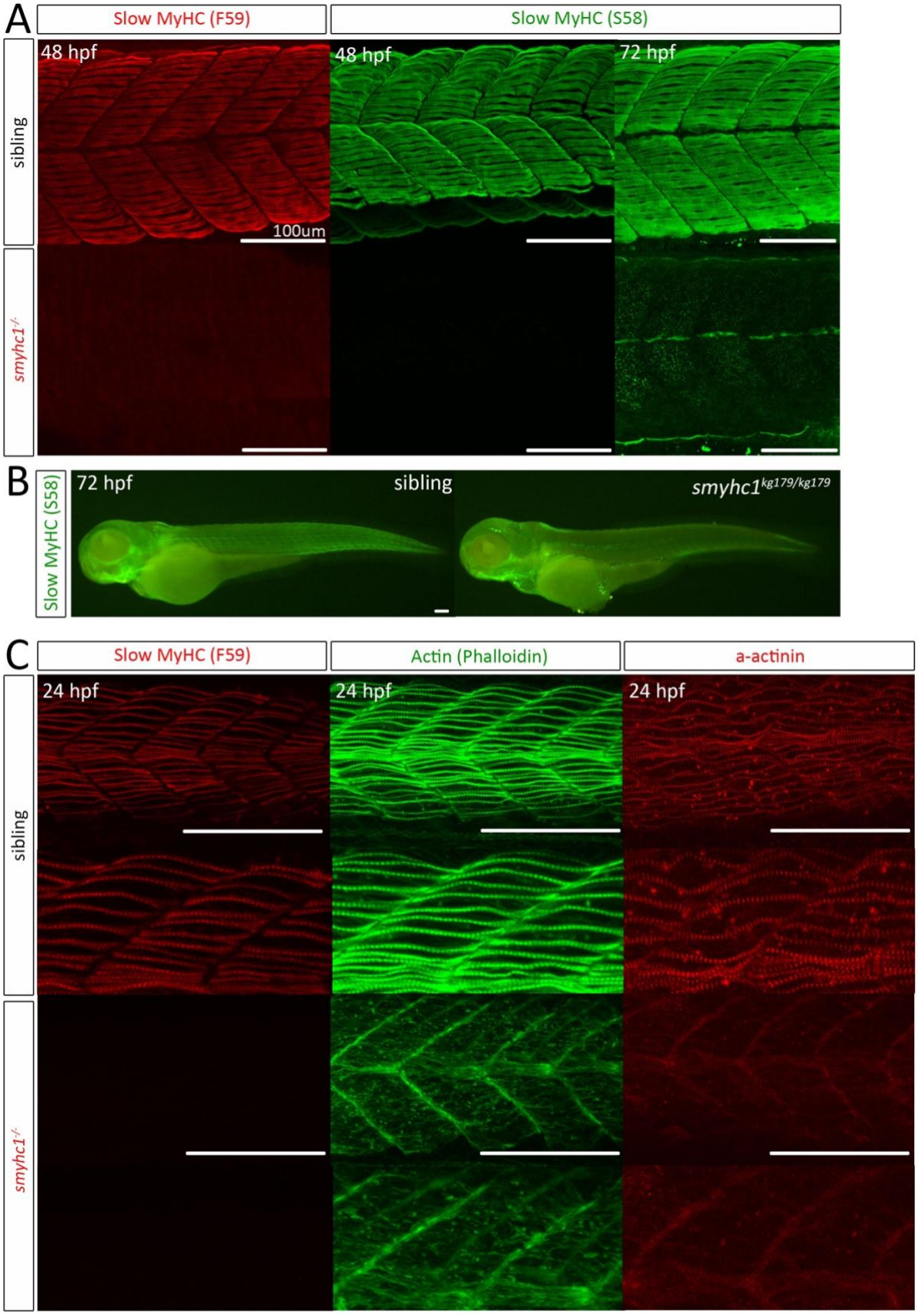


Figure 5.8. Defective sarcomere organisation in *smyhc1*$^{kg179/kg179}$ mutants

A) Slow MyHC immunofluorescence of 48 and 72 hpf larvae from *smyhc1*$^{kg179+}$ in-crosses using F59 and S58 antibodies. Wild type sibling at top and *smyhc1*$^{kg179/kg179}$ mutants below. **B)** Slow MyHC immunofluorescence 72 hpf larvae from *smyhc1*$^{kg179+}$ in-crosses using S58 antibodies. Mutants showing S58 signal exclusively in a subset of muscles: sca, els, iob, sh, scp and icp somite-derived muscles, and dorsal and ventral craniofacial muscles. Wild type sibling on left and *smyhc1*$^{kg179/kg179}$ mutants on right. **C)** Immunofluorescence of 24hpf larvae from *smyhc1*$^{kg179+}$ in-crosses using F59 to label slow MyHC, phalloidin for filamentous actin and a-actinin to mark z-disks. Wild type sibling at top and *smyhc1*$^{kg179/kg179}$ mutants below. Images of wild type and mutant centred on somite 17/18. Abbreviations: dm-head dorsal muscles, els-embryonic lateralis superficialis, icp-infracarinalis posterior, iob- inferior obliquus, sca-supracarinalis anterior, scp-supracarinalis posterior, sh-sternohyoideus; vm-head ventral muscles. Scale bars = 100 µm.

5.2.6. Lack of *smyhc1* does not affect fast fibres

To examine whether defects in fast or slow fibres are caused by a mutation in *smyhc1*, we used the A4.1025 antibody known to detect all MyHC in zebrafish muscle fibres from 24-72 hpf. At 24 and 48 hpf, both wild-type larvae have A4.1025 immunoreactivity in the trunk slow muscle fibres (Fig 5.9A). In *smyhc1*^{kg179} mutant siblings, A4.1025 shows immunoreactivity exclusively in early developing fast muscle fibres at 24 hpf and in 48 hpf fast fibres, no slow fibres were detected using A4.1025 (Fig 5.9A). At 72 hpf, wild-type larvae have A4.1025 immunoreactivity in slow trunk muscle fibres (Fig 5.9B). In *smyhc1*^{kg179} mutant siblings, A4.1025 shows immunoreactivity to fast fibres and a small subset of slow muscles in the els in the horizontal myoseptum and thin muscle fibres at dorsal and ventral somatic extremes (Fig 5.9A). To confirm whether the A4.1025 signal is from fast fibres, co-staining using A4.1025 with F310 on 3 dpf larvae from *smyhc1*$^{kg179/+}$ in-crosses. Wild-type siblings showed A4.1025 immunoreactivity in slow fibres and F310 showed immunoreactivity exclusively in fast muscles (Fig 5.9B). In *smyhc1*^{kg179} mutant siblings, both A4.1025 and F310 showed immunoreactivity exclusively in early developing fast muscle fibres at 24 hpf suggesting mutation in *smyhc1*^{kg179} abolished the majority of slow fibres in the trunk and fast fibres were not affected. Thus, mutation of Smyhc1 abolishes smyhc1 protein but does not affect fast fibre morphology.

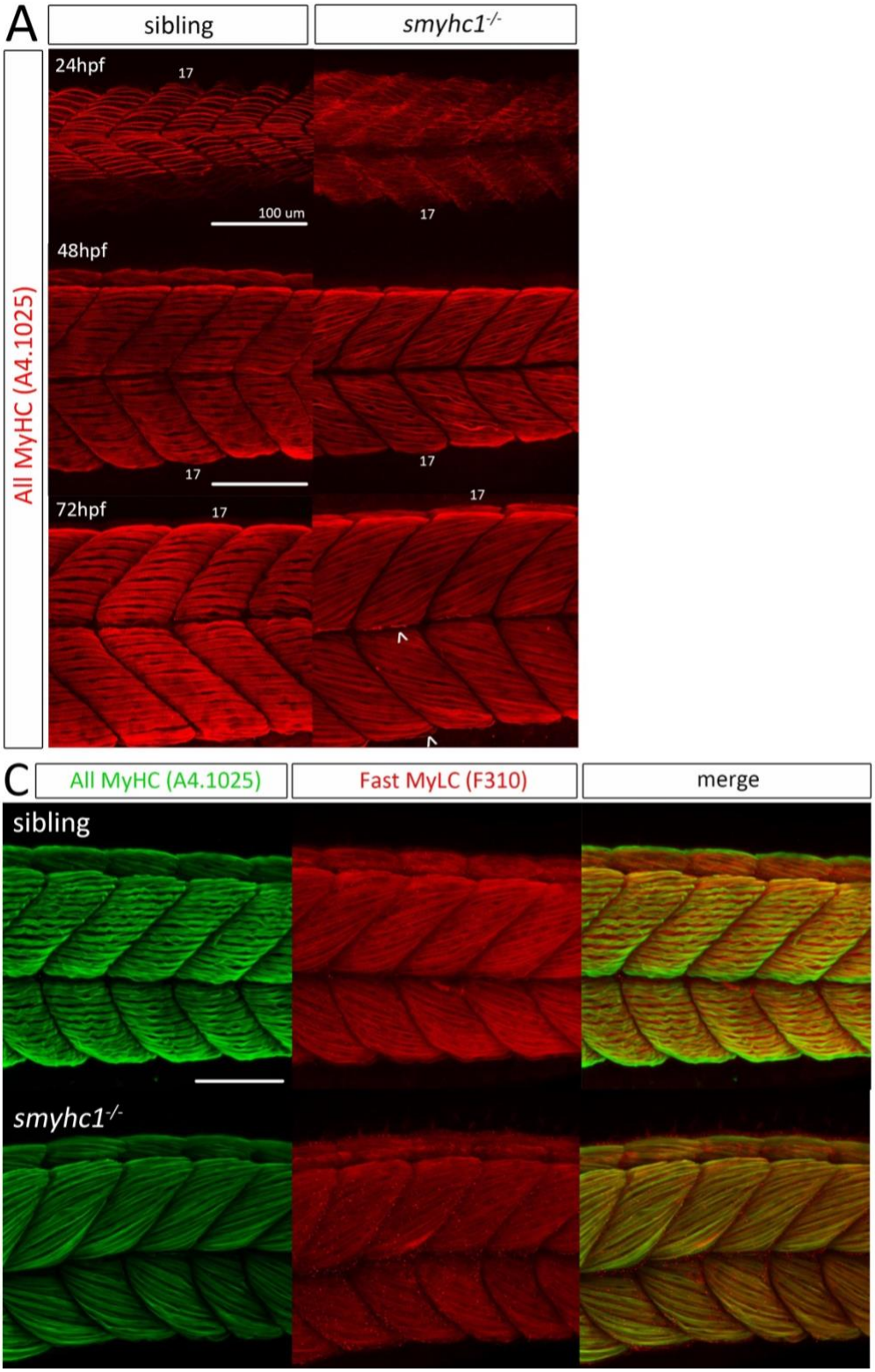


Figure 5.9. Lack of *smyhc1* does not affect fast muscle fibre morphology

A) Total MyHC immunofluorescence of 24, 48 and 72 hpf larvae from *smyhc1*$^{kg179+}$ in-crosses using A4.1025 antibodies. Wild type sibling on the left and *smyhc1*$^{kg179/kg179}$ mutants on the right. Slow fibres at horizontal myoseptum and somitic extremes (arrowheads) **B)** Double immunostaining using A4.1025 to label total MyHC and F310 to label fast specific MyLC. 72 hpf wild type larvae sibling at top and *smyhc1*$^{kg179/kg179}$ mutant below. Images of wild type and mutant centred on somite 17/18. Scale bars = 100 µm.

5.2.7. Slow fibres recover in adult *smyhc1* KO mutants

Since recovery of slow swimming occurred at 20 and 30 dpf, we examined whether slow fibres form in *smyhc1* KO mutants at adult stages (Fig 5.10). To assess whether loss of *smyhc1* had effect on slow muscle growth and muscle structure at adult stages, we performed immunostaining of slow muscle fibres on cross sections of *smyhc1*$^{-/-}$ adults. At adult stages, we used the A4.1025 antibody known to detect all MyHC in zebrafish muscle fibres and F59 and BA-D5 antibodies which are known to detect MyHC in zebrafish slow muscle fibres at 48 hpf (Crow and Stockdale, 1986; Schiaffino *et al.*, 1989; Devoto *et al.*, 1996). *Smyhc1*$^{-/-}$ mutants at 20 dpf, overall muscle appear normal from A4.1025 staining and slow fibres stained by F59 also appear normal, both staining in *smyhc1*$^{-/-}$ mutants were indistinguishable from wild type sibling (Fig 5.10). *Smyhc1*$^{-/-}$ mutants at 30 dpf, slow fibres stained by BA-D5 also appear normal and were indistinguishable from wild type sibling (Fig 5.10). However, phalloidin staining appear brighter in signal in the mutant compared to heterozygous sibling (Fig 5.10). Results show full slow fibre recovery in *smyhc1*$^{-/-}$ mutants at adult stages.

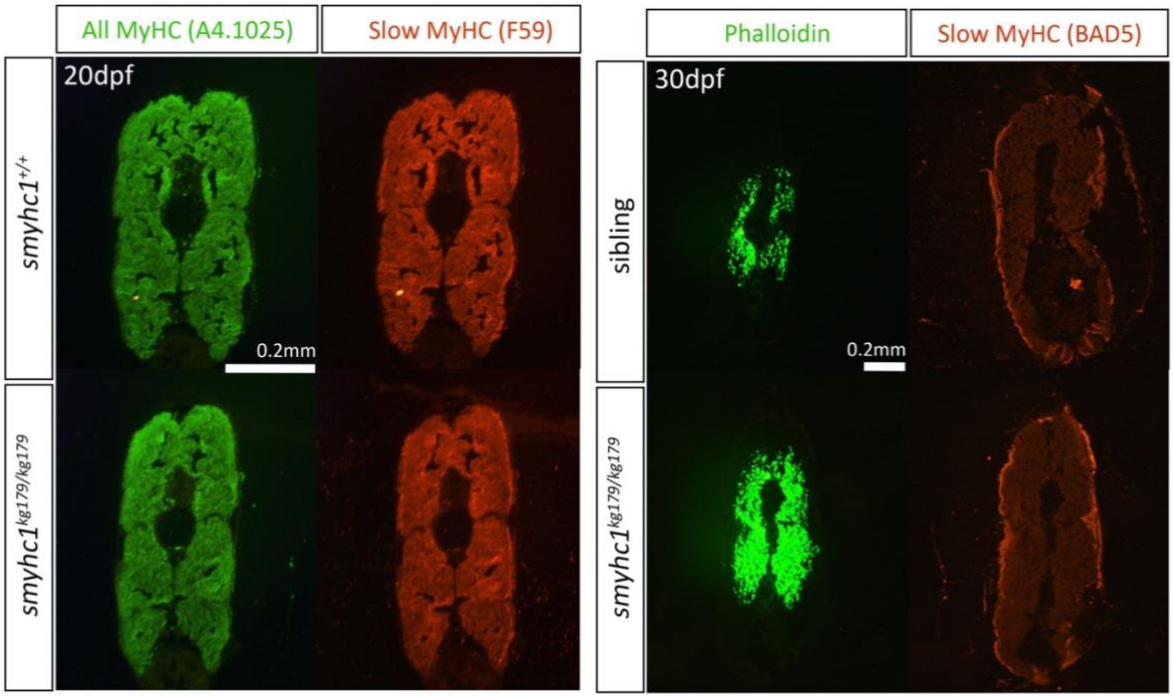


Figure 5.10. Recovery of slow fibres in adult zebrafish stages
Slow MyHC immunofluorescence 20 dpf and 30 dpf zebrafish from *smyhc1*$^{kg179+}$ in-crosses using F59 (at 20 hpf) and BAD5 (at 30dpf) antibodies. At 20 dpf, wild type sibling at top and *smyhc1*$^{kg179/kg179}$ mutant below. At 30 dpf, heterozygous sibling at top and *smyhc1*$^{kg179/kg179}$ mutant below. Scale bars = 200 µm

5.2.8. Large deletion mutations from two gRNAs targeting the smyhc locus

To model defects in later developmental stages, targeting *smyhc1* alone will not describe the role of slow MyHC in sarcomere assembly and function through to adulthood. Since of *smyhc1* is replaced by the expression of *smyhc2* and *smyhc3* in mature slow fibres (Stone Elworthy *et al.*, 2008; Li *et al.*, 2020), targeting the smyhc locus may describe the developmental function of slow MyHC in later stages of development to adulthood.

CRISPR/Cas9 genome editing was used to target *smyhc2* and *smyhc5* to ablate the smyhc locus except for *smyhc1*. To generate null *smyhc2-5* knockout mutants, one gRNA targeted *smyhc2* exon 5 and co-injected with either a second gRNA targeting *smyhc5* exon 1 or exon 36 to delete the majority of *smyhc2-5* locus (Fig 5.11A). There have been successful large deletions of 50 kb+ by using a similar strategy where co-injection of multiple gRNAs in zebrafish led to functionally null mutants (Hoshijima *et al.*, 2019; Kim and Zhang, 2020). The predicted size of PCR fragment between *smyhc2* exon 5 and *smyhc5* exon 1 is from 75.5kb in non-injected embryos to 309bp in injected, the number generated using deletion at the cut site with no additional INDEL mutations. Co-injection of gRNAs targeting *smyhc2* exon 5 and *smyhc5* exon 1 induced large deletion mutations in 1/16 of F0 founder embryos (Fig 5.11B). Sequencing of PCR fragment led to deletion between gRNA cut sites with an additional 14 bp insertion (Fig 5.12A). Mutation removes *smyhc3* and *smyhc4* and removes the majority of the *smyhc5* coding sequence, the start codon for *smyhc2* is removed and thus, predicted to ablate *smyhc2-5.*

The predicted size of PCR fragment between *smyhc2* exon 5 and *smyhc5* exon 36 is from 64.9kb in non-injected embryos to 390bp in injected, the number generated using deletion at the cut site with no additional INDEL mutations. Co-injection of gRNAs targeting *smyhc2* exon 5 and *smyhc5* exon 36 induced large deletion mutations in 2/16 of F0 founder embryos (Fig 5.11C). Sequencing of the first PCR fragment led to deletion between gRNA cut sites with an additional 1bp deletion (Fig 5.12B). The second PCR fragment led to deletion between gRNA cut sites with additional 14bp deletion and 5bp insertion (Fig 5.12B). Here, mutation removes *smyhc3* and *smyhc4* and the *smyhc5* coding sequence connects to *smyhc2,* predicted to generate a non-functional elongated and frameshifted *smyhc2/5* protein. Both mutations are predicted to remove the possibility of *smyhc2-5* function and future work to generate large deletion on *smyhc1* mutant background to study the role of *smyhc1-5* and the role of *smyhc2-5* in sarcomere assembly.

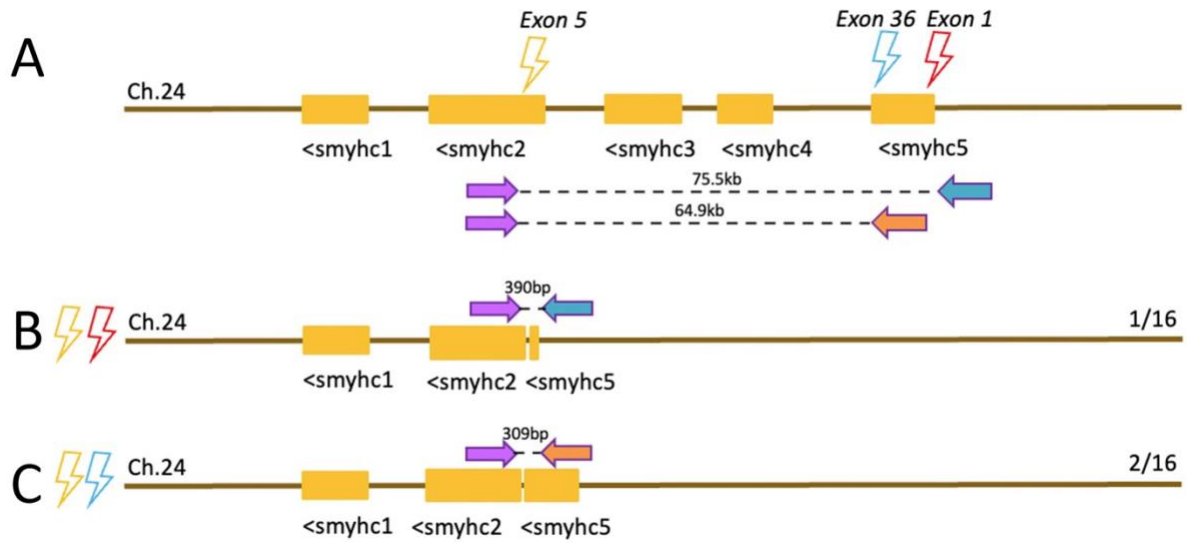


Figure 5.11. Genome editing targeting *smyhc* locus to delete *smyhc2-5*.
A) Schematic of *smyhc* locus on chromosome 24. CRISPR/Cas9 gRNA is designed to target *smyhc2* exon 5 (yellow bolt) and *smyhc5* exon 1 (red bolt) and *smyhc5* exon 36 (blue bolt). Sequencing primers in *smyhc2* exon 5 (purple arrow) and *smyhc5* exon 1 (blue arrow) and *smyhc5* exon 36 (orange arrow). **B)** Predicted deletion of smyhc locus when targeted at smyhc2 exon 5 and smyhc5 exon 36. The distance between sequencing primers *smyhc2* exon 5 (purple arrow) and *smyhc5* exon 1 (blue arrow) is 75.5kb and the distance in predicted *smyhc2-5* deletion is 390bp. 1/16 injected embryos show successful deletion of smyhc2-5 targeting *smyhc2* exon 5 and *smyhc5* exon36. **C)** Predicted deletion of smyhc locus when targeted at smyhc2 exon 5 and smyhc5 exon 1. The distance between sequencing primers *smyhc2* exon 5 (purple arrow) and *smyhc5* exon 36 (red arrow) is 64.9kb and the distance in predicted *smyhc2-5* deletion is 309bp. 2/16 injected embryos show successful deletion of smyhc2-5 targeting *smyhc2* exon 5 and *smyhc5* exon1.

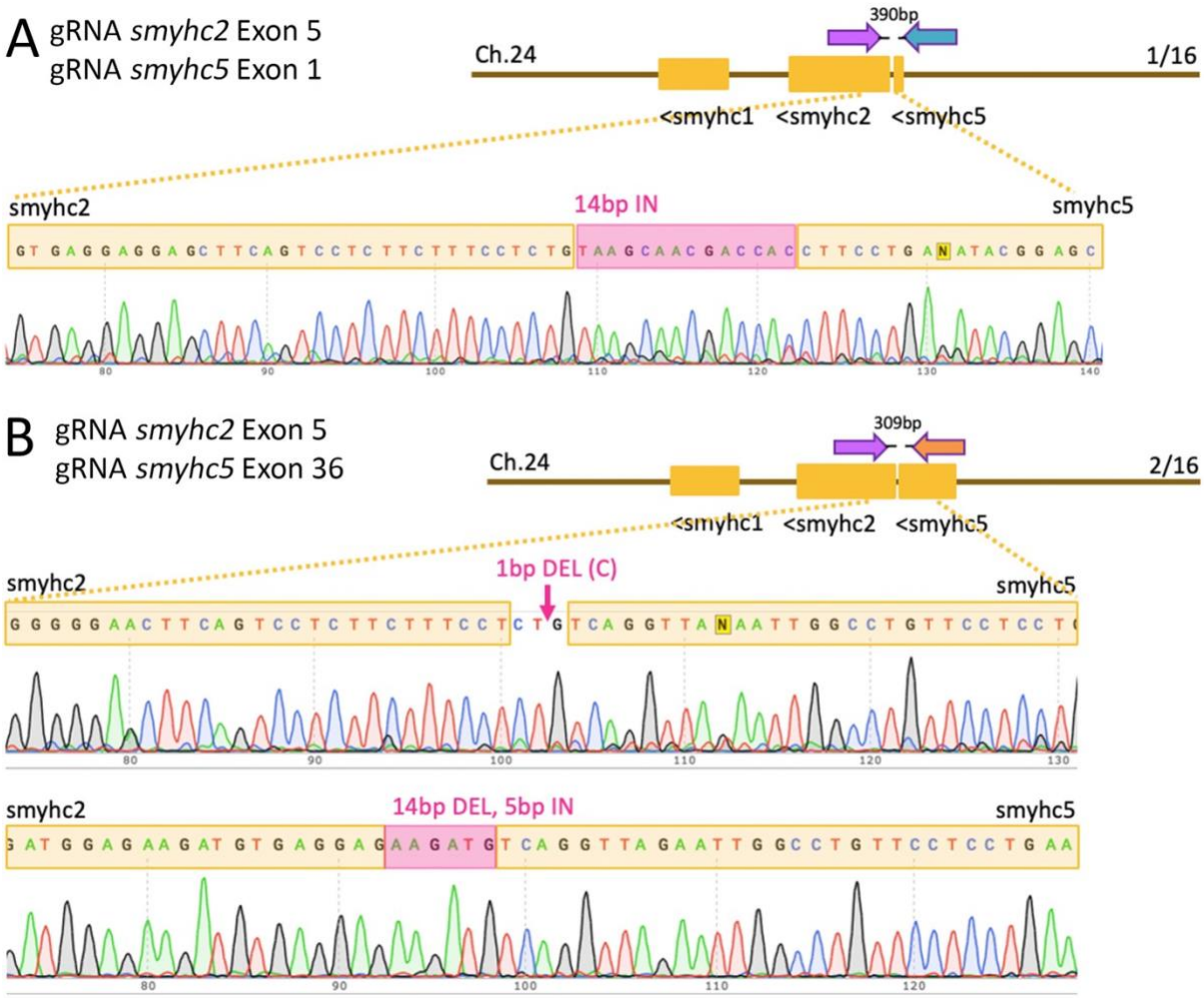


Figure 5.12. Sequencing analysis of *smyhc2-5* deletion.
A) 1/16 injected embryos show successful deletion of *smyhc2-5* targeting *smyhc2* exon 5 and *smyhc5* exon36. The DNA sequence of mutations identified in F0 injected embryos. Mutation led to predicted cut sites at the PAM+3 sequence in *smyhc2* and *smyhc5* with additional 14bp insertion. **B)** 2/16 injected embryos show successful deletion of smyhc2-5 targeting *smyhc2* exon 5 and *smyhc5* exon1. DNA sequence of mutations identified in F0 injected embryos. The sequence presented from 5' to 3' and predicted outcome described above sequence. One mutation led to predicted cut sites at the PAM+3 sequence in *smyhc2* and *smyhc5* with additional 1bp deletion. The second mutation led to predicted cut sites at the PAM+3 sequence in *smyhc2* and *smyhc5* with additional 14bp deletion and 5bp insertion.

5.3. Discussion

Our findings make five important points regarding the role of *smyhc1* and MyHC heterogeneity during skeletal muscle development. First, knockout of smyhc1 using CRISPR/Cas9 technology generated *smyhc1*^{kg179} and *smyhc1*^{kg180} mutants resulted in immotility at the early stages of development. Second, despite immotility at early stages, *smyhc1* mutants show no morphological defects under light microscope observation and are viable and fertile when developing to adulthood. Third, movement defects persist in the early stages but not in adults. Fourth, loss of *smyhc1* abolishes slow fibres and leads to disorganised sarcomere assembly in the early stages of development. Finally, loss of *smyhc1* does not affect fast fibre development.

5.3.1 *Smyhc1* mutants are functionally null with no off-target effects

Loss of Smyhc1 from KO mutants and morphants were shown to have paralysis up to 48 hpf (Codina *et al.*, 2010; Xu *et al.*, 2012; Li *et al.*, 2020; Whittle *et al.*, 2020). Our findings with *smyhc1*^{kg179} and *smyhc1*^{kg180} confirm this conclusion as homozygous *smyhc1*^{kg179} and *smyhc1*^{kg180} mutants were immotile at 24 hpf. However, at adult stages, *smyhc1* KO mutants presented spinal curve defects, reduced food intake and larval lethality (Li *et al.*, 2020; Whittle *et al.*, 2020). Our findings do not show morphological defects or lethality in larval to adult stages.

Several mutants in the *smyhc1* locus have now been reported, not all of which may be genetically simple. Current *smyhc1* KO studies using CRISPR/Cas9 and TALENS have been used to target exon 5 (Li *et al.*, 2020) and exon 16 (Whittle *et al.*, 2020) of *smyhc1* locus, respectively (Fig 5.13). Both *smyhc1* KO alleles in the literature introduced frameshift mutations leading to early stop codons and showed strong *smhyc1* mRNA NMD in mutants. However, the gRNA used by Li et al, 2020 targeted the 3' end of exon 4 of the *smyhc1* locus but also shows 100% identity to *smyhc2-5* sequence (Fig 5.13, Appendix 5.3). As the sequence of these linked genes was not analysed in the chromosome carrying the *smyhc1* mutant allele, linked off-target mutations in *smyhc2-5* cannot be ruled out, particularly as we observed very efficient mutagenesis in this locus (Fig 5.11). Smyhc2 and smyhc3 are predominantly expressed in the head and jaw (S. Elworthy *et al.*, 2008) which such off targets may hinder the ability for zebrafish larvae produce jaw movement and thus, unable to eat. In our mutants, we have sequenced *smyhc2* and *smyhc3* in our *smyhc1* mutants to show that there are no off targets in these genes. Furthermore, the TALENS employed by Whittle et al, 2020, which caused a frameshift mutation leading to an early stop codon in exon 16 may also result in off-target effects as exon 16 show 87-89% sequence similarity

to *myh7* and *smyhc2*. Moreover, an early stop codon in exon 16 may lead to a partially functional truncated protein as the predicted truncated protein contain ATP binding, actin binding, switch 1, switch 2 and loop domains (Appendix. 5.4). MyHC genes show high levels of sequence identity and screening for unique sequences to *smyhc1* is crucial to avoid possible off target effects in other MyHC genes. Our CRISPR/Cas9 mutations were optimised to target exclusively to *smyhc1* and BLAST analysis of both our gRNAs target only to *smyhc1* (Appendix 5.3). Since our mutants did not show larval lethality, spinal curve defects and reduced food intake, may suggest that these are not specific to *smyhc1* function. While these unfortunate issues raise questions about previous work, our finding of a similar phenotype in kg179 and kg180 suggest the previous reports are indeed primarily analysing the effect of loss of Smyhc1 function.

Our INDEL mutations generated using CRISPR/Cas9 targeted the N-terminal region of *smyhc1* and thus lead to loss of majority of functional protein domains. Both *smyhc1*^{kg179} and *smyhc1*^{kg180} mutants show strong nonsense-mediated mRNA decay where ribosomes failed to fully translate mRNA and thus, little truncated (and no full-length) protein is predicted to be produced. There is no known alternative splicing or promotor usage observed and phenotype observed at early stages of development suggest mutant proteins. If spliced variants were produced, they are predicted to be defective and thus, do not incorporate into functional thick filaments. At 72 hpf, mRNAs expressed from *smyhc2* and *smyhc3* are localised in specific subsets of muscles (Stone Elworthy *et al.*, 2008). Our *smyhc1*^{kg179} mutants show slow MyHC immunoreactivity in the same subsets of slow fibres, suggesting that there were no off-target effects in *smyhc2* and *smyhc3.* In adult zebrafish, *smyhc1* expression is replaced by *smyhc2* and *smyhc3* (Stone Elworthy *et al.*, 2008). Consistent with this, our finding that *smyhc1*^{kg179} mutants recover their swimming motility at 20-30 dpf a stage at which *smyhc2* and *smyhc3* become the predominantly expressed slow MyHC genes in slow skeletal muscle (Fig 5.10). Thus, our *smyhc1*^{kg179} and *smyhc1*^{kg180} mutants are likely functionally null with no-off-target effects in *smyhc2*, *smyhc3* or other genes encoding slow MyHCs.

```
                                                  Partial AMO (Xu et al, 2012)
smyhc1      ---------------------------------------------ATGGGTGACGCCGTT    15
smyhc2      ---------------------------------------------ATGGGGGATGCTGTG    15
smyhc3      ------------------------------------------AAGATGGGGGATGCTGTG    18
smyhc4      ---------------------------------------------ATGGGGGATGCTGTG    15
smyhc5      ---------------------------------------------ATGGGGGATGCTCTG    15
myh7        GGTTCTTCTGCCTC-CGCACTTGGTGCACATCAGACAAGGCAATCATGGGGGACGCTCAG    82
myh7l       AAACCTGGAGCTTCCTTCTGCTGTGATTAATCGCTTTGGTTGACAATGGGCGATGCTGAA    120
                                                      ***** ** **

smyhc1      ATGGCAGAGTTTGGGTCTGCTGCTCCCTTCCTGCGCAAGTCTGACAAGGAGCGTCTGGAG    75
smyhc2      ATGGCAGAGTTTGGGCCTGCGGCTCCGTTCTTACGTAAATCAGATAAGGAGCGTCTGGAG    75
smyhc3      ATGGCGGAGTTTGGAGCTGCGGCTCCGTACCTCAGGAAGTCGGACAGGGAGCGTCTGGAG    78
smyhc4      ATGGCGGAGTTTGGAGCTGCGGCTCCGTACCTTAGGAAATCAGACAAGGAGCGTCTGGAG    75
smyhc5      ATGGAGGAGTTTGGAGCTGCGGCTCCGTATCTCAGGAAGTCGGACCGGGAGCGTCTGGAG    75
myh7        ATGGCAGAGTTTGGAGCAGCAGCTTCTTACCTGCGAAAGTCAGATCGAGAGCGTCTGGAA    142
myh7l       ATGTCTGTTTTTGGGGCCGCAGCGCCTTACCTGCGGAAGTCTGAAAAGGAGCGTCTTGAG    180
            ***   *  *****  * ** **  * *   *  * ** ** **    ******** **
                gRNAKO1
smyhc1      GCCCAAACTCGTATTTTTGACATGAAGAAGGAGTGCTTTGTGCCTGACCCTGAGGTTGAG    135
smyhc2      GCCCAAACTCGTCCTTTTGACATGAAGAAGGAGTGTTTCGTGCCTGATCCCGAGGTTGAG    135
smyhc3      GCCCAAACTCGCCCCTTTGACATGAAGAAAGAGTGTTTTGTTCCTGATGCTGACGAGGAG    138
smyhc4      GCCCAAACTCGCCCCTTTGACATGAAGAAAGAGTGTTTTGTTCCTGATGCTGACGAGGAG    135
smyhc5      GCCCAAACTCGCCCCTTTGACATGAAAAAGGAATGTTTCGTCCCGGATACTGATGAAGAG    135
myh7        GCACAAACCCGTCCCTTTGATATGAAAAAGGAGTGTTTTGTGCCTGATCCAGATGAAGAG    202
myh7l       GCGCAGACGAAAGCCTTTGACTTAAAGAAGGAATGCTTTGTGCCGGATGCAATAGAGGAG    240
            ** ** **       *****  * ** ** ** ** ** ** ** **  *    *  ***
```

__ 300 bp__

```
                            gRNAKO2
smyhc1      AACCCATACAAGTGGCTGCCAGTGTACGATTCCTCTGTGGTCAAAGCCTACAGAGGCAAG    435
smyhc2      AACCCATACAAGTGGCTGCCAGTGTACAATCAGGAGGTGGTCGTTGCTTACAGAGGAAAG    435
smyhc3      AACCCCTACAAGTGGCTGCCAGTGTACAATCAGGAGGTGGTCGTTGCTTACAGAGGAAAG    438
smyhc4      AACCCCTACAAGTGGCTGCCAGTGTACAATCAGGAGGTGGTCGTTGCTTACAGAGGAAAG    435
smyhc5      AACCCCTACAAGTGGCTGCCAGTGTACAATCAGGAGGTGGTTCTGGCTTACAGAGGAAAG    435
myh7        AACCCCTACAAGTGGCTGCCGGTGTACAATCAGGAGGTGGTTGTAGCCTATAGAGGGAAA    502
myh7l       AACCCCTACAAGTGGCTGCCGGTGTACAATCAGGAGGTTGTTATAGCCTATAGAGGGAAA    540
            ***** ************** ****** **      ** **    ** ** ***** **
```

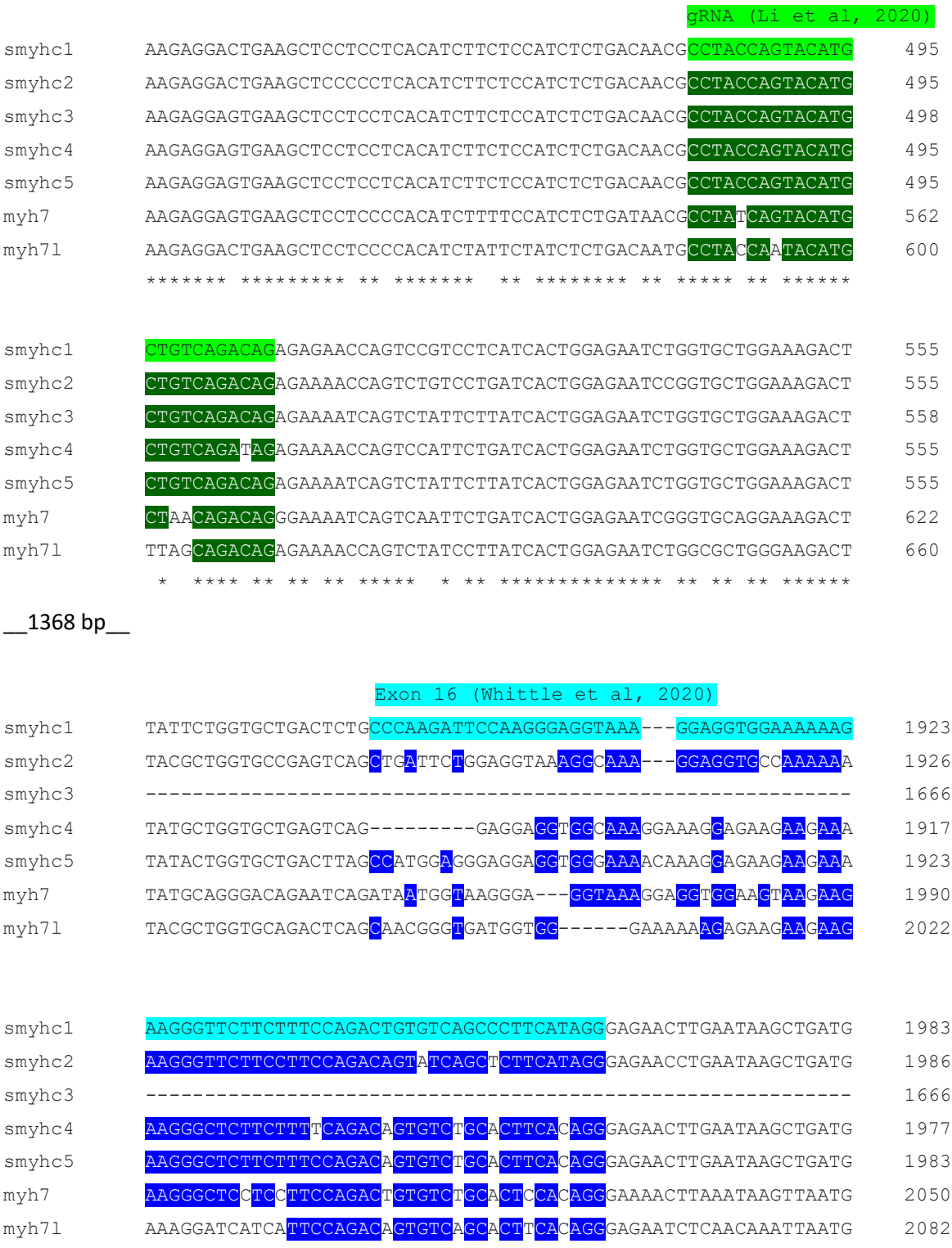


Figure 5.13. Aligned sequencing segments highlighting gRNA used for *smyhc1* KO showing potential off target effects in other *smyhc* and *myh7/myh7l* genes.
Each segment sequenced and subsequently aligned using CLUSTALO. Yellow highlight indicates gRNA used in current thesis. Green highlight indicates morpholino and gRNA used in Xu et al, 2012 and Li et al, 2020, respectively. Blue highlight indicates whole of exon 16 where Whittle et al, 2020 targeted for TALENS genome editing, specific cut location is unknown.

5.3.2. Role of *smyhc1* in sarcomere assembly in early slow fibres

Smyhc1 play a key role in sarcomere assembly during the early stages of development as it is the first MyHC to be expressed in zebrafish slow fibres (Devoto *et al.*, 1996). Studies have been made to model and describe the sequence of events in sarcomere assembly (Rhee, Sanger and Sanger, 1994; Holtzer *et al.*, 1997; Ehler *et al.*, 1999; Gregorio *et al.*, 1999; Rui, Bai and Perrimon, 2010; Fenix *et al.*, 2018). Our findings show that *smyhc1* is essential for thick filament assembly and myofibril organisation in slow fibres during the early stages of development and are consistent with findings in which lack of Smyhc1 lead to defective sarcomere assembly during early stages of development (Li *et al.*, 2020; Whittle *et al.*, 2020).

Myofibrils are first anchored and assembled at the cell periphery, close to the membrane at the MTJ (Kelly and Zacks, 1969; Tokuyasu, 1989). There are integrin adhesion sites known as costameres which connect thin filaments to the MTJ (Pardo, Siliciano and Craig, 1983; Ervasti, 2003; Quach and Rando, 2006). Integrins, α-actinin, vinculin and talin are present at the early stages of the costameres and are suggested to be the site for α-actinin to accumulate at the MTJ (Fujita, Nedachi and Kanzaki, 2007; Du, Sanger and Sanger, 2008). Z-disks are formed initially as aggregates called z-bodies at the myotendinous junction (MTJ) (Tokuyasu and Maher, 1987). Mice and Drosophila have shown ligands at the extracellular matrix are essential for z-disk formation as the first step in sarcomere assembly (Volk, Fessler and Fessler, 1990; Bloor and Brown, 1998). In drosophila, integrins link to a zasp protein for recruitment of α-actinin for z-disk assembly (Au *et al.*, 2004). Our *smyhc1*^{kg179} mutants show accumulation of α-actinin and actin filaments at the somite border supporting a model whereby actin and α-actinin anchor at the MTJ prior to integrating myosin filaments in building and elongating the muscle fibre.

After initiation of sarcomere at the MTJ, elongation of sarcomere can start to build. One model of sarcomere assembly describes the independent assembly of I-Z-I bodies before the integration of myosin, this is known as the "stitching model" of sarcomere assembly (Rhee, Sanger and Sanger, 1994; Holtzer *et al.*, 1997; Van Der Ven *et al.*, 1999; Sanger *et al.*, 2005). A second model describes the formation of stress fibre-like structures utilising non-muscle myosin as a template for sarcomere proteins to assemble, forming a pre-myofibril and are then later replaced with muscle myosin to form mature myofibrils (Rhee, Sanger and Sanger, 1994). A third model describes the role of titin recruited by α-actinin to bind to the Z-disk region and the M-line to act as a template to regulate the alternating patterning of I-Z-I bodies and myosin filaments (Kelly and Zacks, 1969; Tokuyasu and Maher, 1987;

Schwander *et al.*, 2003; Au *et al.*, 2004). Our *smyhc1*^{kg179} mutants show some assembly of actin filaments and clustering of α-actinin at somite borders to initiate sarcomere assembly. However, in the absence of *smyhc1*, the elongation step of the myofibril is absent and thus, myofibrils do not elongate supporting the studies showing integration of myosin as one of the last steps in myofibrillogenesis.

On a cellular level, our staining does not show the weather slow fibres are non-existent in our *smyhc1*^{kg179} or whether the slow fibres remain present but rather the lack of myosin molecules with defective sarcomere formation. One test to confirm whether slow fibres exist in our *smyhc1*^{kg179} mutants is to cross our *smyhc1*^{kg179} mutants to a transgenic line Tg(Ola.Actb:Hsa.HRAS-EGFP)vu119 (ß-actin:GFP) which is a construct containing the β-actin promoter and a membrane targeted EGFP. This line is used to visualise membranes in live larvae (Cooper et al., 2005). Whether slow fibres have resulted in apoptosis, methods such as labelling activated Caspase-3 as one of the signalling molecules involved in cell apoptosis (Sorrells *et al.*, 2013). The next test is to identify the defects in sarcomere assembly at many more timepoints for actin filament formation when there is a lack of *smyhc1*. To test how actin filaments form in our *smyhc1*^{kg179} mutants, I would cross my *smyhc1*^{kg179} mutants to *Tg(acta1:lifeact-GFP;acta1:mCherryCAAX) which* Lifeact-GFP binds to thin filaments through the Lifeact tag. mCherry is directed by the CAAX tag to the sarcolemma (Berger, Hall and Currie, 2015).

5.3.3. Lack of *smyhc1* does not affect sarcomere organisation in adulthood

Despite defects in slow fibre sarcomere assembly in *smyhc1*^{kg179} mutants during early development, there were no defects in the migration of slow muscle precursor into the superficial layer of elongated slow fibres where juvenile to adult *smyhc1*^{kg179} mutants does not show an obvious phenotype. During early slow fibre development, *smyhc1* is predominantly expressed and *smyhc2* and *smyhc3* are expressed in a subset of muscles (Stone Elworthy *et al.*, 2008). In the adult stages, the predominant expression of *smyhc1* diminishes at 42 dpf and is replaced by *smyhc2* and *smyhc3* in mature slow fibres (Stone Elworthy *et al.*, 2008; Li *et al.*, 2020). Our data showing immotility from *smyhc1*^{kg179} mutants at early stages of development followed by the recovery of phenotype in juvenile and adults correlate to the predominant expression of *smyhc1* in young larvae transitioning to *smyhc2* and *smyhc3* in juvenile to adulthood.

During early mouse and human slow fibre development, MyHC-slow and predominantly MyHC-Emb (*MYH3*) are expressed in primary fibres. Lack of MyHC-Emb increases the fast myofiber number and

slow myofiber area (Sharma *et al.*, 2018). Despite early embryonic expression of *smyhc1* in zebrafish slow fibres, our findings suggest that *smyhc1* does not functionally resemble mammalian *MYH3*. *Smyhc1*^{kg179} mutants show defective primary slow fibres but continued survival to develop into secondary fibres in juveniles and adults with no observable difference in fast fibre number or increased slow fibre area. Our *smyhc1*^{kg179} mutants do not show secondary fibre defects as zebrafish *smyhc1* is not homologous to mammalian *MYH3* but rather to mammalian *MYH7*. Zebrafish *smyhc2* and *smyhc3* are also homologous to mammalian *MYH7* and knockout of these genes may describe the juvenile to adult phenotype associated with mutations in human *MYH7*. Currently, there are no knockout studies have been made on these genes. Thus, defective sarcomere organisation in slow fibres from lack of *smyhc1* at the early stages of development is not essential for sarcomere organisation in secondary fibres in juveniles to adulthood. Knockout of the *smyhc* locus will be crucial to identify the role of *MYH7* from early development to adulthood and identify the role of *smyhc2-5* in sarcomere assembly and organisation.

5.3.4. Conclusion

Smyhc1 has been demonstrated to play a role in sarcomere organisation during the early stages of development (Codina *et al.*, 2010; Li *et al.*, 2020; Whittle *et al.*, 2020). Present data show lack of *smyhc1* results in slow muscle immotility during the early stages of development with recovery at 30 dpf. *Smyhc1*^{kg179} mutants show defective sarcomere organisation at the early stages of development and give insight into the role of *smyhc1* in sarcomere assembly after the initiation step whereby Z-disks anchor to the MTJ and subsequently elongate to form the mature myofiber. Phenotypic data from smyhc1 mutants give insight into the early developmental role of mammalian MYH7 however, the subsequent transitional role of *MYH7* from juvenile to adulthood remains in question. Ongoing work to generate a large deletion of the smyhc locus will give insight into the role of *MYH7* orthologs *smyhc1-5* in zebrafish for sarcomere assembly.

Chapter 6

General Discussion

6.1. Summary

The two congenital myopathies that I have focused on in the present work, Laing Distal Myopathy (LDM) and Myosin Storage Myopathy (MSM) are due to sarcomeric gene mutations in *MYH7* (Lamont *et al.*, 2014; Parker and Peckham, 2020). Although there are currently no curative medicines for *MYH7*-related congenital myopathies and available treatments simply target the various symptoms (*Myosin storage myopathy*, 2016; Topaloglu, 2020). The aim of this thesis was to study the potential underlying molecular and cellular mechanisms leading to LDM and MSM by identifying primary biophysical defects in human fibres obtained from affected patients and developing zebrafish models that investigated developmental defects in the quest for treatment design for *MYH7*-related diseases. My main findings were the following: 1) There is no overall alteration in sarcomere organisation in the presence of defective myosin molecules. 2) Mutations affecting the *MYH7* MyBP-C binding domain destabilise the SRX state. 3) Zebrafish genes *smyhc1-5, myh7* and *myh7l* are orthologous to mammalian *MYH7*. 4) Loss of *smyhc1* in zebrafish results in defective sarcomere organisation at the early stages of development, indicating the role of *smyhc1* in sarcomere assembly to elongate the myofiber. 5) Transitional role of *MYH7* from early developmental stages to adulthood remains in question. Overall, current data give early insight into the mechanism for the role of slow myosin in sarcomere assembly.

6.2. Defective slow MyHC does not affect sarcomere organisation in adults.

Myosin molecules are formed through the dimerization of individual myosin units and stabilised through their coiled-coil structure in the light meromyosin (LMM) (McLachlan and Karn, 1982). The heptad repeats *a-g* in the coiled-coil describe the functional purpose for myosin dimerization as described in my introduction. The common amino acids affected in LDM and MSM patients were in amino acids *a* and *d* for the main core of myosin dimerization through the characteristic hydrophobic and in the charged amino acids on the exterior portion of the myosin LMM at positions *b, c* and *f* (McLachlan and Karn, 1982). The charged amino acids within the heptad sequence form a larger 28 amino acid repeat to enable myosin dimers to form larger myosin filaments in sarcomeres (Squire, 1973; Atkinson and Stewart, 1992; Rahmani *et al.*, 2021). Such mutations in *MYH7* affecting either myosin dimerization or myosin filament assembly were unable to distinguish between LDM and MSM.

The LMM structure has been described as an intricate coiled-coil structure and its structure is conserved between vertebrates and invertebrates emphasising the importance of the amino acid arrangement for myosin molecules packing together (Squire, 1973; Rahmani *et al.*, 2021). Despite such conserved intricate structure of the coiled-coil LMM, we demonstrated that in the presence of defective myosin molecules, there was no hindrance for defective myosin molecules to dimerise and pack into thick filaments and slow myosin is not essential for sarcomere organisation. Actin filaments have been shown to form independently of myosin (Lin *et al.*, 1994) and initial steps in sarcomere assembly involve the formation of premyofibrils containing non-muscle myosin II (Rhee, Sanger and Sanger, 1994; Swailes *et al.*, 2006). Muscle myosin, in this case, slow myosin replaces non-muscle myosin as one of the last steps in myofibril formation (Komiyama, Maruyama and Shimada, 1990; Péault *et al.*, 2007) and argues that slow myosin is not essential for sarcomere organisation in myofibrils. Due to the subtle nature of dominant mutations in MSM and LDM patients, missense or single amino acid deletions in defective myosin molecules is intermixed with healthy myosin molecules with a high level of variability. Variability of healthy and defective myosin may have led to full thick filament formation with differences in length that are too subtle to detect through our methods with fluorescence microscopy. Further analysis using more sensitive techniques such as super-resolution microscopy and electron microscopy may be able to detect such small length changes in the thick filament.

Although there was no change in myosin filament length and organisation coupled with the presence of organised actin filaments, our results describing the quality of myosin packing were affected through observations of a change in myosin head positioning. Mutations affecting the LMM at the myomesin or MyBP-C binding site destabilise myosin in the super relaxed (SRX) state. The role of myomesin in the M-band is to regulate and stabilise the packing of myosin filaments into a hexagonal myosin filament lattice (Agarkova *et al.*, 2003; Hu, Ackermann and Kontrogianni-Konstantopoulos, 2015). Loss of myomesin-1 in human cell lines has shown sarcomere disassembly (Hang *et al.*, 2021). Our data show myofibres with organised myosin and actin filaments interlaced in regular intervals and argue that mutation at the myomesin binding site is dispensable for sarcomere organisation. Overall, I argue that LDM and MSM mutations affecting myosin do not affect sarcomere organisation and without a clear analysis of subtle changes, defects in thick filament assembly in the presence of defective myosin molecules remain in question.

6.3. Destabilised SRX state may trigger hypercontractility

The role of MyBP-C in stabilising myosin molecules in the SRX state has been highly studied. Myosin in the conventional "J" motif resembles myosin in a relaxed state whereby myosin heads interact with each other to form an "interacting heads motif" (IHM) and both myosin heads interact with the S2 region (Alamo *et al.*, 2017; Woodhead and Craig, 2020). The absence of MyBP-C has resulted in a shift in the proportion of myosin molecules in the SRX state to predominantly in the disordered (DRX) state (Luther *et al.*, 2008; Zoghbi *et al.*, 2008; McNamara *et al.*, 2016). The MyBP-C and myomesin binding sites overlap in the LMM and since mutations in this region did not lead to sarcomere disassembly, there is a higher possibility that LDM and MSM mutations affect the ability of MyBP-C to bind to myosin molecules at the LMM site. MyBP-C connects to myosin at two sites, the N-terminal MyBP-C domain connects to the myosin head region and C-terminal MyBP-C connects to myosin LMM (Luther *et al.*, 2008; Spudich, 2015). There has been a link between hypertrophic cardiomyopathy (HCM) mutations in *MYH7* and the destabilising effects of myosin in the SRX state and thus lead to hypercontractility in the heart (Alamo *et al.*, 2017; Toepfer *et al.*, 2020). Studies have also shown that in the absence of MyBP-C or the presence of defective MyBP-C molecules in HCM patients, there were also destabilising effects on SRX myosin head positioning as seen in the presence of mutations affecting the MyBP-C binding site on *MYH7* (McNamara *et al.*, 2016; Christopher N Toepfer *et al.*, 2019; Christopher N. Toepfer *et al.*, 2019). We also observed a shift in the proportion of myosin heads in DRX state in fibres in muscle fibres obtained from patients with mutations at the MyBP-C binding domain in the LMM of *MYH7.* Here I argue that the C-terminal MyBP-C binding domain also shows the same destabilising effect of the SRX state as mutations affecting the N-terminal MyBP-C domain and MyBP-C. Current data suggest that the role of both MyBP-C sites and MyBP-C itself is to stabilise the SRX state through the interaction with slow myosin. Mutations affecting this interaction at the MyBP-C site in the head region and MyBP-C itself have led to hypercontractile muscle in HCM patients. Skeletal muscle from LDM and MSM patients may have also shown hypercontractility and show poor ability to relax.

The main clinical phenotype in HCM patients is hypertrophy of the heart ventricle, hypercontractility and myocardial fibrosis. In mouse HCM models, a small molecule drug Mavacamten has been shown to suppress and reverse the symptoms of hypertrophy of the heart ventricle, cardiomyocyte disarray and myocardial fibrosis (Green *et al.*, 2016). Mavacamten have also been shown to reverse the destabilising effects on the IHM of myosin and restored the balance of myosin molecules in their SRX and DRX state in HCM cell lines (Toepfer *et al.*, 2020). Subsequently, cell sizes were reduced and

restored to their original size as wild type cell lines (Toepfer *et al.*, 2020) and are currently used as an effective treatment for HCM patients with mutations in the *MYH7* MyBP-C site and *MyBP-C* (Hegde *et al.*, 2021). Our findings describe the destabilising effects on slow myosin SRX state in LDM patients and may show similar pathological defects in slow skeletal muscle as shown in cardiac muscle from HCM patients. Treatment with Mavacamten may be a candidate treatment for LDM patients with mutations in the MyBP-C binding site.

6.4. Zebrafish *smyhc1* orthologous to human *MYH7*

Early developmental defects are unknown in patients affected by LDM and MSM mutations as clinical phenotypes have only been analysed in adults and children. Zebrafish disease models might prove advantageous as early time points can be studied. Early developing zebrafish larvae are clear, quick development to adulthood and breeding of adult fish gives large clutch sizes. It is essential to identify the zebrafish equivalent of the human *MYH7* gene that would be most likely to give a phenotype in the defined functional muscle.

There have been studies describing zebrafish *smyhc1-5, myh7, myh7l* and *myh6* genes as orthologs to human *MYH6/7* (McGuigan, Phillips and Postlethwait, 2004) but no current data to distinguish orthology between zebrafish myh genes to either *MYH6* or *MYH7* (Liew *et al.*, 1990; Epp *et al.*, 1993; McGuigan, Phillips and Postlethwait, 2004). Mammalian *MYH6* and *MYH7* are located next to each other on the same chromosome and exist from a duplication event (Yamauchi-Takihara *et al.*, 1989; Gulick *et al.*, 1991). In chapter 4.2.3. I identified 4 signature amino acids to distinguish between *MYH6/7* in mammals and ray-finned fish (Fig 4.7). Additionally, in our analysis of gene synteny between humans to zebrafish, zebrafish *smyhc1-5, myh7* and *myh7l* were syntenic to human *MYH7* and zebrafish *myh6* was syntenic to human *MYH6*. The presence of many slow MyHC orthologs to the *MYH7* in mammals (McGuigan, Phillips and Postlethwait, 2004; Watabe and Ikeda, 2006; Ikeda *et al.*, 2007) arose from a teleost genome duplication event (Amores *et al.*, 1998; Meyer and Schartl, 1999; Taylor *et al.*, 2001). Evidence of this duplication event can be seen in zebrafish *smyhc1-5* are syntenic to *myh7* and *myh7l* (Fig 4.8) which have also been observed in goldfish and platyfish (Fig 4.8). Expression patterns between human *MYH7* and zebrafish orthologs *smyhc1-3* and *myh7* demonstrate similarity whereby *MYH7* is expressed both in slow skeletal muscle and in the heart ventricle and Zebrafish have separate myh orthologs expressing *smyhc1-3* only in slow skeletal muscle (Stone Elworthy *et al.*, 2008) and *myh7* expressed in the heart ventricle (Park *et al.*, 2009). Human *MYH6* is predominantly expressed in the heart atrium and compliments the expression pattern of zebrafish

myh6 in the heart atrium (Huang *et al.*, 2005). Thus, I argue that the common ancestor of humans and zebrafish had a pre-existing *MYH6* and *MYH7* gene and show zebrafish are orthologous to human *MYH7* and not to human *MYH6*.

As there is no single ortholog to human *MYH7*, but rather a cluster of orthologous zebrafish genes, the segmented expression pattern of *smyhc1-5, myh7* and *myh7l* prove advantageous in studying developmental defects exclusively associated with slow skeletal muscle. Since zebrafish *smyhc1-3* are only expressed in slow skeletal muscle (Stone Elworthy *et al.*, 2008) the possibility of generating viable mutants is higher as myosin affecting the cardiac muscle is not compromised. *Smyhc1* show broad localisation of expression in the slow skeletal muscle and *smyhc2* and *smyhc3* is expressed in a subset of muscles during the early stages of development. There are knockdown and knockout studies on *smyhc1* revealing early developmental defects in slow muscle giving confidence in generating knockout mutants to study developmental defects associated with mutations in *smyhc1*, the zebrafish ortholog to *MYH7* (Codina *et al.*, 2010; Xu *et al.*, 2012; Li *et al.*, 2020; Whittle *et al.*, 2020).

6.5. *Smyhc1* functions exclusively in early muscle development

Smyhc1 has been demonstrated to play a role in sarcomere organisation during the early stages of development (Codina *et al.*, 2010; Li *et al.*, 2020; Whittle *et al.*, 2020). Consistent with *smyhc1* knockdown and knockout studies, present data show lack of *smyhc1* results in slow muscle immotility during the early stages of development (Codina *et al.*, 2010; Li *et al.*, 2020; Whittle *et al.*, 2020) with full recovery of slow muscle motility at 30 dpf. However, conflicting data is describing *smyhc1* knockout (KO) mutant adults with spinal curve defects, reduced food intake and larval lethality (Li *et al.*, 2020; Whittle *et al.*, 2020). Our findings do not show such morphological defects or lethality in larval to adult stages. Since zebrafish myosin paralogs share a high degree of sequence identity, there are high risks of off-target effects when targeting using CRISPR/Cas9 or TALENS genome editing. Current *smyhc1* KO studies using CRISPR/Cas9 and TALENS have been used to target exon 5 (Li *et al.*, 2020) and exon 16 (Whittle *et al.*, 2020) of the *smyhc1* locus. Both *smyhc1* KO alleles in the literature introduced frameshift mutations leading to early stop codons and showed strong NMD in mutants. However, gRNA from Li et al, 2020 targeting the 3' end of exon 4 of the *smyhc1* locus may have led to off-target mutations in *smyhc2-5* as gRNA design show 100% identity to *smyhc2-5* sequence (Appendix 5.3). Moreover, the TALENS design from Whittle et al, 2020 leading frameshift mutation leading to an early stop codon in exon 16 may result in off-target effects as exon 16 shows 87-89% sequence similarity to myh7 and smyhc2 (Appendix 5.3). Ensuring the highest specificity to *smyhc1*, we

optimised our CRISPR/Cas9 mutations to exclusively target *smyhc1* and BLAST analysis of both our gRNA targets only to *smyhc1* giving confidence that our mutants show minimal to no off-target effects (Appendix 5.3).

Smyhc1 is predominantly expressed during the early stages of development whilst the expression of *smyhc2* and *smyhc3* are localised in a subset of muscles (Stone Elworthy *et al.*, 2008). Our *smyhc1*^{kg179} mutants show predominant loss of slow MyHC signal in the trunk but show some signal in the subset of slow fibres expressing *smyhc2* and *smyhc3* as described by Elworthy et al, 2008. There is a transition phase of *smyhc* expression from juvenile to adult zebrafish whereby *smyhc1* expression is replaced by *smyhc2* and *smyhc3* (Stone Elworthy *et al.*, 2008). In the absence of *smyhc1* at adult stages, our *smyhc1*^{kg179} mutants show full recovery of slow muscle motility and suggest lack of *smyhc1* does not play a template role for the integration of *smyhc2* and *smyhc3*. Since our findings show phenotype during early stages in *smyhc1* KO mutants and full recovery of phenotype when *smyhc1* is no longer required for motility, the lack of phenotype during adult stages reflects the transition from *smyhc1* in early developing larvae to *smyhc2* and *smyhc3* in adults and the possibility of phenotypes in adult *smyhc1* KO mutants are unlikely.

6.6. Role of *smyhc1* in sarcomere assembly

Smyhc1 play a key role in sarcomere assembly during the early stages of development as it is the first MyHC to be expressed in zebrafish slow fibres (Devoto *et al.*, 1996). Studies have been made to model and describe the sequence of events in sarcomere assembly during muscle fibre growth (Rhee, Sanger and Sanger, 1994; Holtzer *et al.*, 1997; Ehler *et al.*, 1999; Gregorio *et al.*, 1999; Rui, Bai and Perrimon, 2010; Fenix *et al.*, 2018). During early zebrafish development, *smyhc1* has been shown as essential for thick filament assembly and myofibril organisation in slow fibres (Li *et al.*, 2020; Whittle *et al.*, 2020). The first step for myofibrillogenesis is for the accumulation of integrins, α-actinin, vinculin and talin at the myotendinous junction (MTJ) which are the somite borders in zebrafish (Kelly and Zacks, 1969; Tokuyasu, 1989). Z-disks are formed initially as aggregates called z-bodies and are also accumulated MTJ (Tokuyasu and Maher, 1987). Mice and drosophila have shown ligands at the extracellular matrix are essential for the z-disk formation as the first step in sarcomere assembly (Volk, Fessler and Fessler, 1990; Bloor and Brown, 1998). Our *smyhc1*^{kg179} mutants show accumulation of α-actinin and actin filaments at the somite border supporting the model describing actin and α-actinin anchoring to the MTJ before integrating myosin filaments in building and elongating the muscle fibre.

After initiation of sarcomere at the MTJ, elongation of sarcomere can start to build. There are three main models describing aspects of sarcomere assembly after the initial anchoring at the MTJ: The stitching model, describing the independent assembly of I-Z-I bodies before myosin integration (Rhee, Sanger and Sanger, 1994; Holtzer *et al.*, 1997; Van Der Ven *et al.*, 1999; Sanger *et al.*, 2005), A pre myofibril model describing the formation of stress fibre-like structures utilising non-muscle myosin as a template for sarcomere proteins to assemble (Rhee, Sanger and Sanger, 1994), the model utilising titin as a molecular ruler, describing template assembly of sarcomere proteins according to the titin molecule (Kelly and Zacks, 1969; Tokuyasu and Maher, 1987; Schwander *et al.*, 2003; Au *et al.*, 2004). All three models suggest the integration of myosin into the sarcomere is the last step in myofibrillogenesis. In our *smyhc1*^{kg179} mutants, initiation of sarcomere assembly showing anchoring Z-bodies to the somite border occurred, but the elongation step of the myofibril remains absent and thus, myofibrils do not elongate supporting the studies showing the integration of myosin as one of the last steps in myofibrillogenesis.

6.7. Role of zebrafish *smyhc* genes in sarcomere assembly

Smyhc1^{kg179} mutants show defective sarcomere organisation at the early stages of development and give insight into the role of one *MYH7* ortholog, *smyhc1.* However, the subsequent transitional role of *MYH7* from juvenile to adulthood remains in question. Since *smyhc2* and *smyhc3* replace *smyhc1* in adult zebrafish, the defective phenotype at the early stages of development show recovery in adult stages. Since humans only have one slow myosin gene from development to adulthood, studying all smyhc orthologs to human *MYH7* may describe the transitional role of slow myosin in the developing muscle. Currently, there are no knockout studies have been made on these genes. Despite defects in slow fibre sarcomere assembly in *smyhc1*^{kg179} mutants during early development, there were no defects in the migration of slow muscle precursor into the superficial layer of elongated slow fibres where juvenile to adult *smyhc1*^{kg179} mutants show does not show an obvious phenotype. During early slow fibre development, *smyhc1* is predominantly expressed and *smyhc2* and *smyhc3* are expressed in a subset of muscles (Stone Elworthy *et al.*, 2008). In the adult stages, the predominant expression of *smyhc1* diminishes at 42 dpf and is replaced by *smyhc2* and *smyhc3* in mature slow fibres (Stone Elworthy *et al.*, 2008; Li *et al.*, 2020). Our data showing immotility from *smyhc1*^{kg179} mutants at early stages of development followed by the recovery of phenotype in juvenile and adults correlate to the predominant expression of *smyhc1* in young larvae transitioning to *smyhc2* and *smyhc3* in juvenile to adulthood. Thus, defective sarcomere organisation in slow fibres from lack of *smyhc1* at the early stages of development is not essential for sarcomere organisation in secondary fibres in juveniles to

adulthood. Knockout of the *smyhc* locus will be crucial to identify the role of *MYH7* from early development to adulthood and identify the role of *smyhc2-5* in sarcomere assembly and organisation. Ongoing work to generate a large deletion of the smyhc locus will give insight into the role of *MYH7* orthologs *smyhc1-5* in zebrafish for sarcomere assembly.

6.8. Limitations and Future Directions

Our results to study the primary biophysical defects in the presence of *MYH7* mutations using human fibres show several limitations when using frozen muscle biopsy specimens from the 19 patients in our study. Initial sample size calculation for human single fibre analysis were not possible prior to receiving patient samples. Muscle samples received were ethically approved and come from European Biobanks (MRC Neuromuscular Centre and Italian Telethon Biobank) and stored in our -80 degrees Celsius freezer (Ethics approval has already been obtained). Despite calculating sample sizes after receiving patient samples to be able to find a difference between patients, there was high variability seen in our results when analysing our muscle fibres. Any differences in the method of obtaining biopsies, post-processing stages for storage of samples, patient health, ethnicity and activity background are unknown between samples and may have led to high variability seen in our results studying our samples. A second factor that may have also led to high variability within our samples may have been the wide range of age between patients and limited number of MSM patient data as current patients with LDM or MSM are rare and/or *de novo* mutations in humans thus, sourcing high number of samples from many patients are not possible. The third limitation to our muscle fibre sample data is the small amounts of muscle fibres per patient received. Such small numbers of fibres per patient limited the number of experiments done and decision to only perform 2 types of experimental assays (thick filament measurement and myosin head positioning).

There were also limitations in generating our zebrafish disease models and *smyhc1* KO lines. Firstly, the generation of *smyhc1* KO mutants limited to only study developmental defects at the early stages but not in the later stages as *smyhc2* and *smyhc3* replace the functional role in adult zebrafish. Zebrafish development usually take 3 months to reach breeding age and generation of homozygous mutants require 3 generations of breeding. Generation of large deletion after generating my initial *smyhc1* mutants limited the amount of time I would have been able to screen such mutations and also to breed them to homozygosity. Not only the lack of time in generating deletion of whole smyhc locus, I was also limited in being able to study the role of each individual *smyhc* gene to be able to fully confirm the potential off target results in other *smyhc* genes seen in Li et al, 2020 and Whittle et al,

2020. Future work to either continue to generate large deletion in both wild type and *smyhc1*$^{-/-}$ mutants or generate individual smyhc mutants to study the role of each smyhc gene over the course of muscle development. Secondly, in attempt to generate specific mutations in disease models, using homologous recombination with short single oligonucleotides are rare events, however not impossible as there have been many studies shown to generate small and specific mutations with this method (Hruscha *et al.*, 2013; Hwang *et al.*, 2013; Armstrong *et al.*, 2016). Another method could be considered in the future is to utilise the recently availability of prime editing (Anzalone *et al.*, 2019), utilising a dead Cas9 protein nicks at target site and a pegRNA (gRNA with prime editing feature) is fused to template DNA for homologous recombination of edited gene. Prime editing minimised the possible INDELS generated by double strand breaks in my previous method utilising the regular Cas9 and subsequent short single oligonucleotide donor and show high success for specific point mutation gene editing. Another method to generate disease models for LDM and MSM in the future would be to insert an expression vector to express defective smyhc genes containing LDM or MSM disease mutations in our *smyhc1* mutants and further future, the large *smyhc* locus deletion mutants. The exact same expression vector method but utilising wild type human *MYH7* to identify whether human *MYH7* would recover defects seen in *smyhc1* mutants and in the larger *smyhc* whole locus deletion.

6.9. Final Conclusion

The two congenital myopathies that I have focused on, LDM and MSM currently no curative medicines and available treatments simply target the various symptoms (*Myosin storage myopathy*, 2016; Topaloglu, 2020). I demonstrated that there is no overall alteration in sarcomere organisation in the presence of defective myosin molecules but rather affect the *MYH7* MyBP-C binding domain destabilise the SRX state. I highlight the possibility of testing the drug Mavacamten or a derivative of this drug to target LDM mutations through stabilising the SRX state, as a similar method for the treatment for HCM. Moving to the future disease models can be generated using zebrafish *smyhc1* KO and *smyhc1-5* whole locus KO with the addition of defective myosin molecules using an expression vector. Overall, current data give early insight into the mechanism for the role of slow myosin in sarcomere assembly in zebrafish and can be used as an accurate disease model to further study the mechanism behind the two diseases LDM and MSM for targeted drug testing.

References

A.T.Geisterfer-Lowrance, A. *et al.* (1990) 'A molecular basis for familial hypertrophic cardiomyopathy: An α β cardiac myosin heavy chain hybrid gene', *Cell*, 62(5), pp. 991–998. doi: 10.1016/0092-8674(90)90273-H.

Adhikari, A. S. *et al.* (2019) 'β-Cardiac myosin hypertrophic cardiomyopathy mutations release sequestered heads and increase enzymatic activity', *Nature Communications*, 10(1), pp. 1–10. doi: 10.1038/s41467-019-10555-9.

Agarkova, I. *et al.* (2003) 'M-band: A safeguard for sarcomere stability?', *Journal of Muscle Research and Cell Motility*, 24(2–3), pp. 191–203. doi: 10.1023/A:1026094924677.

Agarkova, I. *et al.* (2004) 'The molecular composition of the sarcomeric M-band correlates with muscle fiber type', *European Journal of Cell Biology*, 83(5), pp. 193–204. doi: 10.1078/0171-9335-00383.

Alamo, L. *et al.* (2008) 'Three-Dimensional Reconstruction of Tarantula Myosin Filaments Suggests How Phosphorylation May Regulate Myosin Activity', *Journal of Molecular Biology*, 384(4), pp. 780–797. doi: 10.1016/j.jmb.2008.10.013.

Alamo, L. *et al.* (2016) 'Conserved Intramolecular Interactions Maintain Myosin Interacting-Heads Motifs Explaining Tarantula Muscle Super-Relaxed State Structural Basis', *Journal of Molecular Biology*, 428(6), pp. 1142–1164. doi: 10.1016/j.jmb.2016.01.027.

Alamo, L. *et al.* (2017) 'Effects of myosin variants on interacting-heads motif explain distinct hypertrophic and dilated cardiomyopathy phenotypes', *eLife*, 6, pp. 1–31. doi: 10.7554/eLife.24634.

Alberts, B. *et al.* (2015) 'The Cytoskeleton: Myosin and Actin', in *Molecular Cell Biology of The Cell*. 6th Editio, pp. 889–962. Available at: https://www.ncbi.nlm.nih.gov/books/NBK21724/.

Amemiya, C. T. *et al.* (2013) 'The African coelacanth genome provides insights into tetrapod evolution', *Nature*, 496(7445), pp. 311–316. doi: 10.1038/nature12027.

Amores, A. *et al.* (1998) 'Zebrafish hox clusters and vertebrate genome evolution', *Science*, 282(5394), pp. 1711–1714. doi: 10.1126/science.282.5394.1711.

Anderson, R. *et al.* (2018) 'Mavacamten stabilizes a folded-back sequestered super-relaxed state of β-cardiac myosin', *bioRxiv*, p. 266783. doi: 10.1101/266783.

Anzalone, A. V. *et al.* (2019) *Search-and-replace genome editing without double-strand breaks or donor DNA*, *Nature*. doi: 10.1038/s41586-019-1711-4.

Armstrong, G. A. B. *et al.* (2016) 'Homology directed knockin of point mutations in the zebrafish tardbp and fus genes in ALS using the CRISPR/Cas9 system', *PLoS ONE*, 11(3), pp. 1–10. doi: 10.1371/journal.pone.0150188.

Arredondo, J. J. *et al.* (2001) 'Control of Drosophila paramyosin/miniparamyosin gene expression. Differential regulatory mechanisms for muscle-specific transcription', *Journal of Biological Chemistry*, 276(11), pp. 8278–8287. doi: 10.1074/jbc.M009302200.

Atkinson, S. J. and Stewart, M. (1991) 'Expression in Escherichia coli of fragments of the coiled-coil rod domain of rabbit myosin: Influence of different regions of the molecule on aggregation and paracrystal formation', *Journal of Cell Science*, 99(4), pp. 823–836.

Atkinson, S. J. and Stewart, M. (1992) 'Molecular interactions in myosin assembly. Role of the 28-residue charge repeat in the rod', *Journal of Molecular Biology*, 226(1), pp. 7–13. doi: 10.1016/0022-2836(92)90118-4.

Au, Y. *et al.* (2004) 'Solution structure of ZASP PDZ domain: Implications for sarcomere ultrastructure and enigma family redundancy', *Structure*, 12(4), pp. 611–622. doi: 10.1016/j.str.2004.02.019.

Auman, H. J. *et al.* (2007) 'Functional modulation of cardiac form through regionally confined cell shape changes', *PLoS Biology*, 5(3), pp. 0604–0615. doi: 10.1371/journal.pbio.0050053.

Bandman, E. and Rosser, B. W. C. (2000) 'Evolutionary significance of myosin heavy chain heterogeneity in birds', *Microscopy Research and Technique*, 50(6), pp. 473–491. doi: 10.1002/1097-0029(20000915)50:6<473::AID-JEMT5>3.0.CO;2-R.

Barohn, R. J., Brumback, R. A. and Mendell, J. R. (1994) 'Hyaline body myopathy', *Neuromuscular Disorders*. doi: 10.1016/0960-8966(94)90027-2.

Berger, J., Hall, T. E. and Currie, P. D. (2015) 'Novel Transgenic Lines to Label Sarcolemma and Myofibrils of the Musculature', *Zebrafish*, 12(1), pp. 124–125. doi: 10.1089/zeb.2014.1065.

Van den Bergh, P. Y. K. *et al.* (2014) 'Laing early-onset distal myopathy in a Belgian family', *Acta Neurologica Belgica*, 114(4), pp. 253–256. doi: 10.1007/s13760-014-0298-7.

Bernstein, S. I. *et al.* (1983) 'Drosophila muscle myosin heavy chain encoded by a single gene in a cluster of muscle mutations', *Nature*, 302(5907), pp. 393–397. doi: 10.1038/302393a0.

Bitoun, M. *et al.* (2005) 'Mutations in dynamin 2 cause dominant centronuclear myopathy', *Nature Genetics*, 37(11), pp. 1207–1209. doi: 10.1038/ng1657.

Blagden, C. S. *et al.* (1997) 'Notochord induction of zebrafish slow muscle mediated by sonic hedgehog', *Genes and Development*, 11(17), pp. 2163–2175. doi: 10.1101/gad.11.17.2163.

Bloor, J. W. and Brown, N. H. (1998) 'Genetic analysis of the Drosophila α(PS2) integrin subunit reveals discrete adhesive, morphogenetic and sarcomeric functions', *Genetics*, 148(3), pp. 1127–1142. doi: 10.1093/genetics/148.3.1127.

Bohlega, S. *et al.* (2003) 'Autosomal dominant hyaline body myopathy: Clinical variability and pathologic findings', *Neurology*, 61(11), pp. 1519–1523. doi: 10.1212/01.WNL.0000096022.09887.9D.

Bollen, I. A. E. *et al.* (2017) 'Myofilament Remodeling and Function Is More Impaired in Peripartum Cardiomyopathy Compared with Dilated Cardiomyopathy and Ischemic Heart Disease', *American Journal of Pathology*, 187(12), pp. 2645–2658. doi: 10.1016/j.ajpath.2017.08.022.

Boycott, K. M. *et al.* (2013) 'Rare-disease genetics in the era of next-generation sequencing: Discovery to translation', *Nature Reviews Genetics*, 14(10), pp. 681–691. doi: 10.1038/nrg3555.

Bryson-Richardson, R. J. *et al.* (2005) 'Myosin heavy chain expression in zebrafish and slow muscle composition', *Developmental Dynamics*, 233(3), pp. 1018–1022. doi: 10.1002/dvdy.20380.

Cancilla, P. A. *et al.* (1971) 'Familial myopathy with probable lysis of myofibrils in type I fibers', *Neurology*. doi: 10.1212/WNL.21.6.579.

Cazorla, O. *et al.* (2001) 'Titin-based modulation of calcium sensitivity of active tension in mouse skinned cardiac myocytes', *Circulation Research*, 88(10), pp. 1028–1035. doi: 10.1161/hh1001.090876.

Chang, H. H. Y. *et al.* (2017) 'Non-homologous DNA end joining and alternative pathways to double-strand break repair', *Nature Reviews Molecular Cell Biology*, 18(8), pp. 495–506. doi: 10.1038/nrm.2017.48.

Chang, N. *et al.* (2013) 'Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos', *Cell Research*, 23(4), pp. 465–472. doi: 10.1038/cr.2013.45.

Chen, Q. *et al.* (1997) 'Identification of a genomic locus containing three slow myosin heavy chain genes in the chicken', *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1353(2), pp. 148–156. doi: 10.1016/S0167-4781(97)00067-5.

Cheung, A. *et al.* (2002) 'A small-molecule inhibitor of skeletal muscle myosin II', *Nature Cell Biology*, 4(1), pp. 83–88. doi: 10.1038/ncb734.

Chevallier, A., Kieny, M. and Mauger, A. (1977) 'Limb-somite relationship: origin of the limb musculature.', *J Embrol Exp Morphol*, 41, pp. 245–258.

Cho, M., Webster, S. G. and Blau, H. M. (1993) 'Evidence for myoblast-extrinsic regulation of slow myosin heavy chain expression during muscle fiber formation in embryonic development', *Journal of Cell Biology*, 121(4), pp. 795–810. doi: 10.1083/jcb.121.4.795.

Clarke, N. F. *et al.* (2008) 'Mutations in TPM3 are a common cause of congenital fiber type disproportion', *Annals of Neurology*, 63(3), pp. 329–337. doi: 10.1002/ana.21308.

Codina, M. *et al.* (2010) 'Loss of Smyhc1 or Hsp90α1 function results in different effects on myofibril organization in skeletal muscles of zebrafish embryos', *PLoS ONE*, 5(1). doi: 10.1371/journal.pone.0008416.

Cohen, C. and Holmes, K. C. (1963) 'X-ray diffraction evidence for α-helical coiled-coils in native

muscle', *Journal of Molecular Biology*, 6(5), pp. 423–432. doi: 10.1016/S0022-2836(63)80053-4.

Cooke, R. (2011) 'The role of the myosin ATPase activity in adaptive thermogenesis by skeletal muscle', *Biophysical Reviews*, 3(1), pp. 33–45. doi: 10.1007/s12551-011-0044-9.

Cooper, M. S. *et al.* (2005) 'Visualizing morphogenesis in transgenic zebrafish embryos using BODIPY TR methyl ester dye as a vital counterstain for GFP', *Developmental Dynamics*, 232(2), pp. 359–368. doi: 10.1002/dvdy.20252.

Cossu, G. *et al.* (1996) 'Activation of different myogenic pathways: myf-5 is induced by the neural tube and MyoD by the dorsal ectoderm in mouse paraxial mesoderm', *Development*, 122(2), pp. 429–437. doi: 10.1242/dev.122.2.429.

Coureux, P. D. *et al.* (2003) 'A structural state of the myosin V motor without bound nucleotide', *Nature*, 425(6956), pp. 419–423. doi: 10.1038/nature01927.

Crick, F. H. C. (1953) 'The packing of α-helices: simple coiled-coils', *Acta Crystallographica*, 6(8), pp. 689–697. doi: 10.1107/s0365110x53001964.

Cripps, R. M. (1999) 'Assembly of thick filaments and myofibrils occurs in the absence of the myosin head', *The EMBO Journal*, 18(7), pp. 1793–1804. doi: 10.1093/emboj/18.7.1793.

Crow, M. T. and Stockdale, F. E. (1986) 'Myosin expression and specialization among the earliest muscle fibers of the developing avian limb', *Developmental Biology*, 113(1), pp. 238–254. doi: 10.1016/0012-1606(86)90126-0.

Cullup, T. *et al.* (2012) 'Mutations in MYH7 cause Multi-minicore Disease (MmD) with variable cardiac involvement', *Neuromuscular Disorders*, 22(12), pp. 1096–1104. doi: 10.1016/j.nmd.2012.06.007.

Currie, P. D. and Ingham, P. W. (1996) 'Induction of a specific muscle cell type by a hedgehog-like protein in zebrafish', *Nature*, 382(6590), pp. 452–455. doi: 10.1038/382452a0.

D'Agostino, C. *et al.* (2011) 'In sporadic inclusion body myositis muscle fibres TDP-43-positive inclusions are less frequent and robust than p62 inclusions, and are not associated with paired helical filaments', *Neuropathology and Applied Neurobiology*, 37(3), pp. 315–320. doi: 10.1111/j.1365-2990.2010.01108.x.

Daggett, D. F. *et al.* (2007) 'Control of morphogenetic cell movements in the early zebrafish myotome', *Developmental Biology*, 309(2), pp. 169–179. doi: 10.1016/j.ydbio.2007.06.008.

Dahl-Halvarsson, M. *et al.* (2017) 'Myosin storage myopathy in C. Elegans and human cultured muscle cells', *PLoS ONE*, 12(1), pp. 1–14. doi: 10.1371/journal.pone.0170613.

Dahl-Halvarsson, M. *et al.* (2018) 'Drosophila model of myosin myopathy rescued by overexpression of a TRIM-protein family member', *Proceedings of the National Academy of Sciences of the United States of America*, 115(28), pp. E6566–E6575. doi: 10.1073/pnas.1800727115.

Daneshparvar, N. *et al.* (2020) 'CryoEM structure of Drosophila flight muscle thick filaments at 7 A resolution', *Life Science Alliance*, 3(8), pp. 1–14. doi: 10.26508/LSA.202000823.

Darin, N. *et al.* (2007) 'New skeletal myopathy and cardiomyopathy associated with a missense mutation in MYH7', *Neurology*. doi: 10.1212/01.wnl.0000264430.55233.72.

Desjardins, P. R. *et al.* (2002) 'Evolutionary implications of three novel members of the human sarcomeric myosin heavy chain gene family', *Molecular Biology and Evolution*, 19(4), pp. 375–393. doi: 10.1093/oxfordjournals.molbev.a004093.

Devoto, S. H. *et al.* (1996) 'Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation', *Development*, 122(11), pp. 3371–3380. doi: 10.1242/dev.122.11.3371.

Dlugosz, A. A. *et al.* (1984) 'The relationship between stress fiber-like structures and nascent myofibrils in cultured cardiac myocytes.', *Journal of Cell Biology*, 99(6), pp. 2268–2278. doi: 10.1083/jcb.99.6.2268.

Du, A. *et al.* (2003) 'Myofibrillogenesis in the first cardiomyocytes formed from isolated quail precardiac mesoderm', *Developmental Biology*, 257(2), pp. 382–394. doi: 10.1016/S0012-1606(03)00104-0.

Du, A., Sanger, J. M. and Sanger, J. W. (2008) 'Cardiac myofibrillogenesis inside intact embryonic hearts', *Developmental Biology*, 318(2), pp. 236–246. doi: 10.1016/j.ydbio.2008.03.011.

Ehler, E. *et al.* (1999) 'Myofibrillogenesis in the developing chicken heart: Assembly of Z-disk, M-line and the thick filaments', *Journal of Cell Science*, 112(10), pp. 1529–1539. doi: 10.1242/jcs.112.10.1529.

Elworthy, Stone *et al.* (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(12), pp. 2115–2126. doi: 10.1242/dev.015719.

Elworthy, S. *et al.* (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(12), pp. 2115–2126. doi: 10.1242/dev.015719.

Epp, T. A. *et al.* (1993) 'Structural organization of the human cardiac α-myosin heavy chain gene (myh6)', *Genomics*, 18(3), pp. 505–509. doi: 10.1016/S0888-7543(11)80006-6.

Ervasti, J. M. (2003) 'Costameres: The Achilles' heel of Herculean muscle', *Journal of Biological Chemistry*, 278(16), pp. 13591–13594. doi: 10.1074/jbc.R200021200.

Fatkin, D. and Graham, R. M. (2002) 'Molecular Mechanisms of Inherited Cardiomyopathies', *Physiological Reviews*. doi: 10.1152/physrev.00012.2002.

Felsenfeld, A. L., Curry, M. and Kimmel, C. B. (1991) 'The fub-1 mutation blocks initial myofibril formation in zebrafish muscle pioneer cells', *Developmental Biology*, 148(1), pp. 23–30. doi:

10.1016/0012-1606(91)90314-S.

Fenix, A. M. *et al.* (2018) 'Muscle-specific stress fibers give rise to sarcomeres in cardiomyocytes', *eLife*, 7, pp. 1–33. doi: 10.7554/eLife.42144.

Fiorillo, C. *et al.* (2016) 'MYH7-related myopathies: Clinical, histopathological and imaging findings in a cohort of Italian patients', *Orphanet Journal of Rare Diseases*, 11(1). doi: 10.1186/s13023-016-0476-1.

Frazier, A. *et al.* (2008) 'Familial hypertrophic cardiomyopathy associated with cardiac β-myosin heavy chain and troponin I mutations', *Pediatric Cardiology*, 29(4), pp. 846–850. doi: 10.1007/s00246-007-9177-9.

Freiburg, A. *et al.* (2000) 'Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity', *Circulation Research*, 86(11), pp. 1114–1121. doi: 10.1161/01.RES.86.11.1114.

Frisbee, J. C. and Barclay, J. K. (1999) 'Microvascular hematocrit and permeability-surface area product of in situ canine skeletal muscle during fatigue', *Microvascular Research*, 57(2), pp. 203–207. doi: 10.1006/mvre.1998.2118.

Fujita, H., Nedachi, T. and Kanzaki, M. (2007) 'Accelerated de novo sarcomere assembly by electric pulse stimulation in C2C12 myotubes', *Experimental Cell Research*, 313(9), pp. 1853–1865. doi: 10.1016/j.yexcr.2007.03.002.

Fukuda, N. *et al.* (2005) 'Titin-based modulation of active tension and interfilament lattice spacing in skinned rat cardiac muscle', *Pflugers Archiv European Journal of Physiology*, 449(5), pp. 449–457. doi: 10.1007/s00424-004-1354-6.

Fusi, L., Huang, Z. and Irving, M. (2015) 'The Conformation of Myosin Heads in Relaxed Skeletal Muscle: Implications for Myosin-Based Regulation', *Biophysical Journal*, 109(4), pp. 783–792. doi: 10.1016/j.bpj.2015.06.038.

Garfinkel, A. C., Seidman, J. G. and Seidman, C. E. (2018) 'Genetic Pathogenesis of Hypertrophic and Dilated Cardiomyopathy', *Heart Failure Clinics*, 14(2), pp. 139–146. doi: 10.1016/j.hfc.2017.12.004.

George, E. L., Ober, M. B. and Emerson, C. P. (1989) 'Functional domains of the Drosophila melanogaster muscle myosin heavy-chain gene are encoded by alternatively spliced exons', *Molecular and Cellular Biology*, 9(7), pp. 2957–2974. doi: 10.1128/mcb.9.7.2957-2974.1989.

Goebel, H. H. and Blaschek, A. (2011) 'Protein Aggregation in Congenital Myopathies', *Seminars in Pediatric Neurology*, 18(4), pp. 272–276. doi: 10.1016/j.spen.2011.10.009.

Gokhin, D. S. *et al.* (2012) 'Thin-filament length correlates with fiber type in human skeletal muscle', *AJP: Cell Physiology*, 302(3), pp. C555–C565. doi: 10.1152/ajpcell.00299.2011.

Gokhin, D. S. *et al.* (2014) 'Alterations in thin filament length during postnatal skeletal muscle development and aging in mice', *Frontiers in Physiology*, 5(SEP), pp. 1–6. doi: 10.3389/fphys.2014.00375.

Gonzalez-Sanchez, A. and Bader, D. (1985) 'Characterization of a myosin heavy chain in the conductive system of the adult and developing chicken heart', *Journal of Cell Biology*, 100(1), pp. 270–275. doi: 10.1083/jcb.100.1.270.

González-Serratos, H. (1971) 'Inward spread of activation in vertebrate muscle fibres', *The Journal of Physiology*, 212(3), pp. 777–799. doi: 10.1113/jphysiol.1971.sp009356.

Gordon, A. M., Homsher, E. and Regnier, M. (2000) 'Regulation of contraction in striated muscle', *Physiological Reviews*, 80(2), pp. 853–924. doi: 10.1152/physrev.2000.80.2.853.

Green, E. M. *et al.* (2016) 'Heart disease: A small-molecule inhibitor of sarcomere contractility suppresses hypertrophic cardiomyopathy in mice', *Science*, 351(6273), pp. 617–621. doi: 10.1126/science.aad3456.

Gregorio, C. C. *et al.* (1999) 'Muscle assembly: A titanic achievement?', *Current Opinion in Cell Biology*, 11(1), pp. 18–25. doi: 10.1016/S0955-0674(99)80003-9.

Groen, E. J. *et al.* (2007) 'Analysis of the UK diagnostic strategy for limb girdle muscular dystrophy 2A', *Brain*, 130(12), pp. 3237–3249. doi: 10.1093/brain/awm259.

Groves, J. A., Hammond, C. L. and Hughes, S. M. (2005) 'Fgf8 drives myogenic progression of a novel lateral fast muscle fibre population in zebrafish', *Development*, 132(19), pp. 4211–4222. doi: 10.1242/dev.01958.

Gulick, J. *et al.* (1991) 'Isolation and characterization of the mouse cardiac myosin heavy chain genes', *Journal of Biological Chemistry*, 266(14), pp. 9180–9185. doi: 10.1016/s0021-9258(18)31568-0.

Gurevich, D. B. *et al.* (2016) 'Asymmetric division of clonal muscle stem cells coordinates muscle regeneration in vivo', *Science*, 353(6295). doi: 10.1126/science.aad9969.

Gutiérrez-Lovera, C. *et al.* (2017) 'The Potential of Zebrafish as a Model Organism for Improving the Translation of Genetic', *Genes*, 12(12), pp. 1–20. doi: 10.3390/genes8120349.

Hammond, C. L. *et al.* (2007) 'Signals and myogenic regulatory factors restrict pax3 and pax7 expression to dermomyotome-like tissue in zebrafish', *Developmental Biology*, 302(2), pp. 504–521. doi: 10.1016/j.ydbio.2006.10.009.

Hang, C. *et al.* (2021) 'Knockout of MYOM1 in human cardiomyocytes leads to myocardial atrophy via impairing calcium homeostasis', *Journal of Cellular and Molecular Medicine*, 25(3), pp. 1661–1676. doi: 10.1111/jcmm.16268.

Haselgrove, J. C. (1975) 'X-ray evidence for conformational changes in the myosin filaments of

vertebrate striated muscle', *Journal of Molecular Biology*, 92(1), pp. 113–143. doi: 10.1016/0022-2836(75)90094-7.

Hastings, G. A. and Emerson, C. P. (1991) 'Myosin functional domains encoded by alternative exons are expressed in specific thoracic muscles of Drosophila', *Journal of Cell Biology*, 114(2), pp. 263–276. doi: 10.1083/jcb.114.2.263.

Hegde, S. M. *et al.* (2021) 'Effect of Mavacamten on Echocardiographic Features in Symptomatic Patients With Obstructive Hypertrophic Cardiomyopathy', *Journal of the American College of Cardiology*, 78(25), pp. 2518–2532. doi: 10.1016/j.jacc.2021.09.1381.

Hess, N. K. *et al.* (2007) 'Transcriptional regulation of the Drosophila melanogaster muscle myosin heavy-chain gene', *Gene Expression Patterns*, 7(4), pp. 413–422. doi: 10.1016/j.modgep.2006.11.007.

Holtzer, H. *et al.* (1997) 'Independent assembly of 1.6 μm long bipolar MHC filaments and I-Z-I bodies', *Cell Structure and Function*, 22(1), pp. 83–93. doi: 10.1247/csf.22.83.

Homayoun, H. *et al.* (2011) 'Novel mutation in MYH7 gene associated with distal myopathy and cardiomyopathy', *Neuromuscular Disorders*, 21(3), pp. 219–222. doi: 10.1016/j.nmd.2010.12.005.

Hooijman, P., Stewart, M. A. and Cooke, R. (2011) 'A new state of cardiac myosin with very slow ATP turnover: A potential cardioprotective mechanism in the heart', *Biophysical Journal*, 100(8), pp. 1969–1976. doi: 10.1016/j.bpj.2011.02.061.

Hoshijima, K. *et al.* (2019) 'Highly Efficient CRISPR-Cas9-Based Methods for Generating Deletion Mutations and F0 Embryos that Lack Gene Function in Zebrafish', *Developmental Cell*, 51(5), pp. 645-657.e4. doi: 10.1016/j.devcel.2019.10.004.

Howard, J. (1997) 'Molecular motors: Structural adaptations to cellular functions', *Nature*, 389(6651), pp. 561–567. doi: 10.1038/39247.

Hruscha, A. *et al.* (2013) 'Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish', *Development (Cambridge)*, 140(24), pp. 4982–4987. doi: 10.1242/dev.099085.

Hu, L. Y. R., Ackermann, M. A. and Kontrogianni-Konstantopoulos, A. (2015) 'The sarcomeric M-region: A molecular command center for diverse cellular processes', *BioMed Research International*, 2015. doi: 10.1155/2015/714197.

Hu, Z. *et al.* (2016) 'Structure of myosin filaments from relaxed Lethocerus flight muscle by cryo-EM at 6 Å resolution', *Science Advances*, 2(9). doi: 10.1126/sciadv.1600058.

Huang, H. *et al.* (2005) 'NXT2 is required for embryonic heart development in zebrafish', *BMC Developmental Biology*, 5, pp. 1–12. doi: 10.1186/1471-213X-5-7.

Hug, N., Longman, D. and Cáceres, J. F. (2016) 'Mechanism and regulation of the nonsense-mediated decay pathway', *Nucleic Acids Research*, 44(4), pp. 1483–1495. doi: 10.1093/nar/gkw010.

Huxley, H. E. and Brown, W. (1967) 'The low-angle X-ray diagram of vertebrate striated muscle and its behaviour during contraction and rigor', *Journal of Molecular Biology*, 30(2), pp. 383–434. doi: 10.1016/S0022-2836(67)80046-9.

Hwang, W. Y. *et al.* (2013) 'Efficient genome editing in zebrafish using a CRISPR-Cas system', *Nature Biotechnology*, 31(3), pp. 227–229. doi: 10.1038/nbt.2501.

Ikebe, M. *et al.* (2001) 'The Tip of the Coiled-coil Rod Determines the Filament Formation of Smooth Muscle and Nonmuscle Myosin', *Journal of Biological Chemistry*, 276(32), pp. 30293–30300. doi: 10.1074/jbc.M101969200.

Ikeda, D. *et al.* (2007) 'Divergent evolution of the myosin heavy chain gene family in fish and tetrapods: Evidence from comparative genomic analysis', *Physiological Genomics*, 32(1), pp. 1–15. doi: 10.1152/physiolgenomics.00278.2006.

Irving, M. (2017) 'Regulation of Contraction by the Thick Filaments in Skeletal Muscle', *Biophysical Journal*, 113(12), pp. 2579–2594. doi: 10.1016/j.bpj.2017.09.037.

Irving, T. *et al.* (2011) 'Thick-filament strain and interfilament spacing in passive muscle: Effect of titin-based passive tension', *Biophysical Journal*, 100(6), pp. 1499–1508. doi: 10.1016/j.bpj.2011.01.059.

Jasin, M. and Rothstein, R. (2013) 'Repair of strand breaks by homologous recombination.', *Cold Spring Harbor perspectives in biology*, 5(11), pp. 1–18. doi: 10.1101/cshperspect.a012740.

Kassar-Duchossoy, L. *et al.* (2004) 'Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice', *Nature*, 431(7007), pp. 466–471. doi: 10.1038/nature02876.

Kassar-Duchossoy, L. *et al.* (2005) 'Pax3/Pax7 mark a novel population of primitive myogenic cells during development', *Genes and Development*, 19(12), pp. 1426–1431. doi: 10.1101/gad.345505.

Kelly, A. M. and Zacks, S. I. (1969) 'The histogenesis of rat intercostal muscle.', *The Journal of cell biology*, 42(1), pp. 135–153. doi: 10.1083/jcb.42.1.135.

Kelly, K. K., Meadows, S. M. and Cripps, R. M. (2002) 'Drosophila MEF2 is a direct regulator of Actin57B transcription in cardiac, skeletal, and visceral muscle lineages', *Mechanisms of Development*, 110(1–2), pp. 39–50. doi: 10.1016/S0925-4773(01)00586-X.

Kim, B. H. and Zhang, G. J. (2020) 'Generating stable knockout zebrafish lines by deleting large chromosomal fragments using multiple gRNAs', *G3: Genes, Genomes, Genetics*, 10(3), pp. 1029–1037. doi: 10.1534/g3.119.401035.

Kimmel, C. B. *et al.* (1995) 'Stages of embryonic development of the zebrafish.', *Developmental dynamics : an official public*, 203(3), pp. 253–310. doi: 10.1002/aja.1002030302.

Klitzman, B. and Duling, B. R. (1979) 'Microvascular hematocrit and red cell flow in resting and contracting striated muscle', *American Journal of Physiology - Heart and Circulatory Physiology*, 6(4).

doi: 10.1152/ajpheart.1979.237.4.h481.

Knappe, S., Zammit, P. S. and Knight, R. D. (2015) 'A population of Pax7-expressing muscle progenitor cells show differential responses to muscle injury dependent on developmental stage and injury extent', *Frontiers in Aging Neuroscience*, 7(AUG), pp. 1–17. doi: 10.3389/fnagi.2015.00161.

Komiyama, M., Maruyama, K. and Shimada, Y. (1990) 'Assembly of connectin (titin) in relation to myosin and α-actinin in cultured cardiac myocytes', *Journal of Muscle Research and Cell Motility*, 11(5), pp. 419–428. doi: 10.1007/BF01739762.

Koutsopoulos, O. S. *et al.* (2013) 'Dynamin 2 homozygous mutation in humans with a lethal congenital syndrome', *European Journal of Human Genetics*, 21(6), pp. 637–642. doi: 10.1038/ejhg.2012.226.

Labeit, S. and Kolmerer, B. (1995) 'Titins: Giant proteins in charge of muscle ultrastructure and elasticity', *Science*, 270(5234), pp. 293–296. doi: 10.1126/science.270.5234.293.

Laing, N. G. *et al.* (1995) 'Autosomal Dominant Distal Myopathy : Linkage to Chromosome 14', *American Journal of Human Genetics*, 56(2), pp. 422–427.

Laing, N. G. *et al.* (2004) 'Actin mutations are one cause of congenital fibre type disproportion', *Annals of Neurology*, 56(5), pp. 689–694. doi: 10.1002/ana.20260.

Laing, N. G. *et al.* (2005) 'Myosin storage myopathy: Slow skeletal myosin (MYH7) mutation in two isolated cases', *Neurology*. doi: 10.1212/01.WNL.0000150581.37514.30.

Lamont, P. J. *et al.* (2006) 'Laing early onset distal myopathy: Slow myosin defect with variable abnormalities on muscle biopsy', *Journal of Neurology, Neurosurgery and Psychiatry*, 77(2), pp. 208–215. doi: 10.1136/jnnp.2005.073825.

Lamont, P. J. *et al.* (2014) 'Novel Mutations Widen the Phenotypic Spectrum of Slow Skeletal/β-Cardiac Myosin (MYH7) Distal Myopathy', *Human Mutation*, 35(7), pp. 868–879. doi: 10.1002/humu.22553.

Laporte, G. J. *et al.* (no date) 'A gene mutated in X-linked conserved in yeast', pp. 175–182.

Lawlor, M. W. *et al.* (2010) 'Mutations of tropomyosin 3 (TPM3) are common and associated with type 1 myofiber hypotrophy in congenital fiber type disproportion', *Human Mutation*, 31(2), pp. 176–183. doi: 10.1002/humu.21157.

Lewis, K. E. *et al.* (1999) 'Control of muscle cell-type specification in the zebrafish embryo by Hedgehog signalling', *Developmental Biology*, 216(2), pp. 469–480. doi: 10.1006/dbio.1999.9519.

Li, M. and Arner, A. (2015) 'Immobilization of dystrophin and laminin α2-chain deficient zebrafish larvae in vivo prevents the development of muscular dystrophy', *PLoS ONE*, 10(11), pp. 1–14. doi: 10.1371/journal.pone.0139483.

Li, S. *et al.* (2020) 'Defective sarcomere organization and reduced larval locomotion and fish survival in slow muscle heavy chain 1 (smyhc1) mutants', *The FASEB Journal*, 34(1), pp. 1378–1397. doi:

10.1096/fj.201900935RR.

Liew, C. C. *et al.* (1990) 'Complete sequence and organization of the human cardiac β-myosin heavy chain gene', *Nucleic Acids Research*, 18(12), pp. 3647–3651. doi: 10.1093/nar/18.12.3647.

Lin, Z. *et al.* (1994) 'Sequential appearance of muscle-specific proteins in myoblasts as a function of time after cell division: Evidence for a conserved myoblast differentiation program in skeletal muscle', *Cell Motility and the Cytoskeleton*, 29(1), pp. 1–19. doi: 10.1002/cm.970290102.

Lowey, S., Waller, G. S. and Bandman, E. (1991) 'Neonatal and adult myosin heavy chains form homodimers during avian skeletal muscle development', *Journal of Cell Biology*, 113(2), pp. 303–310. doi: 10.1083/jcb.113.2.303.

Lu, B. D. *et al.* (1999) 'Spatial and temporal changes in myosin heavy chain gene expression in skeletal muscle development', *Developmental Biology*, 216(1), pp. 312–326. doi: 10.1006/dbio.1999.9488.

Lupas, A. (1996) 'Coiled coils: new structures and new functions', *Trends in Biochemical Sciences*, 21(10), pp. 375–382. doi: 10.1016/S0968-0004(96)10052-9.

Luther, P. K. *et al.* (2008) 'Understanding the Organisation and Role of Myosin Binding Protein C in Normal Striated Muscle by Comparison with MyBP-C Knockout Cardiac Muscle', *Journal of Molecular Biology*, 384(1), pp. 60–72. doi: 10.1016/j.jmb.2008.09.013.

Lykke-andersen, S. and Jensen, T. H. (2015) 'Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes', *Nature Reviews Molecular Cell Biology*, 16(11), pp. 665–677. doi: 10.1038/nrm4063.

Lyons, G. E. *et al.* (1990) 'The expression of myosin genes in developing skeletal muscle in the mouse embryo', *Journal of Cell Biology*, 111(4), pp. 1465–1476. doi: 10.1083/jcb.111.4.1465.

Ma, D. and Liu, F. (2015) 'Genome Editing and Its Applications in Model Organisms', *Genomics, Proteomics and Bioinformatics*, 13(6), pp. 336–344. doi: 10.1016/j.gpb.2015.12.001.

Ma, W. *et al.* (2018) 'Thick-Filament Extensibility in Intact Skeletal Muscle', *Biophysical Journal*, 115(8), pp. 1580–1588. doi: 10.1016/j.bpj.2018.08.038.

Machida, S. *et al.* (2002) 'Expression of slow skeletal myosin heavy chain 2 gene in Purkinje fiber cells in chick heart', *Biology of the Cell*, 94(6), pp. 389–399. doi: 10.1016/S0248-4900(02)00010-2.

Mahdavi, V., Periasamy, M. and Nadal-Ginard, B. (1982) 'Molecular characterization of two myosin heavy chain genes expressed in the adult heart', *Nature*, 297(5868), pp. 659–664. doi: 10.1038/297659a0.

Málnási-Csizmadia, A. *et al.* (2005) 'Switch movements and the myosin crossbridge stroke', *Journal of Muscle Research and Cell Motility*, 26(1), pp. 31–37. doi: 10.1007/s10974-005-9004-y.

Marín, M.-C., Rodríguez, J.-R. and Ferrús, A. (2004) 'Transcription of Drosophila Troponin I Gene Is

Regulated by Two Conserved, Functionally Identical, Synergistic Elements', *Molecular Biology of the Cell*, 15(3), pp. 1185–1196. doi: 10.1091/mbc.e03-09-0663.

Maron, B. J. *et al.* (1995) 'Prevalence of hypertrophic cardiomyopathy in a general population of young adults: Echocardiographic analysis of 4111 subjects in the CARDIA study', *Circulation*. doi: 10.1161/01.CIR.92.4.785.

Mas, J.-A., García-Zaragoza, E. and Cervera, M. (2004) 'Two Functionally Identical Modular Enhancers in Drosophila Troponin T Gene Establish the Correct Protein Levels in Different Muscle Types', *Molecular Biology of the Cell*, 15(4), pp. 1931–1945. doi: 10.1091/mbc.e03-10-0729.

Mastaglia, F. L. *et al.* (2002) 'Early onset chromosome 14-linked distal myopathy (Laing)', *Neuromuscular Disorders*, 12(4), pp. 350–357. doi: 10.1016/S0960-8966(01)00287-5.

Masuzugawa, S. *et al.* (1997) 'Autosomal dominant hyaline body myopathy presenting as scapuloperoneal syndrome: Clinical features and muscle pathology', *Neurology*. doi: 10.1212/WNL.48.1.253.

Maves, L. (2014) 'Recent advances using zebrafish animal models for muscle disease drug discovery', *Expert Opinion on Drug Discovery*, 9(9), pp. 1033–1045. doi: 10.1517/17460441.2014.927435.

McGuigan, K., Phillips, P. C. and Postlethwait, J. H. (2004) 'Evolution of sarcomeric myosin heavy chain genes: Evidence from fish', *Molecular Biology and Evolution*, 21(6), pp. 1042–1056. doi: 10.1093/molbev/msh103.

McLachlan, A. D. and Karn, J. (1982) 'Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacings in muscle', *Nature*, 299, pp. 226–231.

McNamara, J. W. *et al.* (2015) 'The role of super-relaxed myosin in skeletal and cardiac muscle', *Biophysical Reviews*, 7(1), pp. 5–14. doi: 10.1007/s12551-014-0151-5.

McNamara, J. W. *et al.* (2016) 'Ablation of cardiac myosin binding protein-C disrupts the super-relaxed state of myosin in murine cardiomyocytes', *Journal of Molecular and Cellular Cardiology*, 94, pp. 65–71. doi: 10.1016/j.yjmcc.2016.03.009.

Merrifield, P. A. *et al.* (1989) 'Temporal and tissue-specific expression of myosin heavy chain isoforms in developing and adult avian muscle', *Developmental Genetics*, 10(5), pp. 372–385. doi: 10.1002/dvg.1020100505.

Meyer, A. and Schartl, M. (1999) 'Gene and genome duplications in vertebrates: The one-to-four (-to-eight in fish) rule and the evolution of novel gene functions', *Current Opinion in Cell Biology*, 11(6), pp. 699–704. doi: 10.1016/S0955-0674(99)00039-3.

Miller, D. M., Stockdale, F. E. and Karn, J. (1986) 'Immunological identification of the genes encoding the four myosin heavy chain isoforms of Caenorhabditis elegans', *Proceedings of the National*

Academy of Sciences of the United States of America, 83(8), pp. 2305–2309. doi: 10.1073/pnas.83.8.2305.

Moss, R. L., Fitzsimons, D. P. and Ralphe, J. C. (2015) 'Cardiac MyBP-C regulates the rate and force of contraction in mammalian myocardium', *Circulation Research*, 116(1), pp. 183–192. doi: 10.1161/CIRCRESAHA.116.300561.

Muelas, N. *et al.* (2010) 'MYH7 gene tail mutation causing myopathic profiles beyond Laing distal myopathy', *Neurology*, 75(8), pp. 732–741. doi: 10.1212/WNL.0b013e3181eee4d5.

Myosin storage myopathy (2016) *National Centre for Advancing Translational Sciences*. Available at: https://rarediseases.info.nih.gov/diseases/7148/myosin-storage-myopathy.

Naganawa, Y. and Hirata, H. (2011) 'Developmental transition of touch response from slow muscle-mediated coilings to fast muscle-mediated burst swimming in zebrafish', *Developmental Biology*, 355(2), pp. 194–204. doi: 10.1016/j.ydbio.2011.04.027.

Narusawa, M. *et al.* (1987) 'Slow myosin in developing rat skeletal muscle', *Journal of Cell Biology*, 104(3), pp. 447–459. doi: 10.1083/jcb.104.3.447.

Nguyen, P. D. *et al.* (2017) 'Muscle Stem Cells Undergo Extensive Clonal Drift during Tissue Growth via Meox1-Mediated Induction of G2 Cell-Cycle Arrest', *Cell Stem Cell*, 21(1), pp. 107-119.e6. doi: 10.1016/j.stem.2017.06.003.

Nicot, A. S. *et al.* (2007) 'Mutations in amphiphysin 2 (BIN1) disrupt interaction with dynamin 2 and cause autosomal recessive centronuclear myopathy', *Nature Genetics*, 39(9), pp. 1134–1139. doi: 10.1038/ng2086.

Nord, H. *et al.* (2014) 'Differential regulation of myosin heavy chains defines new muscle domains in zebrafish', *Molecular Biology of the Cell*, 25(8), pp. 1384–1395. doi: 10.1091/mbc.E13-08-0486.

Nowak, K. J., Ravenscroft, G. and Laing, N. G. (2013) 'Skeletal muscle α-actin diseases (actinopathies): pathology and mechanisms', *Acta Neuropathologica*, 125(1), pp. 19–32. doi: 10.1007/s00401-012-1019-z.

Oda, T. *et al.* (2015) 'A de novo mutation of the MYH7 gene in a large Chinese family with autosomal dominant myopathy', *Human Genome Variation*, 2(1), p. 15022. doi: 10.1038/hgv.2015.22.

Ojima, K. *et al.* (2015) 'The importance of subfragment 2 and C-terminus of myosin heavy chain for thick filament assembly in skeletal muscle cells', *Animal Science Journal*, 86(4), pp. 459–467. doi: 10.1111/asj.12310.

Orfanos, Z. and Sparrow, J. C. (2013) 'Myosin isoform switching during assembly of the drosophila flight muscle thick filament lattice', *Journal of Cell Science*, 126(1), pp. 139–148. doi: 10.1242/jcs.110361.

Ortolano, S. *et al.* (2011) 'A novel MYH7 mutation links congenital fiber type disproportion and myosin storage myopathy', *Neuromuscular Disorders*, 21(4), pp. 254–262. doi: 10.1016/j.nmd.2010.12.011.

Overeem, S. *et al.* (2007) 'Symptomatic distal myopathy with cardiomyopathy due to a MYH7 mutation', *Neuromuscular Disorders*, 17(6), pp. 490–493. doi: 10.1016/j.nmd.2007.02.007.

Pardo, J. V., Siliciano, J. D. and Craig, S. W. (1983) 'A vinculin-containing cortical lattice in skeletal muscle: transverse lattice elements ("costameres") mark sites of attachment between myofibrils and sarcolemma.', *Proceedings of the National Academy of Sciences*, 80(4), pp. 1008–1012. doi: 10.1073/pnas.80.4.1008.

Park, J. S. *et al.* (2009) 'Isolation of a ventricle-specific promoter for the zebrafish ventricular myosin heavy chain (vmhc) gene and its regulation by GATA factors during embryonic heart development', *Developmental Dynamics*, 238(6), pp. 1574–1581. doi: 10.1002/dvdy.21964.

Parker, F. and Peckham, M. (2020) 'Disease mutations in striated muscle myosins', *Biophysical Reviews*, 12(4), pp. 887–894. doi: 10.1007/s12551-020-00721-5.

Péault, B. *et al.* (2007) 'Stem and progenitor cells in skeletal muscle development, maintenance, and therapy', *Molecular Therapy*, 15(5), pp. 867–877. doi: 10.1038/mt.sj.6300145.

Pegoraro, E. *et al.* (2007) 'MYH7 gene mutation in myosin storage myopathy and scapulo-peroneal myopathy', *Neuromuscular Disorders*, 17(4), pp. 321–329. doi: 10.1016/j.nmd.2007.01.010.

Pipalia, T. G. *et al.* (2016) 'Cellular dynamics of regeneration reveals role of two distinct Pax7 stem cell populations in larval zebrafish muscle repair', *DMM Disease Models and Mechanisms*, 9(6), pp. 671–684. doi: 10.1242/dmm.022251.

Quach, N. L. and Rando, T. A. (2006) 'Focal adhesion kinase is essential for costamerogenesis in cultured skeletal muscle cells', *Developmental Biology*, 293(1), pp. 38–52. doi: 10.1016/j.ydbio.2005.12.040.

van Raamsdonk, W. *et al.* (1982) 'Differentiation of muscle fiber types in the teleost Brachydanio rerio, the zebrafish - Posthatching development', *Anatomy and Embryology*, 164(1), pp. 51–62. doi: 10.1007/BF00301878.

Rahmani, H. *et al.* (2021) 'The myosin II coiled-coil domain atomic structure in its native environment', *Proceedings of the National Academy of Sciences of the United States of America*, 118(14). doi: 10.1073/pnas.2024151118.

Ran, F. A. *et al.* (2013) 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols*, 8(11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.

Ravenscroft, G. *et al.* (2017) 'New era in genetics of early-onset muscle disease: Breakthroughs and challenges', *Seminars in Cell and Developmental Biology*, 64, pp. 160–170. doi:

10.1016/j.semcdb.2016.08.002.

Rayment, I. and Holden, H. M. (1994) 'The three-dimensional structure of a molecular motor', *Trends in Biochemical Sciences*, 19(3), pp. 129–134. doi: 10.1016/0968-0004(94)90206-2.

Relaix, F. *et al.* (2005) 'A Pax3/Pax7-dependent population of skeletal muscle progenitor cells', *Nature*, 435(7044), pp. 948–953. doi: 10.1038/nature03594.

Rhee, D., Sanger, J. M. J. W. and Sanger, J. M. J. W. (1994) 'The premyofibril: Evidence for its role in myofibrillogenesis', *Cell Motility and the Cytoskeleton*, 28(1), pp. 1–24. doi: 10.1002/cm.970280102.

Robert-Paganin, J., Auguin, D. and Houdusse, A. (2018) 'Hypertrophic cardiomyopathy disease results from disparate impairments of cardiac myosin function and auto-inhibition', *Nature Communications*, 9(1). doi: 10.1038/s41467-018-06191-4.

Robinson, R. *et al.* (2006) 'Mutations in RYR1 in malignant hyperthermia and central core disease', *Human Mutation*, 27(10), pp. 977–989. doi: 10.1002/humu.20356.

Romero, N. B., Sandaradura, S. A. and Clarke, N. F. (2013) 'Recent advances in nemaline myopathy', *Current Opinion in Neurology*, 26(5), pp. 519–526. doi: 10.1097/WCO.0b013e328364d681.

Rossi, A. C. *et al.* (2010) 'Two novel/ancient myosins in mammalian skeletal muscles: MYH14/7b and MYH15 are expressed in extraocular muscles and muscle spindles', *Journal of Physiology*, 588(2), pp. 353–364. doi: 10.1113/jphysiol.2009.181008.

Roy, S. D. *et al.* (2017) 'Myotome adaptability confers developmental robustness to somitic myogenesis in response to fibre number alteration', *Developmental Biology*, 431(2), pp. 321–335. doi: 10.1016/j.ydbio.2017.08.029.

Rui, Y., Bai, J. and Perrimon, N. (2010) 'Sarcomere formation occurs by the assembly of multiple latent protein complexes', *PLoS Genetics*, 6(11). doi: 10.1371/journal.pgen.1001208.

Rushbrook, J. I. *et al.* (1998) 'Protein and mRNA analysis of myosin heavy chains in the developing avian pectoralis major muscle', *Journal of Muscle Research and Cell Motility*, 19(2), pp. 157–168. doi: 10.1023/A:1005360612542.

Sanger, J. W. *et al.* (2005) 'How to build a myofibril', *Journal of Muscle Research and Cell Motility*, 26(6–8), pp. 343–354. doi: 10.1007/s10974-005-9016-7.

Sanger, J. W. J. M. *et al.* (2009) 'Myofibrillogenesis in skeletal muscle cells in zebrafish', *Cell Motility and the Cytoskeleton*, 66(8), pp. 556–566. doi: 10.1002/cm.20365.

Sarelius, I. H. and Duling, B. R. (1982) 'Direct measurement of microvessel hematocrit, red cell flux, velocity, and transit time', *American Journal of Physiology - Heart and Circulatory Physiology*, 12(6). doi: 10.1152/ajpheart.1982.243.6.h1018.

Sarkar, S. S. *et al.* (2020) 'The hypertrophic cardiomyopathy mutations R403Q and R663H increase the

number of myosin heads available to interact with actin', *Science Advances*, 6(14), pp. 1–11. doi: 10.1126/sciadv.aax0069.

Schiaffino, S. *et al.* (1989) 'Three myosin heavy chain isoforms in type 2 skeletal muscle fibres', *Journal of Muscle Research and Cell Motility*, 10(3), pp. 197–205. doi: 10.1007/BF01739810.

Schiaffino, S. *et al.* (2015) 'Developmental myosins: Expression patterns and functional significance', *Skeletal Muscle*, 5(1), pp. 1–14. doi: 10.1186/s13395-015-0046-6.

Schiaffino, S. and Reggiani, C. (2011) 'Fiber types in Mammalian skeletal muscles', *Physiological Reviews*, 91(4), pp. 1447–1531. doi: 10.1152/physrev.00031.2010.

Schoenauer, R. *et al.* (2008) 'Myomesin 3, a Novel Structural Component of the M-band in Striated Muscle', *Journal of Molecular Biology*, 376(2), pp. 338–351. doi: 10.1016/j.jmb.2007.11.048.

Schoenauer, R. *et al.* (2011) 'EH-myomesin splice isoform is a novel marker for dilated cardiomyopathy', *Basic Research in Cardiology*, 106(2), pp. 233–247. doi: 10.1007/s00395-010-0131-2.

Schröter, C. *et al.* (2008) 'Dynamics of zebrafish somitogenesis', *Developmental Dynamics*, 237(3), pp. 545–553. doi: 10.1002/dvdy.21458.

Schwander, M. *et al.* (2003) 'B1 Integrins Regulate Myoblast Fusion and Sarcomere Assembly', *Developmental Cell*, 4(5), pp. 673–685. doi: 10.1016/S1534-5807(03)00118-7.

Sharma, A. *et al.* (2018) 'Myosin Heavy Chain-embryonic is a crucial regulator of skeletal muscle development and differentiation', *bioRxiv*, p. 261685. Available at: https://www.biorxiv.org/content/10.1101/261685v1.full.

Shimizu, T. *et al.* (1985) 'Axial arrangement of the myosin rod in vertebrate thick filaments: Immunoelectron microscopy with a monoclonal antibody to light meromyosin', *Journal of Cell Biology*, 101(3), pp. 1115–1123. doi: 10.1083/jcb.101.3.1115.

Shingde, M. V. *et al.* (2006) 'Myosin storage (hyaline body) myopathy: A case report', *Neuromuscular Disorders*, 16(12), pp. 882–886. doi: 10.1016/j.nmd.2006.09.001.

Siegert, R. *et al.* (2011) 'A myomesin mutation associated with hypertrophic cardiomyopathy deteriorates dimerisation properties', *Biochemical and Biophysical Research Communications*, 405(3), pp. 473–479. doi: 10.1016/j.bbrc.2011.01.056.

Sobrido, M. J. *et al.* (2005) 'Autosomal dominant congenital fibre type disproportion: A clinicopathological and imaging study of a large family', *Brain*, 128(7), pp. 1716–1727. doi: 10.1093/brain/awh511.

Sodek, J. *et al.* (1972) 'Amino-acid sequence of rabbit skeletal tropomyosin and its coiled-coil structure.', *Proceedings of the National Academy of Sciences of the United States of America*, 69(12),

pp. 3800–3804. doi: 10.1073/pnas.69.12.3800.

Sohn, R. L. *et al.* (1997) 'A 29 residue region of the sarcomeric myosin rod is necessary for filament formation', *Journal of Molecular Biology*, 266(2), pp. 317–330. doi: 10.1006/jmbi.1996.0790.

Sorrells, S. *et al.* (2013) 'Analysis of apoptosis in zebrafish embryos by whole-mount immunofluorescence to detect activated caspase 3', *Journal of Visualized Experiments*, (82), pp. 1–8. doi: 10.3791/51060.

Spertus, J. A. *et al.* (2021) 'Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): health status analysis of a randomised, double-blind, placebo-controlled, phase 3 trial', *The Lancet*, 397(10293), pp. 2467–2475. doi: 10.1016/S0140-6736(21)00763-7.

Spudich, J. A. (2015) 'The myosin mesa and a possible unifying hypothesis for the molecular basis of human hypertrophic cardiomyopathy', *Biochemical Society Transactions*, 43(1), pp. 64–72. doi: 10.1042/BST20140324.

Spudich, J. A. (2019) 'Three perspectives on the molecular basis of hypercontractility caused by hypertrophic cardiomyopathy mutations', *Pflugers Archiv European Journal of Physiology*, 471(5), pp. 701–717. doi: 10.1007/s00424-019-02259-2.

Squire, J. M. (1973) 'General model of myosin filament structure', *Journal of Molecular Biology*, 77(2), pp. 291–323. doi: 10.1016/0022-2836(73)90337-9.

Stalpers, X. *et al.* (2011) 'Scoliosis surgery in a patient with "de novo" myosin storage myopathy', *Neuromuscular Disorders*, 21(11), pp. 812–815. doi: 10.1016/j.nmd.2011.05.005.

Stelzer, J. E., Fitzsimons, D. P. and Moss, R. L. (2006) 'Ablation of myosin-binding protein-C accelerates force development in mouse myocardium', *Biophysical Journal*, 90(11), pp. 4119–4127. doi: 10.1529/biophysj.105.078147.

Stewart, M. A. *et al.* (2010) 'Myosin ATP turnover rate is a mechanism involved in thermogenesis in resting skeletal muscle fibers', *Proceedings of the National Academy of Sciences of the United States of America*, 107(1), pp. 430–435. doi: 10.1073/pnas.0909468107.

Stickney, H. L., Barresi, M. J. F. and Devoto, S. H. (2000) 'Somite development in Zebrafish', *Developmental Dynamics*, 219(3), pp. 287–303. doi: 10.1002/1097-0177(2000)9999:9999<::AID-DVDY1065>3.0.CO;2-A.

Suggs, J. A. *et al.* (2017) 'A Drosophila model of dominant inclusion body myopathy type 3 shows diminished myosin kinetics that reduce muscle power and yield myofibrillar defects', *DMM Disease Models and Mechanisms*, 10(6), pp. 761–770. doi: 10.1242/dmm.028050.

Summerbell, D., Halai, C. and Rigby, P. W. J. (2002) 'Expression of the myogenic regulatory factor Mrf4

precedes or is contemporaneous with that of Myf5 in the somitic bud', *Mechanisms of Development*, 117(1–2), pp. 331–335. doi: 10.1016/S0925-4773(02)00208-3.

Sundaram, C. and Megha, S. U. (2012) 'Approach to the Interpretation of Muscle Biopsy', in *Muscle Biopsy*. InTech. doi: 10.5772/39034.

Swailes, N. T. *et al.* (2006) 'Non-muscle myosins 2A and 2B drive changes in cell morphology that occur as myoblasts align and fuse', *Journal of Cell Science*, 119(17), pp. 3561–3570. doi: 10.1242/jcs.03096.

Sweeney, H. L. and Houdusse, A. (2010) 'Structural and Functional Insights into the Myosin Motor Mechanism', *Annual Review of Biophysics*, 39(1), pp. 539–557. doi: 10.1146/annurev.biophys.050708.133751.

Tajbakhsh, S. *et al.* (1997) 'Redefining the genetic hierarchies controlling skeletal myogenesis: Pax- 3 and Myf-5 act upstream of MyoD', *Cell*, 89(1), pp. 127–138. doi: 10.1016/S0092-8674(00)80189-0.

Tajsharghi, H. *et al.* (2003) 'Myosin storage myopathy associated with a heterozygous missense mutation in MYH7', *Annals of Neurology*, 54(4), pp. 494–500. doi: 10.1002/ana.10693.

Tajsharghi, H. *et al.* (2007) 'Homozygous mutation in MYH7 in myosin storage myopathy and cardiomyopathy', *Neurology*, 68(12), p. 962. doi: 10.1212/01.wnl.0000257131.13438.2c.

Tajsharghi, H. and Oldfors, A. (2013) 'Myosinopathies: Pathology and mechanisms', *Acta Neuropathologica*, 125(1), pp. 3–18. doi: 10.1007/s00401-012-1024-2.

Tajsharghi, H., Pilon, M. and Oldfors, A. (2005) 'A Caenorhabditis elegans model of the myosin heavy chain IIa E706R mutation', *Annals of Neurology*, 58(3), pp. 442–448. doi: 10.1002/ana.20594.

Tasca, G. *et al.* (2012) 'New phenotype and pathology features in MYH7-related distal myopathy', *Neuromuscular Disorders*, 22(7), pp. 640–647. doi: 10.1016/j.nmd.2012.03.003.

Taylor, J. S. *et al.* (2001) 'Comparative genomics provides evidence for an ancient genome duplication event in fish', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1414), pp. 1661–1679. doi: 10.1098/rstb.2001.0975.

Thisse, C. and Thisse, B. (2008) 'High-resolution in situ hybridization to whole-mount zebrafish embryos', *Nature Protocols*, 3(1), pp. 59–69. doi: 10.1038/nprot.2007.514.

Thomas, P. K., Schott, G. D. and Morgan-Hughes, J. A. (1975) 'Adult onset scapuloperoneal myopathy', *Journal of Neurology, Neurosurgery and Psychiatry*, 38(10), pp. 1008–1015. doi: 10.1136/jnnp.38.10.1008.

Thompson, R. C. *et al.* (2012) 'Myosin filament assembly requires a cluster of four positive residues located in the rod domain', *FEBS Letters*, 586(19), pp. 3008–3012. doi: 10.1016/j.febslet.2012.06.019.

Tidyman, W. E., Moore, L. A. and Bandman, E. (1997) 'Expression of fast myosin heavy chain transcripts in developing and dystrophic chicken skeletal muscle', *Developmental Dynamics*, 208(4), pp. 491–504.

doi: 10.1002/(SICI)1097-0177(199704)208:4<491::AID-AJA5>3.0.CO;2-D.

Toepfer, C. N. *et al.* (2018) *MYBPC3 Mutations cause Hypertrophic Cardiomyopathy by Dysregulating Myosin: Implications for Therapy*. doi: 10.1097/01.mbc.0000224842.25592.8a.

Toepfer, Christopher N. *et al.* (2019) 'Hypertrophic cardiomyopathy mutations in MYBPC3 dysregulate myosin', *Science Translational Medicine*, 11(476), pp. 1–11. doi: 10.1126/scitranslmed.aat1199.

Toepfer, Christopher N *et al.* (2019) 'Hypertrophic cardiomyopathy mutations in MYBPC3 dysregulate myosin', *Science Translational Medicine*, 11(476). doi: 10.1126/scitranslmed.aat1199.

Toepfer, C. N. *et al.* (2020) 'Myosin Sequestration Regulates Sarcomere Function, Cardiomyocyte Energetics, and Metabolism, Informing the Pathogenesis of Hypertrophic Cardiomyopathy', *Circulation*, pp. 828–842. doi: 10.1161/CIRCULATIONAHA.119.042339.

Tokuyasu, K. T. (1989) 'Immunocytochemical studies of cardiac myofibrillogenesis in early chick embryos. III. Generation of fasciae adherentes and costameres', *Journal of Cell Biology*, 108(1), pp. 43–53. doi: 10.1083/jcb.108.1.43.

Tokuyasu, K. T. and Maher, P. A. (1987) 'Immunocytochemical studies of cardiac myofibrillogenesis in early chick embryos. II. Generation of α-actinin dots within titin spots at the time of the first myofibril formation', *Journal of Cell Biology*, 105(6 I), pp. 2795–2801. doi: 10.1083/jcb.105.6.2795.

Tonino, P. *et al.* (2019) 'Fine mapping titin's C-zone: Matching cardiac myosin-binding protein C stripes with titin's super-repeats', *Journal of Molecular and Cellular Cardiology*, 133(April), pp. 47–56. doi: 10.1016/j.yjmcc.2019.05.026.

Topaloglu, H. (2020) 'Core myopathies - A short review', *Acta Myologica*, 39(4), pp. 266–273. doi: 10.36185/2532-1900-029.

Tosch, V. *et al.* (2006) 'A novel PtdIns3P and PtdIns(3,5)P2 phosphatase with an inactivating variant in centronuclear myopathy', *Human Molecular Genetics*, 15(21), pp. 3098–3106. doi: 10.1093/hmg/ddl250.

Trivedi, D. V. *et al.* (2018) 'Hypertrophic cardiomyopathy and the myosin mesa: viewing an old disease in a new light', *Biophysical Reviews*, 10(1), pp. 27–48. doi: 10.1007/s12551-017-0274-6.

Uro-Coste, E. *et al.* (2009) 'Striking phenotypic variability in two familial cases of myosin storage myopathy with a MYH7 Leu1793pro mutation', *Neuromuscular Disorders*, 19(2), pp. 163–166. doi: 10.1016/j.nmd.2008.11.012.

Van Der Ven, P. F. M. *et al.* (1999) 'Thick filament assembly occurs after the formation of a cytoskeletal scaffold', *Journal of Muscle Research and Cell Motility*, 20(5–6), pp. 569–579. doi: 10.1023/A:1005569225773.

Viswanathan, M. C. *et al.* (2017) 'Myosin storage myopathy mutations yield defective myosin filament

assembly in vitro and disrupted myofibrillar structure and function in vivo', *Human Molecular Genetics*, 26(24), pp. 4799–4813. doi: 10.1093/hmg/ddx359.

Volk, T., Fessler, L. I. and Fessler, J. H. (1990) 'A role for integrin in the formation of sarcomeric cytoarchitecture', *Cell*, 63(3), pp. 525–536. doi: 10.1016/0092-8674(90)90449-O.

Volkmann, N. *et al.* (2000) 'Evidence for cleft closure in actomyosin upon ADP release', *Nature Structural Biology*, 7(12), pp. 1147–1155. doi: 10.1038/82008.

Walsh, R. *et al.* (2009) 'Cardiomyopathy: A systematic review of disease-causing mutations in myosin heavy chain 7 and their phenotypic manifestations', *Cardiology*, 115(1), pp. 49–60. doi: 10.1159/000252808.

Watabe, S. and Ikeda, D. (2006) 'Diversity of the pufferfish Takifugu rubripes fast skeletal myosin heavy chain genes', *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, 1(1 SPEC. ISS.), pp. 28–34. doi: 10.1016/j.cbd.2005.12.001.

Waterman, R. E. (1969) 'Development of the lateral musculature in the teleost, Brachydanio rerio: A fine structural study', *American Journal of Anatomy*, 125(4), pp. 457–493. doi: 10.1002/aja.1001250406.

Waterston, R. H., Fishpool, R. M. and Brenner, S. (1977) 'Mutants affecting paramyosin in Caenorhabditis elegans', *Journal of Molecular Biology*, 117(3), pp. 679–697. doi: 10.1016/0022-2836(77)90064-X.

Waterston, R. H. and Francis, G. R. (1985) 'Genetic analysis of muscle development in Caenorhabditis elegans', *Trends in Neurosciences*, 8(C), pp. 270–276. doi: 10.1016/0166-2236(85)90101-8.

Webster, C. *et al.* (1988) 'Fast muscle fibers are preferentially affected in Duchenne muscular dystrophy', *Cell*, 52(4), pp. 503–513. doi: 10.1016/0092-8674(88)90463-1.

Weiss, A. *et al.* (1999) 'Organization of human and mouse skeletal myosin heavy chain gene clusters is highly conserved', *Proceedings of the National Academy of Sciences of the United States of America*, 96(6), pp. 2958–2963. doi: 10.1073/pnas.96.6.2958.

Whalen, R. G. *et al.* (1981) 'Three myosin heavy-chain isozymes appear sequentially in rat muscle development', *Nature*, 292(5826), pp. 805–809. doi: 10.1038/292805a0.

Whittle, J. *et al.* (2020) ' MYH 3-associated distal arthrogryposis zebrafish model is normalized with para-aminoblebbistatin ', *EMBO Molecular Medicine*, 12(11), pp. 1–14. doi: 10.15252/emmm.202012356.

Wilson, C. *et al.* (2014) 'The myosin inhibitor blebbistatin stabilizes the super-relaxed state in skeletal muscle', *Biophysical Journal*, 107(7), pp. 1637–1646. doi: 10.1016/j.bpj.2014.07.075.

Wolff, C. *et al.* (2004) 'iguana encodes a novel zinc-finger protein with coiled-coil domains essential

for Hedgehog signal transduction in the zebrafish embryo', *Genes and Development*, 18(13), pp. 1565–1576. doi: 10.1101/gad.296004.

Woodhead, J. L. *et al.* (2005) 'Atomic model of a myosin filament in the relaxed state', *Nature*, 436(7054), pp. 1195–1199. doi: 10.1038/nature03920.

Woodhead, J. L. and Craig, R. (2020) 'The mesa trail and the interacting heads motif of myosin II', *Archives of Biochemistry and Biophysics*, 680(October 2019), pp. 12–14. doi: 10.1016/j.abb.2019.108228.

Xu, J. *et al.* (2012) 'Functional Analysis of Slow Myosin Heavy Chain 1 and Myomesin-3 in Sarcomere Organization in Zebrafish Embryonic Slow Muscles', *Journal of Genetics and Genomics*, 39(2), pp. 69–80. doi: 10.1016/j.jgg.2012.01.005.

Yamauchi-Takihara, K. *et al.* (1989) 'Characterization of human cardiac myosin heavy chain genes.', *Proceedings of the National Academy of Sciences*, 86(10), pp. 3504–3508. doi: 10.1073/pnas.86.10.3504.

Yang, Y. G., Obinata, T. and Shimada, Y. (2000) 'Developmental relationship of myosin binding proteins (myomesin, connectin and C-protein) to myosin in chicken somites as studied by immunofluorescence microscopy', *Cell Structure and Function*, 25(3), pp. 177–185. doi: 10.1247/csf.25.177.

Yengo, C. M. *et al.* (1999) 'Intrinsic tryptophan fluorescence identifies specific conformational changes at the actomyosin interface upon actin binding and ADP release', *Biochemistry*, 38(44), pp. 14515–14523. doi: 10.1021/bi991226l.

Yount, R. G. *et al.* (1995) 'Is myosin a "back door" enzyme?', *Biophysical Journal*, 68(4 SUPPL.), pp. 243–246.

Yutzey, K. E., Rhee, J. T. and Bader, D. (1994) 'Expression of the atrial-specific myosin heavy chain AMHC1 and the establishment of anteroposterior polarity in the developing chicken heart', *Development*, 120(4), pp. 871–883. doi: 10.1242/dev.120.4.871.

Zhang, R. and Xu, X. (2009) 'Transient and Transgenic Analysis of the Zebrafish Ventricular Myosin Heavy Chain (vmhc) Promoter: An Inhibitory Mechanism of Ventricle-Specific Gene Expression', *Developmental Dynamics*, 238(6), pp. 1564–1573. doi: 10.1002/dvdy.21929.Transient.

Zhang, S. and Bernstein, S. I. (2001) 'Spatially and temporally regulated expression of myosin heavy chain alternative exons during Drosophila embryogenesis', *Mechanisms of Development*, 101(1–2), pp. 35–45. doi: 10.1016/S0925-4773(00)00549-9.

Zhang, Y. *et al.* (2017a) 'Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system', *Nature Communications*, 8(1), pp. 6–10. doi: 10.1038/s41467-017-00175-6.

Zhang, Y. *et al.* (2017b) 'Programmable base editing of zebrafish genome using a modified CRISPR-

Cas9 system', *Nature Communications*, 8(1), pp. 6–10. doi: 10.1038/s41467-017-00175-6.

Zhao, F. Q., Padrón, R. and Craig, R. (2008) 'Blebbistatin stabilizes the helical order of myosin filaments by promoting the switch 2 closed state', *Biophysical Journal*, 95(7), pp. 3322–3329. doi: 10.1529/biophysj.108.137067.

Zoghbi, M. E. *et al.* (2008) 'Three-dimensional structure of vertebrate cardiac muscle myosin filaments', *Proceedings of the National Academy of Sciences of the United States of America*, 105(7), pp. 2386–2390. doi: 10.1073/pnas.0708912105.

Appendix

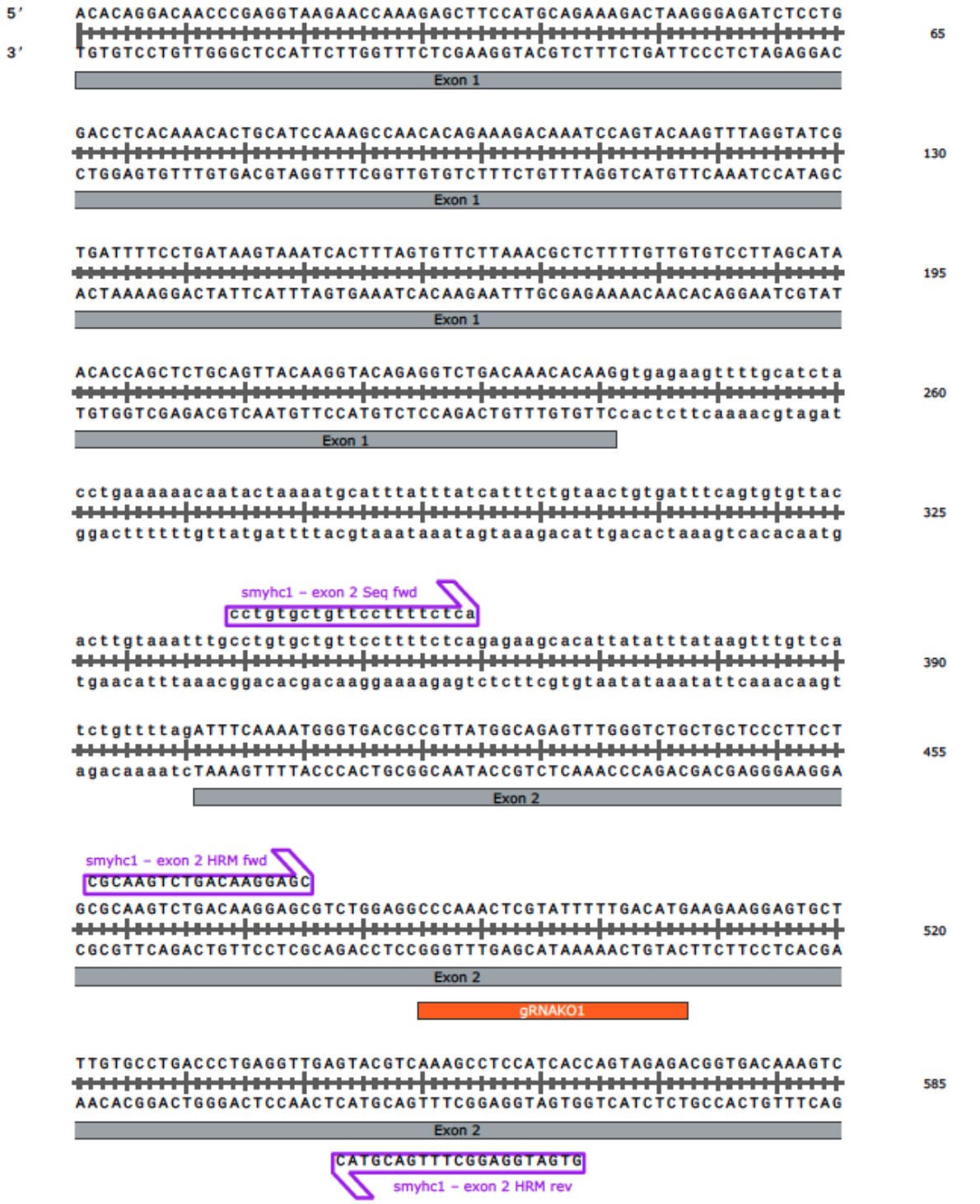
Appendix 1.1 – Literature Review of LDM and MSM

Base Change	Mutation Type	Amino Acid Location	Amino Acid Change	Amino acid position in heptad repeat	Heterozygosity	Region of Protein	Disease Type	Number of Affected Patients	In Vivo Phenotype												Reference
									Early Onset Disease (<25y)	Distal Lower Limb Myopathy	Distal Upper Limb Myopathy	Proximal Myopathy	Delayed Motor Milestones	Scapular Involvement	HyperCKemia	Axial Involvement	Respiratory Involvement	Abnormal Biopsy Findings	Cardiac Involvement	Total Score (/11)	
c.4301G>C	Missense point mutation	1434	p.R1434P	c	Heterozygous	LMM	Laing distal myopathy	3	100%	100%	100%	100%		33%		33%				4.67	Feinstein et al 2016[67]
c.4304T>C	Missense point mutation	1435	p.S1435P	d		LMM	Laing distal myopathy	3	0%	100%	0%	67%			33%	33%	0%	67%	33%	3.33	Fiorillo et al 2016[46]
c.4309G>C	Missense point mutation	1437	p.A1437P	e		LMM	Laing distal myopathy	1	100%	100%	0%	0%				100%	100%	100%	0%	5	Dabaj et al 2018[68]
c.4309G>C	Missense point mutation	1437	p.A1437P	e	Heterozygous	LMM	Laing distal myopathy	5	100%	100%		100%				100%				4	Feinstein et al 2016[67]
c.4315G>C	Missense point mutation	1439	p.A1439P	a	Heterozygous	LMM	Laing distal myopathy	3	100%	100%	100%	100%			100%			100%		6	Park et al 2013[69]
c.4358T>C	Missense point mutation	1453	p.L1453P	a	Heterozygous	LMM	Laing distal myopathy	6	100%	100%	67%	50%	50%	17%		17%	33%		17%	4.5	Lefter et al 2015[70]
c.4442T>C	Missense point mutation	1481	p.L1481P	a		LMM	Laing distal myopathy	6	100%	100%	100%	0%		100%		100%	100%	100%	0%	7	Lamont et al 2014[72]
c.4475T>C	Missense point mutation	1492		e		LMM	Laing distal myopathy	1	100%	100%	100%	0%				100%	100%	100%	0%	6	Dabaj et al 2018[68]
		1500	p.R1500P	f	Heterozygous	LMM	Laing distal myopathy	1	100%	100%	100%	100%	0%	100%	0%				0%	5	Meredith et al 2004[73]
c.4522_4524delGAG	Tricnucleotide deletion	1508	p.E1508del	g	Heterozygous	LMM	Laing distal myopathy	8	100%	100%	100%	100%			0%	0%		100%	0%	5	Reis et al 2015[74]
c.4522_4524delGAG	Tricnucleotide deletion	1508	p.E1508del	g	Heterozygous	LMM	Laing distal myopathy	2	100%	100%	50%	100%			0%	50%	0%	100%	0%	5	van den Bergh et al 2014[71]
c.4522_4524delGAG	Tricnucleotide deletion	1508	p.E1508del	g		LMM	Laing distal myopathy	1	100%	0%	100%	100%		0%		100%	100%	100%	100%	7	Lamont et al 2014[72]
c.4522_4524delGAG	Tricnucleotide deletion	1508	p.E1508del	g		LMM	Laing distal myopathy	2	100%	100%	100%	100%		0%		100%	0%	100%	0%	6	Lamont et al 2014[72]
c.4522_4524delGAG	Tricnucleotide deletion	1508	p.E1508del	g	Heterozygous	LMM	Laing distal myopathy	5	100%	100%	60%	20%			100%	60%		100%	100%	6.4	Dubourg et al 2011[76]
c.4622A>C	Missense point mutation	1541	p.Q1541P	e		LMM	Laing distal myopathy	8	100%	100%	100%	0%		0%				100%		4	Lamont et al 2014[72]
c.4645G>C	Missense point mutation	1549	p.A1549P	f	Heterozygous	LMM	Laing distal myopathy	4	100%	100%	75%	67%			25%			25%	0%	3.92	Ferbert et al 2017[77]
c.4679G>C	Missense point mutation	1560	p.R1560P	c	Heterozygous	LMM	Laing distal myopathy	16	79%	75%	19%				0%	0%		100%	0%	2.72	Carbonell-Corvillo et al 201
c.4772T>C	Missense point mutation	1591	p.L1591P	e		LMM	Laing distal myopathy	3	100%	100%	100%	100%			100%	100%		100%	0%	7	Tasca et al 2012[12]
c.4790T>G	Missense point mutation	1597	p.L1597R	d	Heterozygous	LMM	Laing distal myopathy	1	100%	100%	100%	100%			100%		100%	100%	0%	7	Clarke et al 2013[79]
c.4790T>G	Missense point mutation	1597	p.L1597R	d	Heterozygous	LMM	Laing distal myopathy	1	100%	100%		100%			100%	100%	100%	100%	0%	7	Clarke et al 2013[79]
c.4795A>C	Missense point mutation	1599	p.T1599P	f		LMM	Laing distal myopathy	6	100%	100%	100%	0%		0%						3	Lamont et al 2014[72]
c.4802T>C	Missense point mutation	1601		a		LMM	Laing distal myopathy	1	100%	100%	100%	100%				100%	0%	100%	0%	6	Dabaj et al 2018[68]
c.4807G>C	Missense point mutation	1603	p.A1603P	c		LMM	Laing distal myopathy	2	100%	100%	100%	100%				100%	50%	100%	0%	6.5	Dabaj et al 2018[68]
c.4807G>C	Missense point mutation	1603	p.A1603P	c		LMM	Laing distal myopathy	4	100%	100%	100%	0%			0%	100%	0%	50%	0%	4.5	Fiorillo et al 2016[46]
c.4823G>C	Missense point mutation	1608	p.R1608P	a		LMM	Laing distal myopathy	2	100%	100%	100%	100%		0%		100%	100%	100%	100%	8	Lamont et al 2014[72]
c.4814G>C	Missense point mutation	1611	p.A1611D	d		LMM	Laing distal myopathy	1	100%	100%	100%	100%				100%	0%	100%	0%	6	Dabaj et al 2018[68]
c.4835T>C	Missense point mutation	1612	p.L1612P	e		LMM	Laing distal myopathy	1	100%	100%	0%	100%		0%		100%	0%		0%	4	Lamont et al 2014[72]
c.4849_4851delAAG	Tricnucleotide deletion	1617	p.K1617del	c	Heterozygous	LMM	Laing distal myopathy	14	100%	100%	92%	42%			0%	50%		17%	0%	4	Oda et al 2015[80]
c.4849_4851delAAG	Tricnucleotide deletion	1617	p.K1617del	c		LMM	Laing distal myopathy	9	100%	100%	100%	100%		0%		100%	0%	100%	0%	6	Lamont et al 2014[72]

Base Change	Mutation Type	Amino Acid Location	Amino Acid Change	Amino acid position in heptad repeat	Heterozygosity	Region of Protein	Disease Type	Number of Affected Patients	In Vivo Phenotype												Reference
									Early Onset Disease (<25y)	Distal Lower Limb Myopathy	Distal Upper Limb Myopathy	Proximal Myopathy	Delayed Motor Milestones	Scapular Involvement	HyperCKemia	Axial Involvement	Respiratory Involvement	Abnormal Biopsy Findings	Cardiac Involvement	Total Score (/11)	
c.4849_4851delAAG	Tricnucleotide deletion	1617	p.K1617del	c		LMM	Laing distal myopathy	6	100%	100%	100%	100%		0%		100%	0%	100%	0%	6	Lamont et al 2014[72]
c.4849_4851delAAG	Tricnucleotide deletion	1617	p.K1617del	c		LMM	Laing distal myopathy	1	100%	100%	100%	100%		0%				100%		5	Lamont et al 2014[72]
c.4849_4851delAAG	Tricnucleotide deletion	1617	p.K1617del	c		LMM	Laing distal myopathy	1	100%	100%	0%	100%		0%		0%	100%	100%	0%	5	Lamont et al 2014[72]
c.4849_4851delAAG	Tricnucleotide deletion	1617	p.K1617del	c		LMM	Laing distal myopathy	1	100%	100%	0%	100%		0%		100%	100%	100%	100%	7	Lamont et al 2014[72]
c.4849_4851delAAG	Tricnucleotide deletion	1617	p.K1617del	c	Heterozygous	LMM	Laing distal myopathy	4	100%	100%	100%	100%				100%		100%	0%	6	Komlosi et al 2014[81]
c.4906G>C	Missense point mutation	1636	p.A1636P	a		LMM	Laing distal myopathy	20	100%	100%	100%	100%		0%		100%	0%	100%	0%	6	Lamont et al 2014[72]
c.4937T>C	Missense point mutation	1646	p.L1646P	d		LMM	Laing distal myopathy	6	100%	100%	100%	0%		100%		100%	0%		0%	5	Lamont et al 2014[72]
c.4985G>C	Missense point mutation	1662	p.R1662P	f		LMM	Laing distal myopathy	3	100%	100%	0%	0%		0%						2	Lamont et al 2014[72]
c.5005_5007delGAG	Tricnucleotide deletion	1669	p.E1669del	f		LMM	Laing distal myopathy	7	100%	100%	100%	100%		0%		100%	0%		0%	5	Lamont et al 2014[72]
c.5059-5061del	Tricnucleotide deletion	1687	p.E1687del	d	Heterozygous	LMM	Laing distal myopathy	8	100%	100%	0%	0%		100%	0%	100%	0%	100%	0%	5	Li et al 2018[82]
		1706	p.L1706P	a	Heterozygous	LMM	Laing distal myopathy	1	100%	100%		100%								3	Meredith et al 2004[73]
c.5186_5188delAGA	Tricnucleotide deletion	1728	p.K1729del	c	Heterozygous	LMM	Laing distal myopathy	8	100%	100%	80%			20%	0%				0%	3	Roda et al 2014[84]
c.5186_5188dupAGA	Trinucleotide duplication	1729	p.K1729dup	c		LMM	Laing distal myopathy	3	100%	100%	0%	100%		0%		0%	0%	100%	0%	4	Lamont et al 2014[72]
c.5186_5188dupAGA	Trinucleotide duplication	1729	p.K1729dup	c		LMM	Laing distal myopathy	32	50%	100%	100%	100%						100%	3%	4.53	Udd et al 2009[85]
c.5352_5354delGAA	Tricnucleotide deletion	1784	p.K1784del	b		LMM	Laing distal myopathy	3	0%	100%		100%			0%				0%	2	Tasca et al 2012[12]
c.5378_5380delTGC	Tricnucleotide deletion	1793	p.L1793del	d		LMM	Laing distal myopathy	1	100%	100%	100%	100%		0%		100%	0%		100%	6	Lamont et al 2014[72]
c.5401G>A	Missense point mutation	1801	p.E1801K	e		LMM	Laing distal myopathy	4	75%	75%	0%	75%			100%	25%	0%	50%	100%	5	Fiorillo et al 2016[46]
c.5401G>A	Missense point mutation	1801	p.E1801K	e	Heterozygous	LMM	Laing distal myopathy	3	100%	100%		100%			100%			100%	67%	5.67	Ruggiero et al 2015[88]
c.5401G>A	Missense point mutation	1801	p.E1801K	e		LMM	Laing distal myopathy	2	100%	100%	0%	0%		0%		100%	0%	100%	100%	5	Lamont et al 2014[72]
c.5566G>A	Missense point mutation	1856	p.E1856K	c		LMM	Laing distal myopathy	4	33%	50%	50%	50%	25%		0%			100%	75%	3.83	Finsterer et al 2014[90]
c.4399C>G	Missense point mutation	1467	p.L1467V	a	Heterozygous	LMM	Myosin storage myopat	3	100%	100%		100%			0%		100%	100%	100%	6	Cullup et al 2012[71]
c.4763G>C	Missense point mutation	1588	p.R1588P	b	Heterozygous	LMM	Myosin storage myopat	1	100%	100%	100%							100%		4	Cullup et al 2012[71]
c.5352_5354delGAA	Tricnucleotide deletion	1784	p.K1784del	b	Heterozygous	LMM	Myosin storage myopat	1	100%	100%				100%	100%	100%	100%	100%	0%	7	Stalpers et al 2011[18]
c.5378T>C	Missense point mutation	1793	p.L1793P	d	Heterozygous	LMM	Myosin storage myopat	2	100%	50%						100%	50%			3	Dye et al 2006[87]
c.5458C>T	Missense point mutation	1820	p.R1820W	b	Homozygous	LMM	Myosin storage myopat	2	0%					100%			50%	100%	50%	3	Yuceyar et al 2015[89]
c.5533C>T	Missense point mutation	1845	p.R1845W	f	Heterozygous	LMM	Myosin storage myopat	1		100%		100%			100%			100%		4	Li et al 2018[82]
c.5533C>T	Missense point mutation	1845	p.R1845W	f	Heterozygous	LMM	Myosin storage myopat	4	50%	100%	50%	50%	25%	50%	50%			100%		4.75	Pegoraro et al 2007[20]
c.5533C>T	Missense point mutation	1845	p.R1845W	f	Heterozygous	LMM	Myosin storage myopat	11	100%	100%	100%	100%		100%	100%		100%	100%	100%	9	Tajsharghi et al 2003[17]
c.23014C>T	Missense point mutation	1845	p.R1845W	f	Heterozygous	LMM	Myosin storage myopat	1	100%	100%		100%		100%	100%				100%	6	Laing et al 2005[22]
c.23014C>T	Missense point mutation	1845	p.R1845W	f	Heterozygous	LMM	Myosin storage myopat	2	50%	50%	50%	100%		100%	100%		50%		0%	5	Shingde et al 2006[19]
c.23014C>T	Missense point mutation	1846	p.R1845W	f	Heterozygous	LMM	Myosin storage myopat	1	0%	100%		100%		100%	100%	100%		100%	100%	7	Laing et al 2005[22]
c.24012G>A	Missense point mutation	1883	p.E1883K	b	Homozygous	LMM	Myosin storage myopat	3	67%	67%		67%			100%	100%	67%		100%	5.67	Tajsharghi et al 2007[32]
c.25596A>T	Missense point mutation	1904	p.H1904L	f	Heterozygous	LMM	Myosin storage myopat	10	100%	100%	100%	100%		100%	0%			100%	0%	6	Bohlega et al 2004[92]
c.5740G>A	Missense point mutation	1914	p.E1914K	e		LMM	Myosin storage myopat	1	100%	100%	0%	100%		0%			0%		100%	4	Lamont et al 2014[72]
c.5807A>T	Missense point mutation	1936	p.X1936LfsX32		Heterozygous	LMM	Myosin storage myopat	1	100%	100%		100%		100%		100%			0%	5	Banfai et al 2017[94]
c.5807A>G	Missense point mutation	1936	p.X1936WfsX32		Heterozygous	LMM	Myosin storage myopat	12		100%		100%						100%		3	Ortolano et al 2011[25]

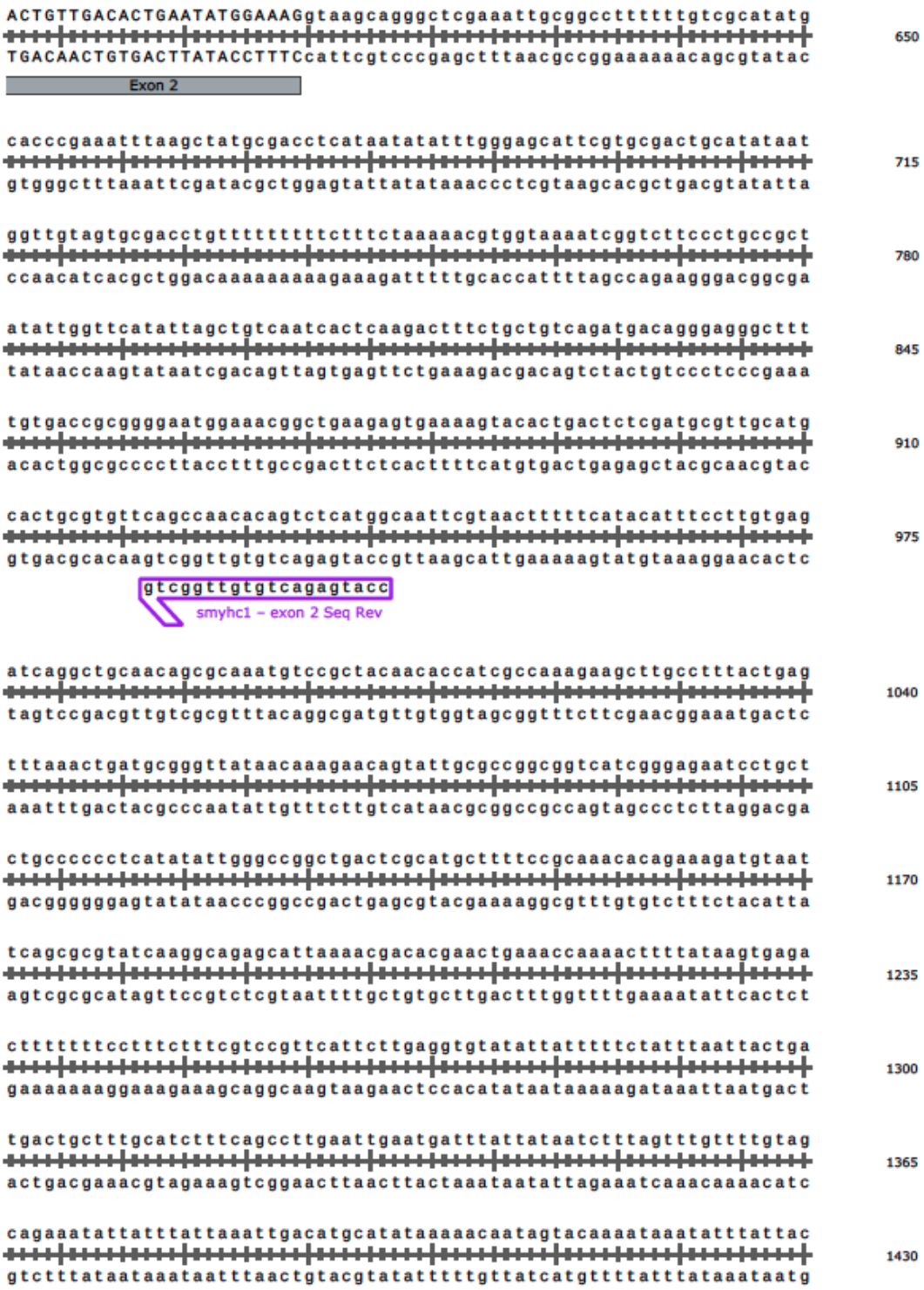
Appendix 2.1 – Primer design for *smyhc1*

Sequence: smyhc1 primer design.dna (Linear / 3636 bp)
Features: 7 total
Primers: 8 total

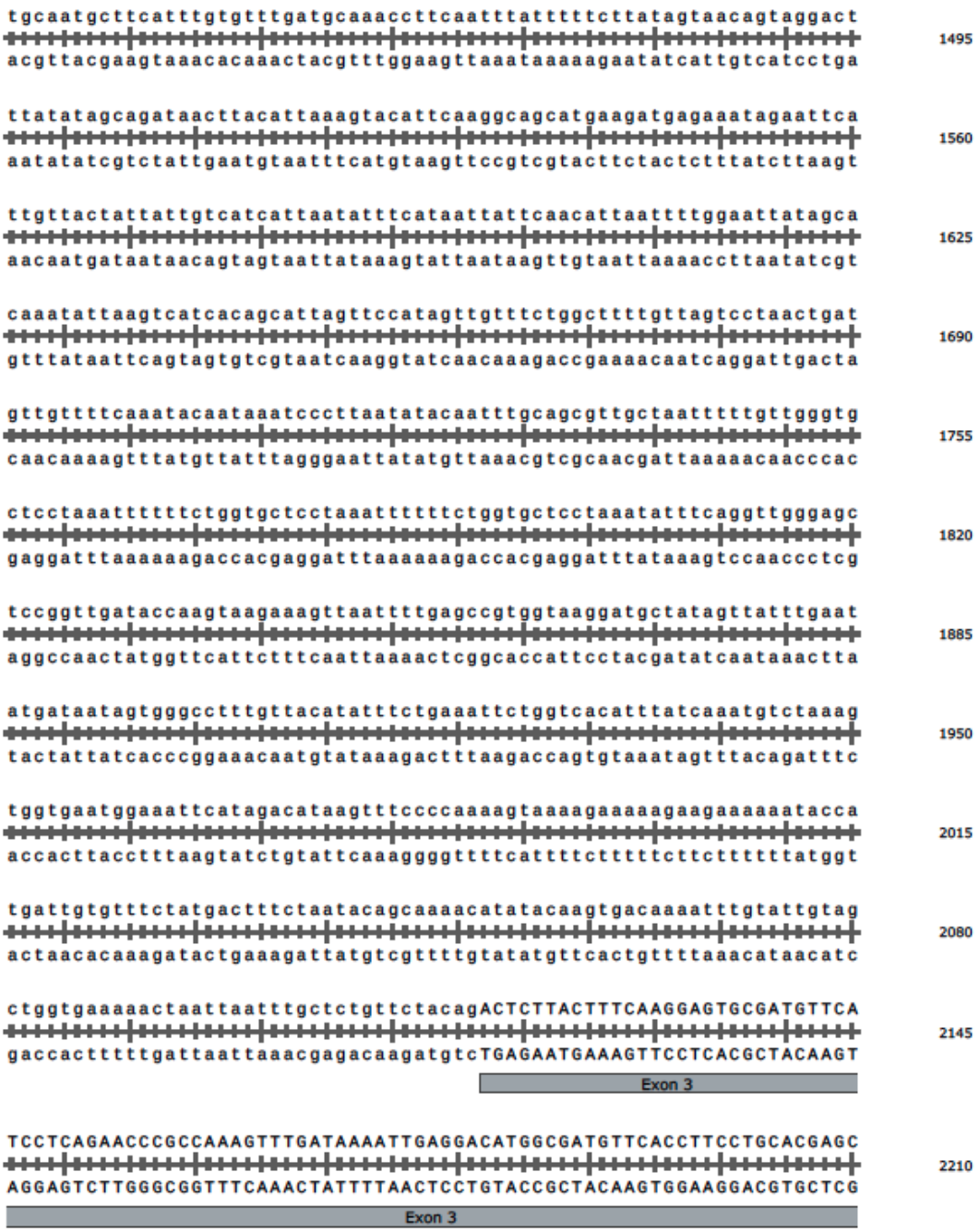


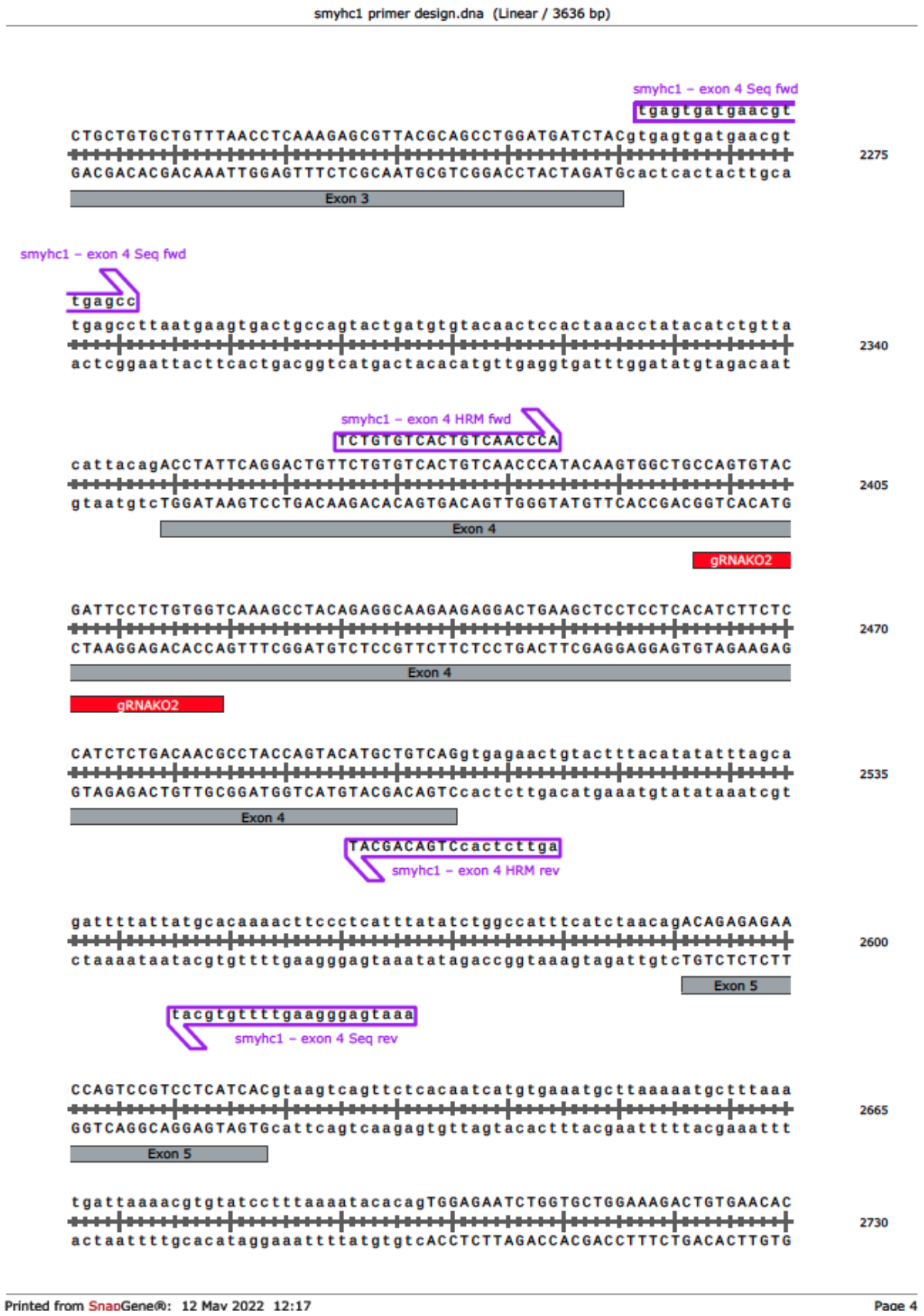
Printed from SnapGene®: 12 May 2022 12:17 Page 1

smyhc1 primer design.dna (Linear / 3636 bp)



smyhc1 primer design.dna (Linear / 3636 bp)




smyhc1 primer design.dna (Linear / 3636 bp)
smyhc1 – exon 4 Seq fwd
tgagtgatgaacgt
CTGCTGTGCTGTTTAACCTCAAAGAGCGTTACGCAGCCTGGATGATCTACgtgagtgatgaacgt
GACGACACGACAAATTGGAGTTTCTCGCAATGCGTCGGACCTACTAGATGcactcactacttgca
2275
Exon 3
smyhc1 – exon 4 Seq fwd
tgagcc
tgagccttaatgaagtgactgccagtactgatgtgtacaactccactaaacctatacatctgtta
actcggaattacttcactgacggtcatgactacacatgttgaggtgatttggatatgtagacaat
2340
smyhc1 – exon 4 HRM fwd
TCTGTGTCACTGTCAACCCA
cattacagACCTATTCAGGACTGTTCTGTGTCACTGTCAACCCATACAAGTGGCTGCCAGTGTAC
gtaatgtcTGGATAAGTCCTGACAAGACACAGTGACAGTTGGGTATGTTCACCGACGGTCACATG
2405
Exon 4
gRNAKO2
GATTCCTCTGTGGTCAAAGCCTACAGAGGCAAGAAGAGGACTGAAGCTCCTCCTCACATCTTCTC
CTAAGGAGACACCAGTTTCGGATGTCTCCGTTCTTCTCCTGACTTCGAGGAGGAGTGTAGAAGAG
2470
Exon 4
gRNAKO2
CATCTCTGACAACGCCTACCAGTACATGCTGTCAGgtgagaactgtactttacatatatttagca
GTAGAGACTGTTGCGGATGGTCATGTACGACAGTCcactcttgacatgaaatgtatataaatcgt
2535
Exon 4
TACGACAGTCcactcttga
smyhc1 – exon 4 HRM rev
gattttattatgcacaaaacttccctcatttatatctggccatttcatctaacagACAGAGAGAA
ctaaaataatacgtgttttgaagggagtaaatatagaccggtaaagtagattgtcTGTCTCTCTT
2600
Exon 5
tacgtgttttgaagggagtaaa
smyhc1 – exon 4 Seq rev
CCAGTCCGTCCTCATCACgtaagtcagttctcacaatcatgtgaaatgcttaaaaatgctttaaa
GGTCAGGCAGGAGTAGTGcattcagtcaagagtgttagtacactttacgaatttttacgaaattt
2665
Exon 5
tgattaaaacgtgtatccttaaaatacacagTGGAGAATCTGGTGCTGGAAAGACTGTGAACAC
actaattttgcacataggaaattttatgtgtcACCTCTTAGACCACGACCTTTCTGACACTTGTG
2730
Printed from SnapGene®: 12 May 2022 12:17
Page 4

Appendix 3.1 – Mant-ATP Assay Average Data

3.1.1. - SRX and DRX values from slow fibres

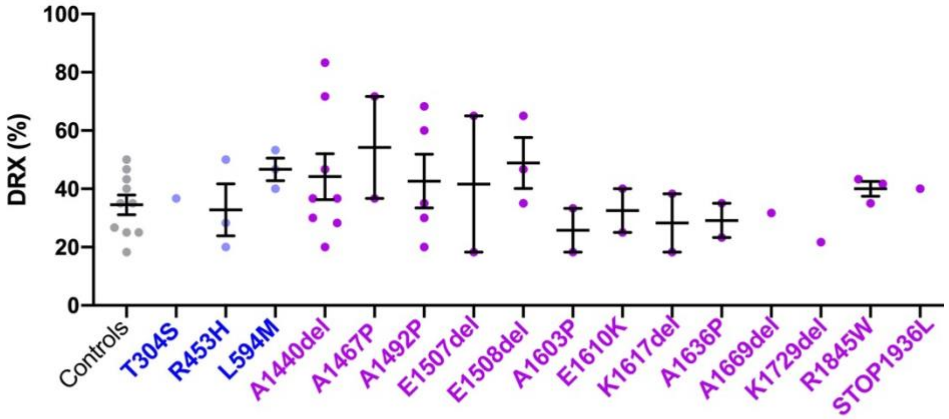
	P1	P2	DRX	SRX	n
Controls	63.78	36.22	40%	60%	18
T304S	65.00	35.00	42%	58%	11
R453H	64.40	35.60	41%	59%	10
L594M	62.45	37.55	37%	63%	11
A1440del	73.30	26.70	56%	45%	10
A1467P	74.00	26.00	57%	43%	8
A1492P	75.30	24.70	59%	41%	10
E1507del	62.91	37.09	38%	62%	11
E1508del	71.71	28.29	53%	47%	7
A1603P	75.71	24.29	60%	40%	14
E1610K	76.00	24.00	60%	40%	10
K1617del	72.40	27.60	54%	46%	10
A1636P	74.73	25.36	58%	42%	11
L1657P	76.25	23.75	60%	39%	8
E1669del	76.27	23.73	60%	40%	11
K1729del	76.75	23.25	61%	39%	8
R1845W	74.00	26.00	57%	43%	11
A1883E	61.40	38.60	36%	64%	10
STOP1936L	76.00	23.89	60%	40%	9

3.1.2. - SRX and DRX values from fast fibres

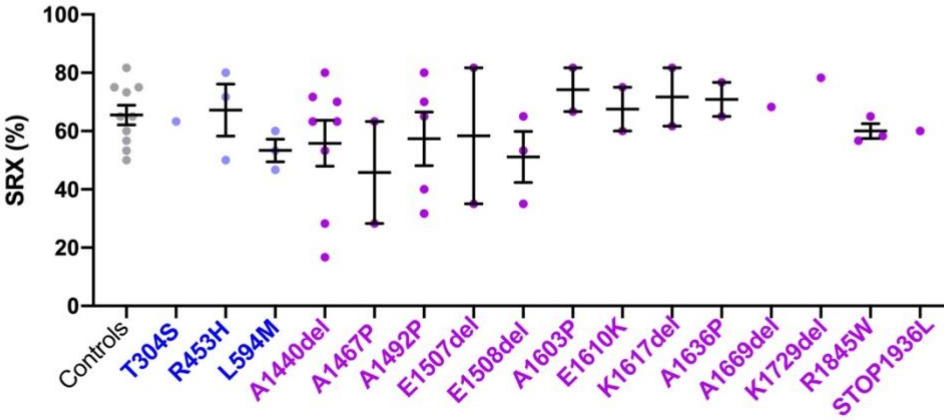
A

	P1	P2	DRX	SRX	n
Controls	60.70	39.30	35%	66%	10
T304S	62.00	38.00	37%	63%	1
R453H	59.67	40.33	33%	67%	3
L594M	68.00	32.00	47%	53%	3
A1440del	68.57	31.43	48%	52%	7
A1467P	65.67	34.33	43%	57%	3
A1492P	65.60	34.40	43%	57%	5
E1507del	65.00	35.00	42%	58%	3
E1508del	69.33	30.67	49%	51%	3
A1603P	55.50	44.50	26%	74%	2
E1610K	59.50	40.50	33%	68%	2
K1617del	57.00	43.00	28%	72%	2
A1636P	57.50	42.50	29%	71%	2
L1657P	78.50	21.00	65.00%	35.00%	2
E1669del	59.00	41.00	32%	68%	1
K1729del	53.00	47.00	22%	78%	1
R1845W	64.00	36.00	40%	60%	3
A1883E	0.00	0.00	0%	0%	0
STOP1936L	64.00	36.00	40%	60%	1

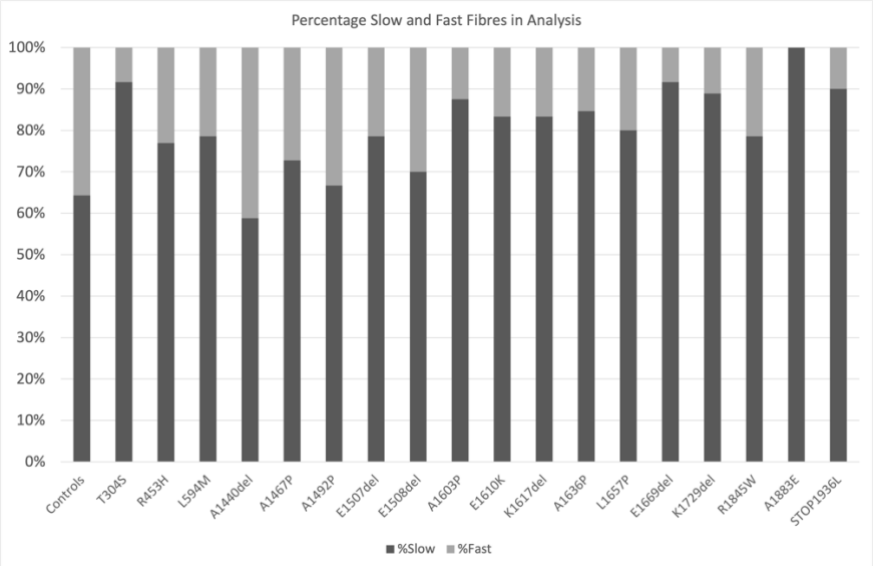
B



C



3.1.3. - Proportion of DRX increases in patients with LMM mutations.



3.1.4. - Proportions of fast and slow fibres analysed in Mant-ATP assay

	Total N	no. slow fibres	no. fast fibres	% slow	% fast
Controls	28	18	10	64%	36%
T304S	12	11	1	92%	8%
R453H	13	10	3	77%	23%
L594M	14	11	3	79%	21%
A1440del	17	10	7	59%	41%
A1467P	11	8	3	73%	27%
A1492P	15	10	5	67%	33%
E1507del	14	11	3	79%	21%
E1508del	10	7	3	70%	30%
A1603P	16	14	2	88%	13%
E1610K	12	10	2	83%	17%
K1617del	12	10	2	83%	17%
A1636P	13	11	2	85%	15%
L1657P	10	8	2	80%	20%
E1669del	12	11	1	92%	8%
K1729del	9	8	1	89%	11%
R1845W	14	11	3	79%	21%
A1883E	10	10	0	100%	0%
STOP1936L	10	9	1	90%	10%

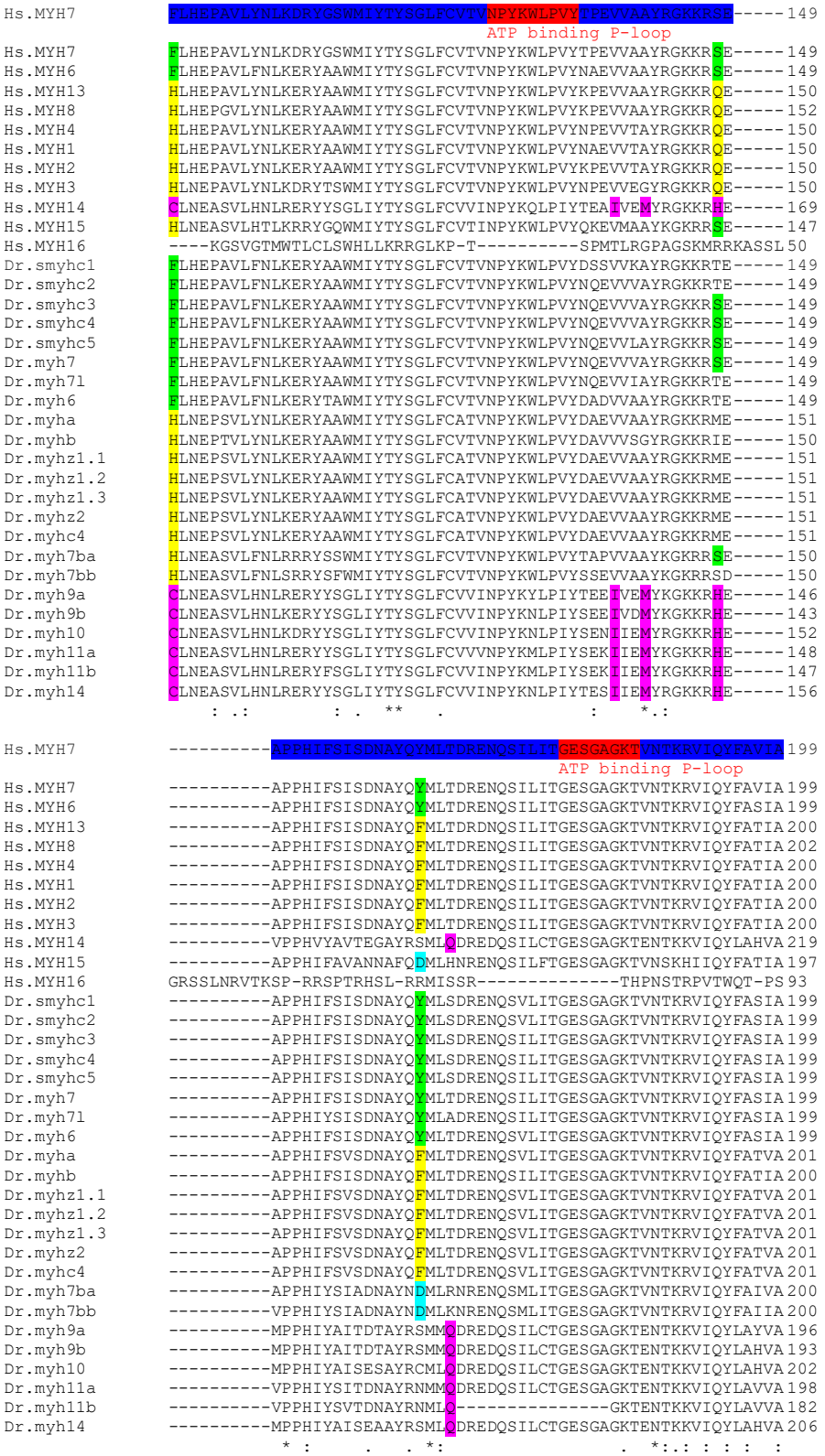
Appendix 4.1 – MYH6 and MYH7 signature amino acids

Amino acid		35	282	318	1111
MYH Protein/region		S1			LMM
Mammals	**MYH7**	K	D	T	L
Coelocanth	myh7	K	D	T	L
Xenopus	myh7	K	D	T	L
Rainbow Trout	myh7	K	D	T	L
Atlantic Salmon	myh7	K	D	T	L
Northern Pike	myh7	K	D	T	L
Atlantic Herring	myh7	K	D	T	L
Zebrafish	**smyhc1**	K	D	T	L
Zebrafish	**smyhc2**	K	D	T	L
Zebrafish	**smyhc3**	K	D	T	L
Zebrafish	**smyhc4**	K	D	T	L
Zebrafish	**smyhc5**	K	D	T	L
Platyfish	**smyhc1**	R	D	T	L
Platyfish	**smyhc2**	R	D	T	L
Tilapia	**smyhc1**	K	D	T	L
Medaka	**smyhc1**	R	D	T	L
Zebrafish	**myh7**	K	D	T	L
Platyfish	**myh7**	K	D	T	L
Zebrafish	**myhl**	K	D	T	L
Cod	**myh7l**	K	D	T	L
Tilapia	**myh7l**	K	D	T	-

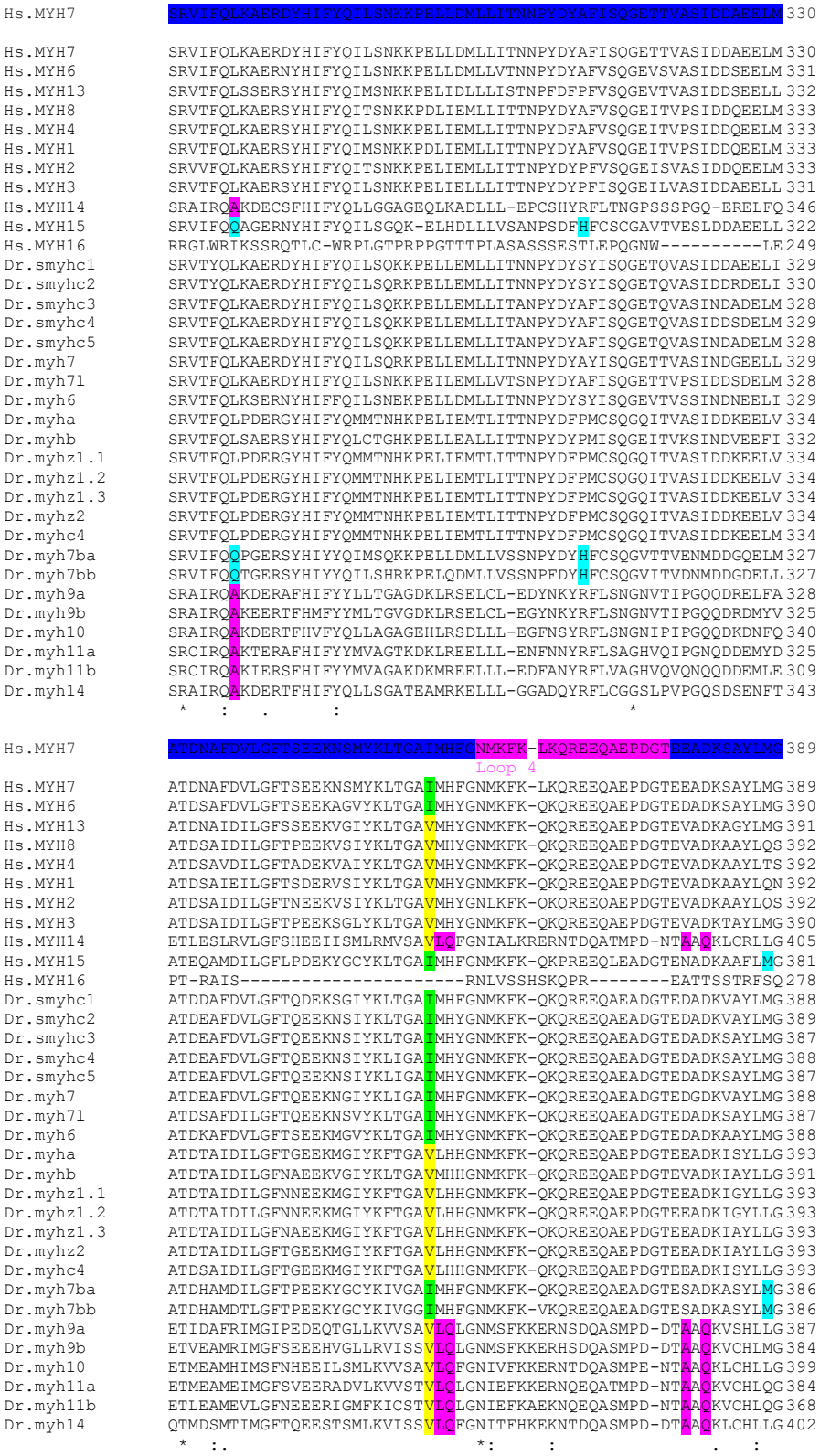
Amino acid		35	282	318	1111
MYH Protein/region		S1			LMM
Mammals	**MYH6**	T	N	V	N
Coelocanth	myh6	K	N	V	H
Xenopus	myh6	T	D	V	H
Rainbow Trout	myh6	-	D	I	N
Atlantic Salmon	myh6	R	D	I	N
Northern Pike	myh6	R	D	I	N
Atlantic Herring	myh6	V	D	I	-
Zebrafish	myh6	T	N	V	N
Platyfish	myh6	T	N	V	N
Tilapia	myh6	T	N	V	S
Cod	myh6	T	N	V	N

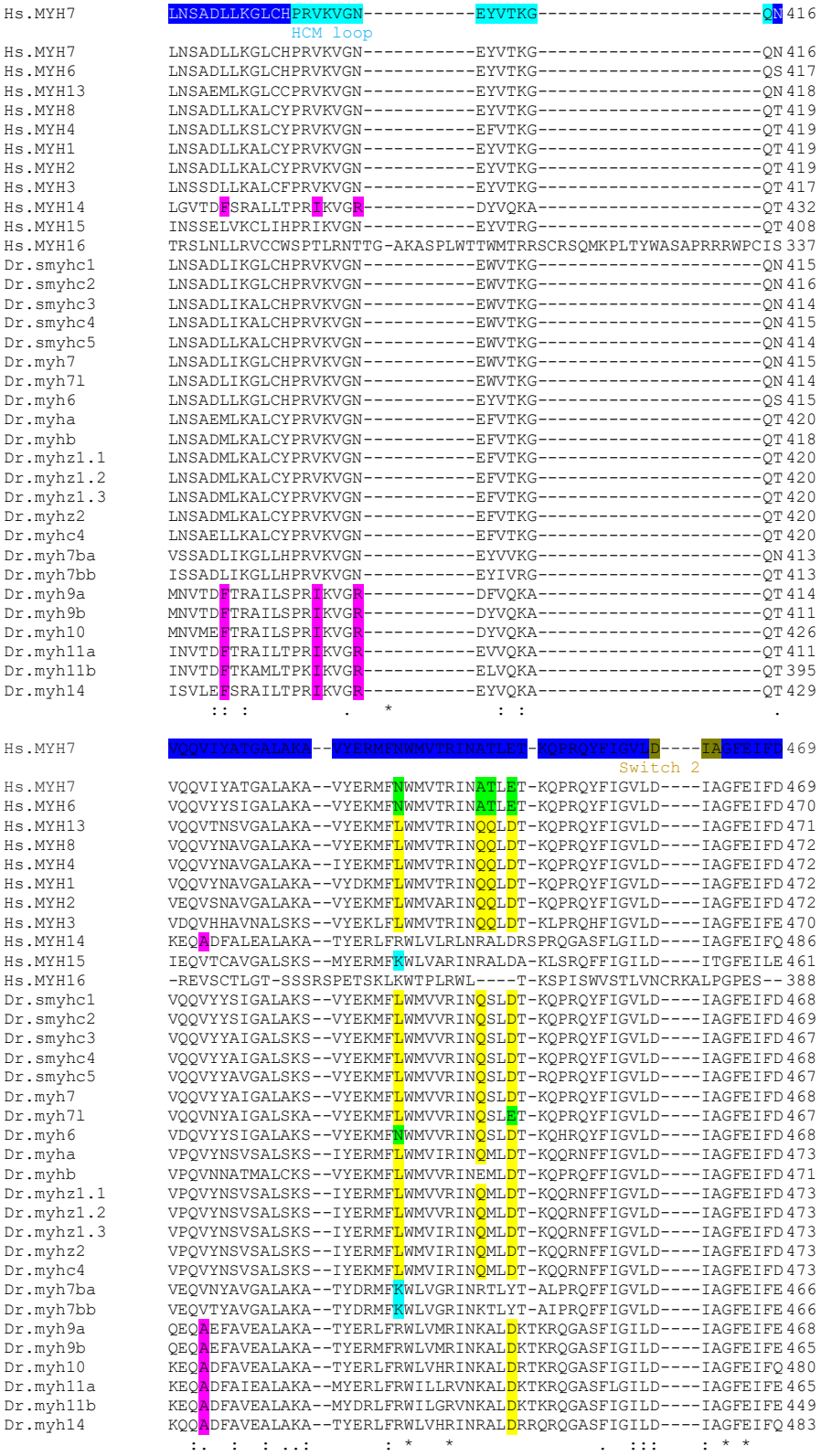
Appendix 4.2 – CLUSTALO Human MYH vs Zebrafish MYH proteins

```
Hs.MYH7          --------MGD----SEMAVFGAAAPYLRKSEKE------------RLEAQTRPFDLKKD 36
                                                                 SH3-like domain
Hs.MYH7          --------MGD----SEMAVFGAAAPYLRKSEKE------------RLEAQTRPFDLKKD 36
Hs.MYH6          --------MTD----AQMADFGAAAQYLRKSEKE------------RLEAQTRPFDIRTE 36
Hs.MYH13         -------MSSD----AEMAIFGEAAPYLRKPEKE------------RIEAQNRPFDSKKA 37
Hs.MYH8          -----MSASSD----AEMAVFGEAAPYLRKSEKE------------RIEAQNKPFDAKTS 39
Hs.MYH4          -------MSSD----SEMAIFGEAAPFLRKSEKE------------RIEAQNKPFDAKTS 37
Hs.MYH1          -------MSSD----SEMAIFGEAAPFLRKSERE------------RIEAQNKPFDAKTS 37
Hs.MYH2          -------MSSD----SELAVFGEAAPFLRKSERE------------RIEAQNRPFDAKTS 37
Hs.MYH3          -------MSSD----TEMEVFGIAAPFLRKSEKE------------RIEAQNQPFDAKTY 37
Hs.MYH14         MAAVTMSVPGRKAPPRPGPVPEAAQPFLFTPRGPSAGGGPG-----SGTSPQVEWTARRL 55
Hs.MYH15         --------------MDLSDLGEAAAFLRRSEAE------------LLLLQATALDGKKK 33
Hs.MYH16         ------------------------------------------------------------ 0
Dr.smyhc1        --------MGD----AVMAEFGSAAPFLRKSDKE------------RLEAQTRIFDMKKE 36
Dr.smyhc2        --------MGD----AVMAEFGPAAPFLRKSDKE------------RLEAQTRPFDMKKE 36
Dr.smyhc3        --------MGD----AVMAEFGAAAPYLRKSDRE------------RLEAQTRPFDMKKE 36
Dr.smyhc4        --------MGD----AVMAEFGAAAPYLRKSDKE------------RLEAQTRPFDMKKE 36
Dr.smyhc5        --------MGD----ALMEEFGAAAPYLRKSDRE------------RLEAQTRPFDMKKE 36
Dr.myh7          --------MGD----AQMAEFGAAASYLRKSDRE------------RLEAQTRPFDMKKE 36
Dr.myh7l         --------MGD----AEMSVFGAAAPYLRKSEKE------------RLEAQTKAFDLKKE 36
Dr.myh6          --------MGD----ALMAEFGKAAPFLRKSDKE------------RLEAQTRAFDIKTE 36
Dr.myha          -------MSTD----AEMAVYGKAAIYLRKPEKE------------RIEAQNKPFDAKSA 37
Dr.myhb          -------MSGD----PEMECFGPAAVYLRKPEKE------------RIEAQNRPFDAKTA 37
Dr.myhz1.1       -------MSTD----AEMAVYGKAAIYLRKPEKE------------RVEAQNKPFDAKTA 37
Dr.myhz1.2       -------MSTD----AEMAVYGKAAIYLRKPEKE------------RIEAQNKPFDAKTA 37
Dr.myhz1.3       -------MSTD----AEMAVYGKAAIYLRKPEKE------------RIEAQNKPFDAKTA 37
Dr.myhz2         -------MSTD----AEMAIYGKAAIFLRKPEKE------------RIEAQSKPFDAKTA 37
Dr.myhc4         -------MSTD----AEMAVYGKAAIYLRKPEKE------------RIEAQNKPFDAKSA 37
Dr.myh7ba        -------MSRM----LDMKEFGEAAPFLRKSDLE------------LLAAQTVAFDGKKR 37
Dr.myh7bb        -------MSRF----MELREFGEAATFLRKTNLE------------QLAAQSHAFDGKKR 37
Dr.myh9a         ------------------XAKMSDAEKFLYADRNTI-----------NDPLAQADWATKKL 32
Dr.myh9b         --------------------MSDVDKFLYVDRNLV-----------NNPLAQADWATKKL 29
Dr.myh10         -------MPEM----AQRSGQEDPERYLFVDRAVV-----------YNPTTQADWTAKKL 38
Dr.myh11a        ---------------MTKKGLSDDEKFLFTDKDFI-----------NSPVAQADWSAKKL 34
Dr.myh11b        ---------------MTMQDNDDSNKFLFLDSEFK-----------NSGVAQADWSTRKM 34
Dr.myh14         -----MSRP-------AGGSINDVACFL------STGAGPGSPTSVFSASSQADWAAKRL 42


Hs.MYH7          VFVPDDKQEFVKAKIVSRE-GGKVTAETEYGK-TVTVKEDQVMQQNPPKFDKIEDMAMLT 94
                 SH3-like domain
Hs.MYH7          VFVPDDKQEFVKAKIVSRE-GGKVTAETEYGK-TVTVKEDQVMQQNPPKFDKIEDMAMLT 94
Hs.MYH6          CFVPDDKEEFVKAKILSRE-GGKVIAETENGK-TVTVKEDQVLQQNPPKFDKIEDMAMLT 94
Hs.MYH13         CFVADNKEMYVKGMIQTRE-NDKVIVKTLDDR-MLTLNNDQVFPMNPPKFDKIEDMAMMT 95
Hs.MYH8          VFVAEPKESYVKSTIQSKE-GGKVTVKTEGGA-TLTVREDQVFPMNPPKYDKIEDMAMMT 97
Hs.MYH4          VFVVDPKESYVKAIVQSRE-GGKVTAKTEAGA-TVTVKEDQVFSMNPPKYDKIEDMAMMT 95
Hs.MYH1          VFVVDPKESFVKATVQSRE-GGKVTAKTEAGA-TVTVKDDQVFPMNPPKYDKIEDMAMMT 95
Hs.MYH2          VFVAEPKESFVKGTIQSRE-GGKVTVKTEGGA-TLTVKDDQVFPMNPPKYDKIEDMAMMT 95
Hs.MYH3          CFVVDSKEEYAKGKIKSSQ-DGKVTVETEDNR-TLVVKPEDVYAMNPPKFDRIEDMAMLT 95
Hs.MYH14         VWVPSELHGFEAAALRDEG-EEEAEVELAESGRRLRLPRDQIQRMNPPKFSKAEDMAELT 114
Hs.MYH15         CWIPDGENAYIEAEVKGSEDDGTVIVETADGE-SLSIKEDKIQQMNPPEFEMIEDMAMLT 92
Hs.MYH16         CLVGT---------------------------------------------------------- 5
Dr.smyhc1        CFVPDPEVEYVKASITSRD-GDKVTVDTEYGK-TLTFKECDVHPQNPPKFDKIEDMAMFT 94
Dr.smyhc2        CFVPDPEVEYVKASVTSRD-GDKVTVETEFGK-TVTVKEVDCHPQNPPKFDKIEDMAMFT 94
Dr.smyhc3        CFVPDADEEYLKATVISRD-GDKATCETSKGT-TVTVKECDVHPQNPPKFDKIEDMAMFT 94
Dr.smyhc4        CFVPDADEEYLKATVISRD-GDKVTCETSKKT-TVTVKECDVHPQNPPKFDKIEDMAMFT 94
Dr.smyhc5        CFVPDTDEEYVKGSIISRD-GDKVTCETEKGK-TVTVKECDVHPQNPPKFDKIEDMAMFT 94
Dr.myh7          CFVPDPDEEYVKASIVSRE-GDKVTVQTEKRK-TVTVKEADIHPQNPPKFDKIEDMAMFT 94
Dr.myh7l         CFVPDAIEEFVKATVVSRE-GDKVTVETQGGK-TVTVKEADVLQQNPPKFDKIEDMAMLT 94
Dr.myh6          CFVVDEKVEYVKGQIQNKD-GGKVTVKTEDGR-TVTVKDGDVHPQNPPKFDKIEDMAMLT 94
Dr.myha          CYVVDDKELYVKGTIKSRD-GGKVTVITLDTKEERVAKEEDVHPMNPPKFDKIEDMAMMT 96
Dr.myhb          YFVSEPKEMYLKGVLKSKE-GGKATVQTLCGKT-LTVKEDEIFPMNPPKFDKIEDMAMMT 95
Dr.myhz1.1       CYVVDDKELYVKGTIKSKD-GGKVTVITLDTKEEKVVKEDDVHPMNPPKFDKIEDMAMMT 96
Dr.myhz1.2       CYVVDDKELYVKGTIKSRD-GGKVTVITLDTKEERVAKEDDVHPMNPPKFDKIEDMAMMT 96
Dr.myhz1.3       CYVVDDKELYVKGTIKSRD-GGKVTVITLDTKEERVAKEDDVHPMNPPKFDKIEDMAMMT 96
Dr.myhz2         CYVVDDKELYVKGTIKSKD-GGKVTVVTLDTQTEKVVKEDDVHPMNPPKFDKIEDMAMMT 96
Dr.myhc4         CYVVDDKELYVKGTIKSKD-GGKVTVITLDTKEERVVKEDDVHPMNPPKYDKIEDMAMMT 96
Dr.myh7ba        AWIPDDKDAYIEVEIKQID-GDRVEVETKDGK-CLTVKEDDIQQMNPPKFDLIEDMAMLT 95
Dr.myh7bb        VWIPDEKEAYIEVEIKDTD-GDKVMVETKDGR-MLTVKEEDIQQMNPPKFDLMEDMAMLT 95
Dr.myh9a         VWVPSEKLGFEAGSIKEET-GDECLVELADSGKKIKVNKDDIQKMNPPKFSKVEDMAELT 91
Dr.myh9b         VWVPSERLGFEAGSLKEEH-GDEVVVELADSGKKIRVNKDDIQKMNPPKFSKVEDMAELT 88
Dr.myh10         VWVPSERHGFEAASIREER-GEEVLVELAENGKKAMVNKDDIQKMNPPKFSKVEDMAELT 97
Dr.myh11a        VWVPSEKHGFESASIKEEH-GDEVLVELMDNGKKITVNKDDIQKMNPPKFSKVEDMAELT 93
Dr.myh11b        VWIPSERDGFQSASIKEET-GNEVVVEL-DNGQKVTVSKDDIQKMNPPKFNKVEDMAALT 92
Dr.myh14         VWVPSEKHGFESASIREER-GDEVEVELTDSGRKLTLLREELQRMNPPRFSKVEDMADLT 101
                   :
```


ATP binding P-loop
ATP binding P-loop

```
Hs.MYH7          AIGDRSK-KD-----Q--SPGKGTLEDQII-----------QANPALEAFGNAKTVRND 239
                 Loop 1                                    Switch 1
Hs.MYH7          AIGDRSK-KD-----Q--SPGKGTLEDQII-----------QANPALEAFGNAKTVRND 239
Hs.MYH6          AIGDRGK-KD-----N-ANANKGTLEDQII-----------QANPALEAFGNAKTVRND 240
Hs.MYH13         VTGDKK--KE-----TQPGKMQGTLEDQII-----------QANPLLEAFGNAKTVRND 241
Hs.MYH8          VTGEKKK-D-------ESGKMQGTLEDQII-----------SANPLLEAFGNAKTVRND 242
Hs.MYH4          VTGEKKK-EE-----PASGKMQGTLEDQII-----------SANPLLEAFGNAKTVRND 242
Hs.MYH1          VTGEKKK-EE-----VTSGKMQGTLEDQII-----------SANPLLEAFGNAKTVRND 242
Hs.MYH2          VTGEKKK-EE-----ITSGKIQGTLEDQII-----------SANPLLEAFGNAKTVRND 242
Hs.MYH3          ATGDLAK-K-------KDSKMKGTLEDQII-----------SANPLLEAFGNAKTVRND 240
Hs.MYH14         SSPKGRKEPGV----------PGELERQLL-----------QANPILEAFGNAKTVKND 257
Hs.MYH15         AMIES-------------RKKQGALEDQIM-----------QANTILEAFGNAKTLRND 232
Hs.MYH16         -MRPVS-TIC-----ANATPT-GSIPTRACSA-RSTPTSGCPSTGPVWLTCTRA---RSA 141
Dr.smyhc1        AVS-G-K-KD-----A-ASEKKGTLEDQII-----------QANPALEAFGNAKTIRND 238
Dr.smyhc2        AAPGG-K-KD-----P-SQEKKGTLEDQII-----------QCNPALEAFGNAKTIRND 239
Dr.smyhc3        ASP-T-K-K------E-TTEKKGTLEDQII-----------QCNPALEAFGNAKTIRND 237
Dr.smyhc4        AGS-S-K-KD-----S-SSEKKGTLEDQII-----------QCNPALEAFGNAKTIRND 238
Dr.smyhc5        ASP-T-K-K------E-TTEKKGTLEDQII-----------QCNPALEAFGNAKTIRND 237
Dr.myh7          AGGS--A-KK-----E-GAEKKGTLEDQII-----------QANPALEAFGNAKTIRND 238
Dr.myh7l         ASGGK-K--------D-QDKNKGTLEDQII-----------QANPALEAFGNAKTIRND 237
Dr.myh6          AAGGSAG-K--------KDSSKGTLEDQII-----------QANPALEAFGNAKTLRND 238
Dr.myha          VQGGDKK-KE-----QTPGKMQGSLEDQII-----------AANPLLEAYGNAKTVRND 243
Dr.myhb          VAGKQK--QE-----PIPGKMQGSLEDQII-----------AANPLLEAYGNAKTVRND 241
Dr.myhz1.1       VQGPEKK-KE-----QASGKMQGSLEDQII-----------AANPLLEAYGNAKTVRND 243
Dr.myhz1.2       VQGPEKK-KE-----QAAGKMQGSLEDQII-----------AANPLLEAYGNAKTVRND 243
Dr.myhz1.3       VQGPEKK-KE-----QAAGKMQGSLEDQII-----------AANPLLEAYGNAKTVRND 243
Dr.myhz2         VQGGDKK-KE-----QAAGKMQGSLEDQII-----------AANPLLEAYGNAKTVRND 243
Dr.myhc4         VQGGDKK-KE-----QAPGKMQGSLEDQII-----------AANPLLEAYGNAKTVRND 243
Dr.myh7ba        ALGEA------------AAKKGGTLEDQII-----------EANPAMEAFGNAKTLRND 236
Dr.myh7bb        ALGEA------------GGKKGGTLEDQII-----------EANPAMEAFGNAKTLRND 236
Dr.myh9a         SSFKTKKDQSS------IALSHGELEKQLL-----------QANPILEAFGNAKTVKND 238
Dr.myh9b         SSHKTKKDQSS------SVLSHGELEKQLL-----------QANPILEAFGNAKTVKND 235
Dr.myh10         SSHKGRKDHNIPPESPKAVKLQGELERQLL-----------QANPILESFGNAKTVKND 250
Dr.myh11a        SSHKGKKDMS-----------AGELEKQLL-----------QANPILEAFGNAKTIKND 235
Dr.myh11b        SSHKGKKEAT-----------SGELEKQLL-----------QANPILEAFGNAKTIKND 219
Dr.myh14         SSHKSGT-LGRPKDTVVQTVQYGELERQLL-----------QANPILEAFGNAKTVKND 253
                                      * :  :    *          .    :  .*   :.

Hs.MYH7          NSSRFGKFIRIHFGATGKLASADIETYLLEK----------------------------270
                 Switch 1
Hs.MYH7          NSSRFGKFIRIHFGATGKLASADIETYLLEK----------------------------270
Hs.MYH6          NSSRFGKFIRIHFGATGKLASADIETYLLEK----------------------------271
Hs.MYH13         NSSRFGKFIRIHFGATGKLASADIETYLLEK----------------------------272
Hs.MYH8          NSSRFGKFIRIHFGTTGKLASADIETYLLEK----------------------------273
Hs.MYH4          NSSRFGKFIRIHFGATGKLASADIETYLLEK----------------------------273
Hs.MYH1          NSSRFGKFIRIHFGTTGKLASADIETYLLEK----------------------------273
Hs.MYH2          NSSRFGKFIRIHFGTTGKLASADIETYLLEK----------------------------273
Hs.MYH3          NSSRFGKFIRIHFGTTGKLASADIETYLLEK----------------------------271
Hs.MYH14         NSSRFGKFIRINFDVAGYIVGANIETYLLEK----------------------------288
Hs.MYH15         NSSRFGKFIRMHFGARGMLSSVDIDIYLLEK----------------------------263
Hs.MYH16         QRCRLTSSPSLTTPTTTCLWIVRISLC-SPENLVLVRLRTRRRSSSTLPTLEELANRPQI 200
Dr.smyhc1        NSSRFGKFIRIHFGVSGKLASADIETYLLEK----------------------------269
Dr.smyhc2        NSSRFGKFIRIHFGVSGKLSSADIETYLLEK----------------------------270
Dr.smyhc3        NSSRFGKFIRIHFAASGKLASADIETYLLEK----------------------------268
Dr.smyhc4        NSSRFGKFIRIHFAASGKLASADIETYLLEK----------------------------269
Dr.smyhc5        NSSRFGKFIRIHFAANGKLASADIETYLLEK----------------------------268
Dr.myh7          NSSRFGKFIRIHFGASGKLASADIETYLLEK----------------------------269
Dr.myh7l         NSSRFGKFIRIHFDTRGKLASADIETYLLEK----------------------------268
Dr.myh6          NSSRFGKFIRIHFGTSGKLSSADIETYLLEK----------------------------269
Dr.myha          NSSRFGKFIRIHFGTTGKLASADIETYLLEK----------------------------274
Dr.myhb          NSSRFGKFIRIHFGTTGKLASADIETYLLEK----------------------------272
Dr.myhz1.1       NSSRFGKFIRIHFGTSGKLASADIETYLLEK----------------------------274
Dr.myhz1.2       NSSRFGKFIRIHFGTSGKLASADIETYLLEK----------------------------274
Dr.myhz1.3       NSSRFGKFIRIHFGTSGKLASADIETYLLEK----------------------------274
Dr.myhz2         NSSRFGKFIRIHFGTSGKLASADIETYLLEK----------------------------274
Dr.myhc4         NSSRFGKFIRIHFGTTGKLASADIETYLLEK----------------------------274
Dr.myh7ba        NSSRFGKFIRIHFGPTGKLASADIDIYLLEK----------------------------267
Dr.myh7bb        NSSRFGKFIRIHFGPTGKLASADIDIYLLEK----------------------------267
Dr.myh9a         NSSRFGKFIRINFDVNGYIVGANIETYLLEK----------------------------269
Dr.myh9b         NSSRFGKFIRINFDVNGYIVGANIETYLLEK----------------------------266
Dr.myh10         NSSRFGKFIRINFDVTGYIVGANIETYLLEK----------------------------281
Dr.myh11a        NSSRFGKFIRINFDVTGFIVGANIETYLLEK----------------------------266
Dr.myh11b        NSSRFGKFIKINFDNTGYIVGANIETYLLEK----------------------------250
Dr.myh14         NSSRFGKFIRINFDVAGYIVGANIETYLLEK----------------------------284
                 : .*: .   :       :  . *.     :
```


Hs.MYH7 SRVIFQLKAERDYHIFYQILSNKKPELLDMLLITNNPYDYAFISQGETTVASIDDAEELM 330
Hs.MYH7 SRVIFQLKAERDYHIFYQILSNKKPELLDMLLITNNPYDYAFISQGETTVASIDDAEELM 330
Hs.MYH6 SRVIFQLKAERNYHIFYQILSNKKPELLDMLLVTNNPYDYAFVSQGEVSVASIDDSEELM 331
Hs.MYH13 SRVTFQLSSERSYHIFYQIMSNKKPELIDLLLISTNPFDFPFVSQGEVTVASIDDSEELL 332
Hs.MYH8 SRVTFQLKAERSYHIFYQITSNKKPDLIEMLLITTNPYDYAFVSQGEITVPSIDDQEELM 333
Hs.MYH4 SRVTFQLKAERSYHIFYQILSNKKPELIEMLLITTNPYDFAFVSQGEITVPSIDDQEELM 333
Hs.MYH1 SRVTFQLKAERSYHIFYQIMSNKKPDLIEMLLITTNPYDYAFVSQGEITVPSIDDQEELM 333
Hs.MYH2 SRVVFQLKAERSYHIFYQITSNKKPELIEMLLITTNPYDYPFVSQGEISVASIDDQEELM 333
Hs.MYH3 SRVTFQLKAERSYHIFYQILSNKKPELIELLLITTNPYDYPFISQGEILVASIDDAEELL 331
Hs.MYH14 SRAIRQAKDECSFHIFYQLLGGAGEQLKADLLL-EPCSHYRFLTNGPSSSPGQ-ERELFQ 346
Hs.MYH15 SRVIFQQAGERNYHIFYQILSGQK-ELHDLLLVSANPSDFHFCSCGAVTVESLDDAEELL 322
Hs.MYH16 RRGLWRIKSSRQTLC-WRPLGTPRPPGTTTPLASASSSESTLEPQGNW----------LE 249
Dr.smyhc1 SRVTYQLKAERDYHIFYQILSQKKPELLEMLLITNNPYDYSYISQGETQVASIDDAEELI 329
Dr.smyhc2 SRVTYQLKAERDYHIFYQILSQRKPELLEMLLITNNPYDYSYISQGETQVASIDDRDELI 330
Dr.smyhc3 SRVTFQLKAERDYHIFYQILSQKKPELLEMLLITANPYDYAFISQGETQVASINDADELM 328
Dr.smyhc4 SRVTFQLKAERDYHIFYQILSQKKPELLEMLLITANPYDYAFISQGETQVASIDDSDELM 329
Dr.smyhc5 SRVTFQLKAERDYHIFYQILSQKKPELLEMLLITANPYDYAFISQGETQVASINDADELM 328
Dr.myh7 SRVTFQLKAERDYHIFYQILSQRKPELLEMLLITNNPYDYAYISQGETTVASINDGEELL 329
Dr.myh7l SRVTFQLKAERDYHIFYQILSNKKPEILEMLLVTSNPYDYAFISQGETTVPSIDDSDELM 328
Dr.myh6 SRVTFQLKSERNYHIFFQILSNEKPELLDMLLITNNPYDYSYISQGEVTVSSINDNEELI 329
Dr.myha SRVTFQLPDERGYHIFYQMMTNHKPELIEMTLITTNPYDFPMCSQGQITVASIDDKEELV 334
Dr.myhb SRVTFQLSAERSYHIFYQLCTGHKPELLEALLITTNPYDYPMISQGEITVKSINDVEEFI 332
Dr.myhz1.1 SRVTFQLPDERGYHIFYQMMTNHKPELIEMTLITTNPYDFPMCSQGQITVASIDDKEELV 334
Dr.myhz1.2 SRVTFQLPDERGYHIFYQMMTNHKPELIEMTLITTNPYDFPMCSQGQITVASIDDKEELV 334
Dr.myhz1.3 SRVTFQLPDERGYHIFYQMMTNHKPELIEMTLITTNPYDFPMCSQGQITVASIDDKEELV 334
Dr.myhz2 SRVTFQLPDERGYHIFYQMMTNHKPELIEMTLITTNPYDFPMCSQGQITVASIDDKEELV 334
Dr.myhc4 SRVTFQLPDERGYHIFYQMMTNHKPELIEMTLITTNPYDFPMCSQGQITVASIDDKEELM 334
Dr.myh7ba SRVIFQQPGERSYHIYYQIMSQKKPELLDMLLVSSNPYDYHFCSQGVTTVENMDDGQELM 327
Dr.myh7bb SRVIFQQTGERSYHIYYQILSHRKPELQDMLLVSSNPFDYHFCSQGVITVDNMDDGDELL 327
Dr.myh9a SRAIRQAKDERAFHIFYYLLTGAGDKLRSELCL-EDYNKYRFLSNGNVTIPGQQDRELFA 328
Dr.myh9b SRAIRQAKEERTFHMFYYMLTGVGDKLRSELCL-EGYNKYRFLSNGNVTIPGQQDRDMYV 325
Dr.myh10 SRAIRQAKDERTFHVFYQLLAGAGEHLRSDLLL-EGFNSYRFLSNGNIPIPGQQDKDNFQ 340
Dr.myh11a SRCIRQAKTERAFHIFYYMVAGTKDKLREELLL-ENFNNYRFLSAGHVQIPGNQDDEMYD 325
Dr.myh11b SRCIRQAKIERSFHIFYYMVAGAKDKMREELLL-EDFANYRFLVAGHVQVQNQQDDEMLE 309
Dr.myh14 SRAIRQAKDERTFHIFYQLLSGATEAMRKELLL-GGADQYRFLCGGSLPVPGQSDSENFT 343
Hs.MYH7 ATDNAFDVLGFTSEEKNSMYKLTGAIMHFGNMKFK-LKQREEQAEPDGTEEADKSAYLMG 389
Loop 4
Hs.MYH7 ATDNAFDVLGFTSEEKNSMYKLTGAIMHFGNMKFK-LKQREEQAEPDGTEEADKSAYLMG 389
Hs.MYH6 ATDSAFDVLGFTSEEKAGVYKLTGAIMHYGNMKFK-QKQREEQAEPDGTEEADKSAYLMG 390
Hs.MYH13 ATDNAIDILGFSSEEKVGIYKLTGAVMHYGNMKFK-QKQREEQAEPDGTEVADKAGYLMG 391
Hs.MYH8 ATDSAIDILGFTPEEKVSIYKLTGAVMHYGNMKFK-QKQREEQAEPDGTEVADKAAYLQS 392
Hs.MYH4 ATDSAVDILGFTADEKVAIYKLTGAVMHYGNMKFK-QKQREEQAEPDGTEVADKAAYLTS 392
Hs.MYH1 ATDSAIEILGFTSDERVSIYKLTGAVMHYGNMKFK-QKQREEQAEPDGTEVADKAAYLQN 392
Hs.MYH2 ATDSAIDILGFTNEEKVSIYKLTGAVMHYGNLKFK-QKQREEQAEPDGTEVADKAAYLQS 392
Hs.MYH3 ATDSAIDILGFTPEEKSGLYKLTGAVMHYGNMKFK-QKQREEQAEPDGTEVADKTAYLMG 390
Hs.MYH14 ETLESLRVLGFSHEEIISMLRMVSAVLQFGNIALKRERNTDQATMPD-NTAAQKLCRLLG 405
Hs.MYH15 ATEQAMDILGFLPDEKYGCYKLTGAIMHFGNMKFK-QKPREEQLEADGTENADKAAFLMG 381
Hs.MYH16 PT-RAIS----------RNLVSSHSKQPR--------EATTSSTRFSQ 278
Dr.smyhc1 ATDDAFDVLGFTQDEKSGIYKLTGAIMHFGNMKFK-QKQREEQAEADGTEDADKVAYLMG 388
Dr.smyhc2 ATDEAFDVLGFTQEEKNSIYKLTGAIMHYGNMKFK-QKQREEQAEADGTEDADKVAYLMG 389
Dr.smyhc3 ATDEAFDVLGFTQEEKNSIYKLTGAIMHYGNMKFK-QKQREEQAEADGTEDADKSAYLMG 387
Dr.smyhc4 ATDEAFDVLGFTQEEKNSIYKLIGAIMHYGNMKFK-QKQREEQAEADGTEDADKSAYLMG 388
Dr.smyhc5 ATDEAFDVLGFTQEEKNSIYKLIGAIMHYGNMKFK-QKQREEQAEADGTEDADKSAYLMG 387
Dr.myh7 ATDEAFDVLGFTQEEKNGIYKLIGAIMHFGNMKFK-QKQREEQAEADGTEDGDKVAYLMG 388
Dr.myh7l ATDSAFDILGFTQEEKNSVYKLTGAIMHYGNMKFK-QKQREEQAEADGTEDADKSAYLMG 387
Dr.myh6 ATDKAFDVLGFTSEEKMGVYKLTGAIMHYGNMKFK-QKQREEQAEPDGTEDADKAAYLMG 388
Dr.myha ATDTAIDILGFTGEEKMGIYKFTGAVLHHGNMKFK-QKQREEQAEPDGTEEADKISYLLG 393
Dr.myhb ATDTAIDILGFNAEEKVGIYKLTGAVMHHGNMKFK-QKQREEQAEPDGTEVADKIAYLLG 391
Dr.myhz1.1 ATDTAIDILGFNNEEKMGIYKFTGAVLHHGNMKFK-QKQREEQAEPDGTEEADKIGYLLG 393
Dr.myhz1.2 ATDTAIDILGFNNEEKMGIYKFTGAVLHHGNMKFK-QKQREEQAEPDGTEEADKIGYLLG 393
Dr.myhz1.3 ATDTAIDILGFNAEEKMGIYKFTGAVLHHGNMKFK-QKQREEQAEPDGTEEADKIAYLLG 393
Dr.myhz2 ATDTAIDILGFTGEEKMGIYKFTGAVLHHGNMKFK-QKQREEQAEPDGTEEADKIAYLLG 393
Dr.myhc4 ATDSAIDILGFTGEEKMGIYKFTGAVLHHGNMKFK-QKQREEQAEPDGTEEADKISYLLG 393
Dr.myh7ba ATDHAMDILGFTPEEKYGCYKIVGAIMHFGNMKFK-QKQREEQAEADGTESADKASYLMG 386
Dr.myh7bb ATDHAMDTLGFTPEEKYGCYKIVGGIMHFGNMKFK-VKQREEQAEADGTESADKASYLMG 386
Dr.myh9a ETIDAFRIMGIPEDEQTGLLKVVSAVLQLGNMSFKKERNSDQASMPD-DTAAQKVSHLLG 387
Dr.myh9b ETVEAMRIMGFSEEEHVGLLRVISSVLQLGNMSFKKERHSDQASMPD-DTAAQKVCHLMG 384
Dr.myh10 ETMEAMHIMSFNHEEILSMLKVVSAVLQFGNIVFKKERNTDQASMPE-NTAAQKLCHLLG 399
Dr.myh11a ETMEAMEIMGFSVEERADVLKVVSTVLQLGNIEFKKERNQEQATMPD-NTAAQKVCHLQG 384
Dr.myh11b ETLEAMEVLGFNEEERIGMFKICSTVLQLGNIEFKAEKNQEQASMPD-NTAAQKVCHLQG 368
Dr.myh14 QTMDSMTIMGFTQEESTSMLKVISSVLQFGNITFHKEKNTDQASMPD-DTAAQKLCHLLG 402


HCM loop
Switch 2

```
Hs.MYH7          FNSFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDFGMDLQACIDLI 524
                                 Relay
Hs.MYH7          FNSFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDFGMDLQACIDLI 524
Hs.MYH6          FNSFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDFGMDLQACIDLI 525
Hs.MYH13         FNSLEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWEFIDFGMDLAACIELI 526
Hs.MYH8          FNSLEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDFGMDLAACIELI 527
Hs.MYH4          FNSLEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWEFIDFGMDLAACIELI 527
Hs.MYH1          FNSLEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDFGMDLAACIELI 527
Hs.MYH2          FNSLEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDFGMDLAACIELI 527
Hs.MYH3          YNSLEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDFGMDLAACIELI 525
Hs.MYH14         LNSFEQLCINY-----TNEKLQQLFNHTMFVLEQEEYQREGIPWTFLDFGLDLQPCIDLI 541
Hs.MYH15         YNSLEQLCINF-----TNEKLQQFFNWHMFVLEQEEYKKESIEWVSIGFGLDLQACIDLI 516
Hs.MYH16         -KLAMSLCKKARTWNSAKTPLGLWARLSMTRCSSGWWPGLTRPWTPRCRGSSSLECWTSL 447
Dr.smyhc1        FNTFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIDWEFIDFGMDLQACIELI 523
Dr.smyhc2        FNTFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWEFIDFGMDLQACIELI 524
Dr.smyhc3        FNTFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDFGMDLQACIDLI 522
Dr.smyhc4        FNTFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIDWEFIDFGMDLQACIDLI 523
Dr.smyhc5        FNTFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWTFIDFGMDLQACIDLI 522
Dr.myh7          FNTFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWEFIDFGMDLQACIDLI 523
Dr.myh7l         FNTFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWEFIDFGMDLQACIDLI 522
Dr.myh6          FNTFEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIDWEFIDFGMDLQSCIDLI 523
Dr.myha          FNSMEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIVWEFIDFGMDLAACIELI 528
Dr.myhb          FNSLEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIEWEFIDFGMDLAACIELI 526
Dr.myhz1.1       FNSMEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIVWEFIDFGMDLAACIELI 528
Dr.myhz1.2       FNSMEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIVWEFIDFGMDLAACIELI 528
Dr.myhz1.3       FNSMEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIVWEFIDFGMDLAACIELI 528
Dr.myhz2         FNSMEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIVWEFIDFGMDLAACIELI 528
Dr.myhc4         FNSMEQLCINF-----TNEKLQQFFNHHMFVLEQEEYKKEGIVWEFIDFGMDLAACIELI 528
Dr.myh7ba        LNSFEQLCINF-----TNEKLQQFFNHHMFILEQEEYKREGIEWTFIDFGLDLQACIDLI 521
Dr.myh7bb        FNNFEQMCINF-----TNEKLQQFFNHHMFILEQEEYKTEGIEWTFIDFGLDLQACIDLI 521
Dr.myh9a         LNSFEQLCINY-----TNEKLQQLFNHTMFILEQEEYQREGIEWSFIDFGLDLQPCIELI 523
Dr.myh9b         LNSFEQLCINY-----TNEKLQQLFNHTMFILEQEEYQREGIEWSFIDFGLDLQPCIDLI 520
Dr.myh10         LNSFEQLCINY-----TNEKLQQLFNHTMFILEQEEYQREGIEWSFIDFGLDLQPCIDLI 535
Dr.myh11a        NNSFEQLCINY-----TNEKLQQLFNHTMFILEQEEYQREGIEWNFIDFGLDLQPCIELI 520
Dr.myh11b        DNSFEQLCINY-----TNEKLQQLFNHTMFILEQEEYKKEGIEWSFIDFGLDLQPCIELI 504
Dr.myh14         LNSFEQLCINY-----TNEKLQQLFNHTMFVLEQEEYQREGIEWNFIDFGLDLQPCIDLI 538
                  :   .:* :      ::  *    .  *   ..  :      *     * .   *   :

Hs.MYH7          E-KPMG-----I-----------MSILEEECMF-------------------------- 540

Hs.MYH7          E-KPMG-----I-----------MSILEEECMF-------------------------- 540
Hs.MYH6          E-KPMG-----I-----------MSILEEECMF-------------------------- 541
Hs.MYH13         E-KPMG-----I-----------FSILEEECMF-------------------------- 542
Hs.MYH8          E-KPLG-----I-----------FSILEEECMF-------------------------- 543
Hs.MYH4          E-KPMG-----I-----------FSILEEECMF-------------------------- 543
Hs.MYH1          E-KPMG-----I-----------FSILEEECMF-------------------------- 543
Hs.MYH2          E-KPMG-----I-----------FSILEEECMF-------------------------- 543
Hs.MYH3          E-KPMG-----I-----------FSILEEECMF-------------------------- 541
Hs.MYH14         E-RPANPPG--L-----------LALLDEECWF-------------------------- 560
Hs.MYH15         E-KPMG-----I-----------LSILEEECMF-------------------------- 532
Hs.MYH16         ALRSLSSTALSSYASTSPTRSCSSSSTTTCSCWSRRSTRGKASSGSSSTLASTFRPASTC 507
Dr.smyhc1        E-KPMG-----I-----------MSILEEECMF-------------------------- 539
Dr.smyhc2        E-KPMG-----I-----------MSILEEECMF-------------------------- 540
Dr.smyhc3        E-KPMG-----I-----------MSILEEECMF-------------------------- 538
Dr.smyhc4        E-KPMG-----I-----------MSILEEECMF-------------------------- 539
Dr.smyhc5        E-KPMG-----I-----------MSILEEECMF-------------------------- 538
Dr.myh7          E-KPMG-----I-----------MSILEEECMF-------------------------- 539
Dr.myh7l         E-KPMG-----I-----------MSILEEECMF-------------------------- 538
Dr.myh6          E-KPLG-----I-----------MSILEEECMF-------------------------- 539
Dr.myha          E-KPMG-----I-----------FSILEEECMF-------------------------- 544
Dr.myhb          E-KPMG-----I-----------FSILEEECMF-------------------------- 542
Dr.myhz1.1       E-KPLG-----I-----------FSILEEECMF-------------------------- 544
Dr.myhz1.2       E-KPLG-----I-----------FSILEEECMF-------------------------- 544
Dr.myhz1.3       E-KPLG-----I-----------FSILEEECMF-------------------------- 544
Dr.myhz2         E-KPLG-----I-----------FSILEEECMF-------------------------- 544
Dr.myhc4         E-KPMG-----I-----------FSILEEECMF-------------------------- 544
Dr.myh7ba        E-KPLG-----I-----------MSILEEECMF-------------------------- 537
Dr.myh7bb        E-KPLG-----I-----------LSILEEECMF-------------------------- 537
Dr.myh9a         E-KPNGPPG--I-----------LALLDEECWF-------------------------- 542
Dr.myh9b         E-KPASPPG--I-----------LALLDEECWF-------------------------- 539
Dr.myh10         E-RPANPPG--V-----------LALLDEECWF-------------------------- 554
Dr.myh11a        E-RPNNPPG--I-----------LALLDEECWF-------------------------- 539
Dr.myh11b        E-RPNNPPG--I-----------LALLDEECWF-------------------------- 523
Dr.myh14         E-RPAHPPG--V-----------LALLDEECWF-------------------------- 557
                   :                     :    .*
```

```
Hs.MYH7          ---PKATDMTFKAKLFD--------------NHLGKSANFQKPRNIK-GKPEAHFSLIHY 582
                                                 Loop 3
Hs.MYH7          ---PKATDMTFKAKLFD--------------NHLGKSANFQKPRNIK-GKPEAHFSLIHY 582
Hs.MYH6          ---PKATDMTFKAKLYD--------------NHLGKSNNFQKPRNIK-GKQEAHFSLIHY 583
Hs.MYH13         ---PKATDTSFKNKLYD--------------QHLGKSNNFQKPKPAK-GKAEAHFSLVHY 584
Hs.MYH8          ---PKATDTSFKNKLYD--------------QHLGKSANFQKPKVVK-GKAEAHFSLIHY 585
Hs.MYH4          ---PKATDTSFKNKLYE--------------QHLGKSNNFQKPKPAK-GKPEAHFSLVHY 585
Hs.MYH1          ---PKATDTSFKNKLYE--------------QHLGKSNNFQKPKPAK-GKPEAHFSLIHY 585
Hs.MYH2          ---PKATDTSFKNKLYD--------------QHLGKSANFQKPKVVK-GKAEAHFALIHY 585
Hs.MYH3          ---PKATDTSFKNKLYD--------------QHLGKSNNFQKPKVVK-GRAEAHFSLIHY 583
Hs.MYH14         ---PKATDKSFVEKVAQ---------------EQGGHPKFQRPRHLR---DQADFSVLHY 599
Hs.MYH15         ---PKATDLTFKTKLFD--------------NHFGKSVHLQKPKPDK-KKFEAHFELVHY 574
Hs.MYH16         WKSPWASSPSWRNSASSPKPPMPRSRQPCTTTTWASPATS-SPRGARARGP--RSTSSWF 564
Dr.smyhc1        ---PKASDQTFKAKLYD--------------NHLGKNPTFQKPRIVK-GRPEAHFALVHY 581
Dr.smyhc2        ---PKASDATFKAKLYD--------------NHLGKNPNFQKPRIVK-GRPEAHFALVHY 582
Dr.smyhc3        ---PKASDATFKAKLYD--------------NHLGKSNNFQKPRIVK-GKPEAHFSLVHY 580
Dr.smyhc4        ---PKASDATFKAKLYD--------------NHLGKSNNFQKPRIVK-GKPEAHFSLVHY 581
Dr.smyhc5        ---PKASDATFKAKLYD--------------NHLGKSNNFQKPRIVK-GKPEAHFSLVHY 580
Dr.myh7          ---PKASDSTFKAKLYD--------------NHLGKSNNFQKPRAIK-GKPESHFSLVHY 581
Dr.myh7l         ---PKASDATFKAKLYD--------------NHLGKSNNFQKPRLVK-GKPEAHFALVHY 580
Dr.myh6          ---PKASDQTFKAKLYD--------------NHLGKTNIFQKPRAVK-GKAEAHFALSHY 581
Dr.myha          ---PKATDTSFKNKLYD--------------QHLGKCNAFQKPRPQK-GKAEAHFSLVHY 586
Dr.myhb          ---PKATDTTFKNKLHD--------------QHLGKTNCFQKPKPAK-GKAEAHFSLVHY 584
Dr.myhz1.1       ---PKATDTSFKNKLYD--------------QHLGKCNAFQKPKPAK-GKAEAHFSLVHY 586
Dr.myhz1.2       ---PKATDTSFKNKLYD--------------QHLGKCNAFQKPKPAK-GKAEAHFSLVHY 586
Dr.myhz1.3       ---PKATDTSFKNKLYD--------------QHLGKCNAFQKPKPAK-GKAEAHFSLVHY 586
Dr.myhz2         ---PKATDTSFKNKLYD--------------QHLGKCNAFQKPKPAK-GKAEAHFSLVHY 586
Dr.myhc4         ---PKATDVSFKNKLYD--------------QHLGKCNAFQKPRPQK-GKAEAHFSLVHY 586
Dr.myh7ba        ---PKATDNSFKAKLFD--------------NHLGKSANFQKPRPDKKRKYEAHFELVHY 580
Dr.myh7bb        ---PKATESSFKAKLYD--------------NLLGKSPNFLKPRPDKKRKYDTHFELVHY 580
Dr.myh9a         ---PKATDKSFVEKVVQ---------------ELGNNPKFQKPKKLK---DDADFCIIHY 581
Dr.myh9b         ---PKATDKSFVEKVLQ---------------EQGTHPKFHKPKKLK---DEADFCIIHY 578
Dr.myh10         ---PKATDKTFVDKLVQ---------------EQGTHGKFQKPRQLK---DKADFCIIHY 593
Dr.myh11a        ---PKATDVSFVEKLCN---------------THANHTKFAKPKQLK---DKTEFSVQHY 578
Dr.myh11b        ---PKATDVSFVEKLTN---------------THSSHCKFSKPKNLK---EKTFFTVQHY 562
Dr.myh14         ---PRATDRSFVDKLSA---------------EQGSHSKFMRPRQLK---EEADFSIIHY 596
                    * *:. ::  .                    .       *:  :            :

Hs.MYH7          AGIVDYNI---IGWLQKNKDPLNETVVGLYQKSSLKLLSTLFANYAGADAP-I------- 631
                                                 Loop 2
Hs.MYH7          AGIVDYNI---IGWLQKNKDPLNETVVGLYQKSSLKLLSTLFANYAGADAP-I------- 631
Hs.MYH6          AGTVDYNI---LGWLEKNKDPLNETVVALYQKSSLKLMATLFSSYATADTGDS------- 633
Hs.MYH13         AGTVDYNI---AGWLDKNKDPLNETVVGLYQKSSLKLLSFLFSNYAGAET-GDSG----- 635
Hs.MYH8          AGTVDYNI---TGWLDKNKDPLNDTVVGLYQKSAMKTLASLFSTYASAEA---DS----- 634
Hs.MYH4          AGTVDYNI---AGWLDKNKDPLNETVVGLYQKSAMKTLAFLFSGAQTAEAE--GG----- 635
Hs.MYH1          AGTVDYNI---AGWLDKNKDPLNETVVGLYQKSAMKTLALLFVGATGAEAE--AG----- 635
Hs.MYH2          AGVVDYNI---TGWLEKNKDPLNETVVGLYQKSAMKTLAQLFSGAQTAEGEGAGG----- 637
Hs.MYH3          AGTVDYSV---SGWLEKNKDPLNETVVGLYQKSSNRLLAHLYATFATADA---DS----- 632
Hs.MYH14         AGKVDYKA---NEWLMKNMDPLNDNVAALLHQSTDRLTAEIWKDVEGIVGLEQVSSLGDG 656
Hs.MYH15         AGVVPYNI---SGWLEKNKDLLNETVVAVFQKSSNRLLASLFENYMSTDSAIPF------ 625
Hs.MYH16         TTQAPWDITSQAGWRRTKTP-MK----QWWACSRNRVWQSW--PFSSKKRRLQP------ 611
Dr.smyhc1        AGTVDYNI---SNWLVKNKDPLNETVVGLFQKSTVKLLSFLFAGYSGADSAQDSK----- 633
Dr.smyhc2        AGTVDYNI---SNWLVKNKDPLNETVVGLFQKSTLKLLGTLFANYAGAESADSGG----- 634
Dr.smyhc3        AGTVDYNI---NNWLVKNKDPLNETVVGLYQKSTMKMLSILFANYAGAESAAEGG----- 632
Dr.smyhc4        AGTVDYNI---NNWLVKNKDPLNETVVGLYQKSTMKLLSNLFAGYAGAE---SGG----- 630
Dr.smyhc5        AGTVDYNI---NNWLVKNKDPLNETVVGLYQKSTMKLLSNLFANYTGADLAMEGG----- 632
Dr.myh7          AGTVDYNI---NNWLVKNKDPLNETVVGLFQKSTVKLLSMLFANYAGTESD-NGK----- 632
Dr.myh7l         AGTVDYNI---NNWLVKNKDPLNETVVGLYQKSSLKLLSNLFANYAGADSATGDG----- 632
Dr.myh6          AGTVDYNI---AGWLVKNKDPLNETVVGLYQKSSLKLLSLLFSSYAGSDGGEKSG----- 633
Dr.myha          AGTVDYNV---SGWLDKNKDPLNESVVQLYQKSSVKLLATLYPPV--VED---T------ 632
Dr.myhb          AGTVDYNI---SGWLDKNKDPLNDSVVQLYQKSSVKLLCHLYAAHASTEA---ES----- 633
Dr.myhz1.1       AGTVDYNI---SGWLDKNKDPLNESVVQLYQKSSVKLLATLYPPV--VEE---TG----- 633
Dr.myhz1.2       AGTVDYNI---SGWLDKNKDPLNESVVQLYQKSSVKLLATLYPPV--VEE---TG----- 633
Dr.myhz1.3       AGTVDYNI---SGWLDKNKDPLNESVVQLYQKSSVKLLATLYPPV--VEE---TG----- 633
Dr.myhz2         AGTVDYNI---SGWLDKNKDPLNESVVQLYQKSSVKLLATLYPPV--VEE---TG----- 633
Dr.myhc4         AGTVDYNV---NGWLDKNKDPLNESVVQLYQKSSVKLLATLYPPV--VEE---TG----- 633
Dr.myh7ba        AGVVPYNI---IGWLDKNKDPLNETVVICFQKSSNKLLASLYEKYVSSDSASDPK----- 632
Dr.myh7bb        AGVVPYNI---NGWLDKNRDPLNETVVGIFQRSSNKLMSSLFENFISLDSGSEAK----- 632
Dr.myh9a         AGKVDYKA---NEWLMKNMDPLNDNVATLLNQSVDKFVSELWKDVDRIVGLDKVAGMGE- 637
Dr.myh9b         AGKVDYKA---DEWLMKNMDPLNDNVATLLNQSTDRFVSELWKDVDRIVGLDKVAGMSE- 634
Dr.myh10         AGRVDYKA---DEWLMKNMDPLNDNVATLLHQSTDKFVAELWKDVDRIVGLDQVAGMNE- 649
Dr.myh11a        AGRVDYNA---VAWLTKNMDPLNDNVTALLNNSSNPFVQDLWKDADRVVGLETIAKMSD- 634
Dr.myh11b        AGKVDYNA---MSWLTKNMDPLNDNVTALLSNSSSAFIQDIWKDVDRVVGLETMAKMAKS 619
Dr.myh14         AGKVDYKA---KEWLVKNMDPLNDNVASLLHQSSDPFISELWREVERIVGLDQVSSGENS 653
                 :  . :.      *  .:   ::         *
```

```
Hs.MYH7          ----EKGKGKAKKGSSFQTVSALHRENLNKLMTNLRSTHPHFVRCIIPNETKSPGVMDNP 687
                     Loop 2                  Actin binding domain          Converter
Hs.MYH7          ----EKGKGKAKKGSSFQTVSALHRENLNKLMTNLRSTHPHFVRCIIPNETKSPGVMDNP 687
Hs.MYH6          ----GKSKGGKKKGSSFQTVSALHRENLNKLMTNLRTTHPHFVRCIIPNERKAPGVMDNP 689
Hs.MYH13         ----GSKKGGKKKGSSFQTVSAVFRENLNKLMTNLRSTHPHFVRCLIPNETKTPGVMDHY 691
Hs.MYH8          ----SAKKGAKKKGSSFQTVSALFRENLNKLMTNLRSTHPHFVRCIIPNETKTPGAMEHE 690
Hs.MYH4          ----GGKKGGKKKGSSFQTVSALFRENLNKLMTNLRSTHPHFVRCIIPNETKTPGAMEHE 691
Hs.MYH1          ----GGKKGGKKKGSSFQTVSALFRENLNKLMTNLRSTHPHFVRCIIPNETKTPGAMEHE 691
Hs.MYH2          ----GAKKGGKKKGSSFQTVSALFRENLNKLMTNLRSTHPHFVRCIIPNETKTPGAMEHE 693
Hs.MYH3          ----GKKKVAKKKGSSFQTVSALFRENLNKLMSNLRTTHPHFVRCIIPNETKTPGAMEHS 688
Hs.MYH14         ------PPGGRPRRGMFRTVGQLYKESLSRLMATLSNTNPSFVRCIVPNHEKRAGKLEPR 710
Hs.MYH15         ------GEKKRKKGASFQTVASLHKENLNKLMTNLKSTAPHFVRCINPNVNKIPGILDPY 679
Hs.MYH16         -----EARSRREAPPS-Q--SPISTGSS-TS--PPSIAAPHFVRCIIPNEFKQSGVIDAH 660
Dr.smyhc1        ---GGKGGG-KKKGSSFQTVSALHRENLNKLMTNLRSTHPHFVRCLIPNETKTPGAMENP 689
Dr.smyhc2        ---KGKGGA-KKKGSSFQTVSALHRENLNKLMTNLRSTHPHFVRCLIPNETKTPGAMENP 690
Dr.smyhc3        ---GGKGKEKKKKGSSFQTVSALHRENLNKLMTNLRSTHPHFVRCIIPNETKTPGAMENP 689
Dr.smyhc4        ---GGKGKEKKKKGSSFQTVSALHRENLNKLMTNLRSTHPHFVRCIIPNETKTPGAMENP 687
Dr.smyhc5        ---GGKTKEKKKKGSSFQTVSALHRENLNKLMTNLRSTHPHFVRCIIPNETKTPGAMENP 689
Dr.myh7          ---GGKGGGSKKKGSSFQTVSALHRENLNKLMTNLRSTHPHFVRCIIPNETKTPGAMENP 689
Dr.myh7l         ---G--KKEKKKKGSSFQTVSALHRENLNKLMTNLRSTHPHFVRCLIPNETKTPGAMENP 687
Dr.myh6          ------GKGAKKKGSSFQTVSALHRENLNKLMTNLKTTHPHFVRCLIPNESKIPGIMDNC 687
Dr.myha          -----SKKGGKKKGGSMQTVSSQFRENLGKLMTNLRSTHPHFVRCLIPNESKTPGLMENF 687
Dr.myhb          ----GGKKGGKKKGGSFQTVSALFRENLGKLMTNLRSTHPHFVRCLIPNESKTPGLMENF 689
Dr.myhz1.1       ----GGKKGGKKKGGSMQTVSSQFRENLGKLMTNLRSTHPHFVRCLIPNESKTPGLMENF 689
Dr.myhz1.2       ----GGKKGGKKKGGSMQTVSSQFRENLGKLMTNLRSTHPHFVRCLIPNESKTPGLMENF 689
Dr.myhz1.3       ----GGKKGGKKKGGSMQTVSSQFRENLGKLMTNLRSTHPHFVRCLIPNESKTPGLMENF 689
Dr.myhz2         ----GGKKGGKKKGGSMQTVSSQFRENLGKLMTNLRSTHPHFVRCLIPNESKTPGLMENF 689
Dr.myhc4         ----GGKKGGKKKGGSMQTVSSQFRENLGKLMTNLRSTHPHFVRCLIPNESKTPGLMENF 689
Dr.myh7ba        ----TGGKEKRKKAASFQTVSQLHKENLNKLMTNLRSTQPHFVRCIIPNETKTPGIMDSF 688
Dr.myh7bb        ----PGSKEKRKKGASFQTVSQLHKENLNKLMTNLRSTQPHFVRCIIPNEAKNPGMMEPF 688
Dr.myh9a         ---SLHG-AVKTRKGMFRTVGQLYKEQLMNLMTTLRNTNPNFVRCIIPNHEKKAGKLAHH 693
Dr.myh9b         ----LPG-AFKTRKGMFRTVGQLYKEQLSKLMATLRNTNPNFVRCIIPNHEKKAGKLDPH 689
Dr.myh10         ---TAFGAAYKTKKGMFRTVGQLYKESLTKLMATLRNTNPNFVRCIIPNHEKRAGKLEPH 706
Dr.myh11a        ---SLAPSASKTKKGMFRTVGQLYKESLAKLMTTLHNTQPNFVRCIIPNHEKRAGKLDAH 691
Dr.myh11b        D--SSAPAASKSKKGMFRTVGQLYKESLGKLMTTLHNTQPNFVRCIIPNHEKRAGKIDAH 677
Dr.myh14         GPVSFGAAGLKTKKGMFRTVGQLYKESLTKLMATLRNTNPNFLRCIIPNHEKRAGKLSPH 713
                                 :  .    .          : * *:**: **  *  * :

Hs.MYH7          LVMHQLRCNGVLEGIRICRKGFPNRILYGDFRQRYRILNPAAIPEGQFIDSRKGAEKLLS 747
                 Converter
Hs.MYH7          LVMHQLRCNGVLEGIRICRKGFPNRILYGDFRQRYRILNPAAIPEGQFIDSRKGAEKLLS 747
Hs.MYH6          LVMHQLRCNGVLEGIRICRKGFPNRILYGDFRQRYRILNPVAIPEGQFIDSRKGTEKLLS 749
Hs.MYH13         LVMHQLRCNGVLEGIRICRKGFPSRILYADFKQRYRILNASAIPEGQFIDSKNASEKLLN 751
Hs.MYH8          LVLHQLRCNGVLEGIRICRKGFPSRILYGDFKQRYKVLNASAIPEGQFIDSKKASEKLLA 750
Hs.MYH4          LVLHQLRCNGVLEGIRICRKGFPSRILYADFKQRYKVLNASAIPEGQFIDSKKASEKLLG 751
Hs.MYH1          LVLHQLRCNGVLEGIRICRKGFPSRILYADFKQRYKVLNASAIPEGQFIDSKKASEKLLG 751
Hs.MYH2          LVLHQLRCNGVLEGIRICRKGFPSRILYADFKQRYKVLNASAIPEGQFIDSKKASEKLLA 753
Hs.MYH3          LVLHQLRCNGVLEGIRICRKGFPNRILYGDFKQRYRVLNASAIPEGQFIDSKKACEKLLA 748
Hs.MYH14         LVLDQLRCNGVLEGIRICRQGFPNRILFQEFRQRYEILTPNAIPKG-FMDGKQACEKMIQ 769
Hs.MYH15         LVLQQLRCNGVLEGTRICREGFPNRLQYADFKQRYCILNPRTFPKSKFVSSRKAAEELLG 739
Hs.MYH16         LIMHQLACNGVLEGIRICRKGFPNRLQYPEFKQRYQVLNPNVIPQG-FVDNKKASELLLA 719
Dr.smyhc1        LVMHQLRCNGVLEGIRICRKGFPNRILYGDFKQRYRILNPSANPEGQFIDNKKAAEKLLG 749
Dr.smyhc2        LVMHQLRCNGVLEGIRICRKGFPNRILYGDFKQRYRILNPSAIPEGQFIDNKKGAEKLLG 750
Dr.smyhc3        LVMHQLRCNGVLEGIRICRKGFPNRILYGDFKQRYRILNPAAIPEGQFIDSRKGAEKLLG 749
Dr.smyhc4        LVMHQLRCNGVLEGIRICRKGFPNRILYGDFKQRYRILNPAAIPEGQFIDSRKGAEKLLG 747
Dr.smyhc5        LVMHQLRCNGVLEGIRICRKGFPNRILYGDFKQRYRILNPAAIPDGQFIDSRKGAEKLLG 749
Dr.myh7          LVMHQLRCNGVLEGIRICRKGFPNRILYGDFKQRYRILNPAAIPEGQFIDSRKGAEKLLG 749
Dr.myh7l         LVMHQLRCNGVLEGIRICRKGFPNRILYGDFKQRYRILNPAAIPEGQFIDSKKGAEKLLG 747
Dr.myh6          LVMHQLRCNGVLEGIRICRKGFPNRILYGDFKQRYRILNASAIPEGQFIENKKSAEKLLG 747
Dr.myha          LVIHQLRCNGVLEGIRICRKGFPSRILYADFKQRYKVLNASVIPEGQFIDNKKASEKLLG 747
Dr.myhb          LVIHQLRCNGVLEGIRICRKGFPSRILYGDFKQRYKVLNASVIPEGQFIDNKKASEKLLG 749
Dr.myhz1.1       LVIHQLRCNGVLEGIRICRKGFPSRILYADFKQRYKVLNASVIPEGQFIDNKKASEKLLG 749
Dr.myhz1.2       LVIHQLRCNGVLEGIRICRKGFPSRILYADFKQRYKVLNASVIPEGQFIDNKKASEKLLG 749
Dr.myhz1.3       LVIHQLRCNGVLEGIRICRKGFPSRILYGDFKQRYKVLNASVIPEGQFIDNKKASEKLLG 749
Dr.myhz2         LVIHQLRCNGVLEGIRICRKGFPSRILYGDFKQRYKVLNASVIPEGQFIDNKKASEKLLG 749
Dr.myhc4         LVIHQLRCNGVLEGIRICRKGFPSRILYGDFKQRYKVLNASVIPEGQFIDNKKASEKLLG 749
Dr.myh7ba        MVLHQLRCNGVLEGIRICRKGFPNRMLYAEFKQRYRILNPHAIPDDKFVDSRKAAEKLLA 748
Dr.myh7bb        LVLHQLRCNGVLEGIRICRKGFPNRILYAEFKQRYRILNPLAIPEDTYVDSRKAVEKLLG 748
Dr.myh9a         LVLDQLRCNGVLEGIRICRQGFPNRIVFQEFRQRYEILTPNAIPKG-FMDGKQACVLMVK 752
Dr.myh9b         LVLDQLRCNGVLEGIRICRQGFPNRIVFQEFRQRYEILTPNSIPKG-FMDGKQACMLMIR 748
Dr.myh10         LVLDQLRCNGVLEGIRICRQGFPNRIVFQEFRQRYEILTPNAIPKG-FMDGKQACERMIR 765
Dr.myh11a        LVLEQLRCNGVLEGIRICRQGFPNRIVFQEFRQRYEILAANAIPKG-FMDGKQACCLMIK 750
Dr.myh11b        LVLDQLSCNGVLEGIRICRQGFPNRIVFHEFRQRYEVLAAGSIPKG-FMDGKQSCTLMIK 736
Dr.myh14         LVLDQLRCNGVLEGIRICRQGFPNRIPFQEFRQRYEILTPNAIPRT-FMDGKHASELMIS 772
                 :::.** ******* ****:***.*: : :*:*** :*     *   ::..::.   ::
```

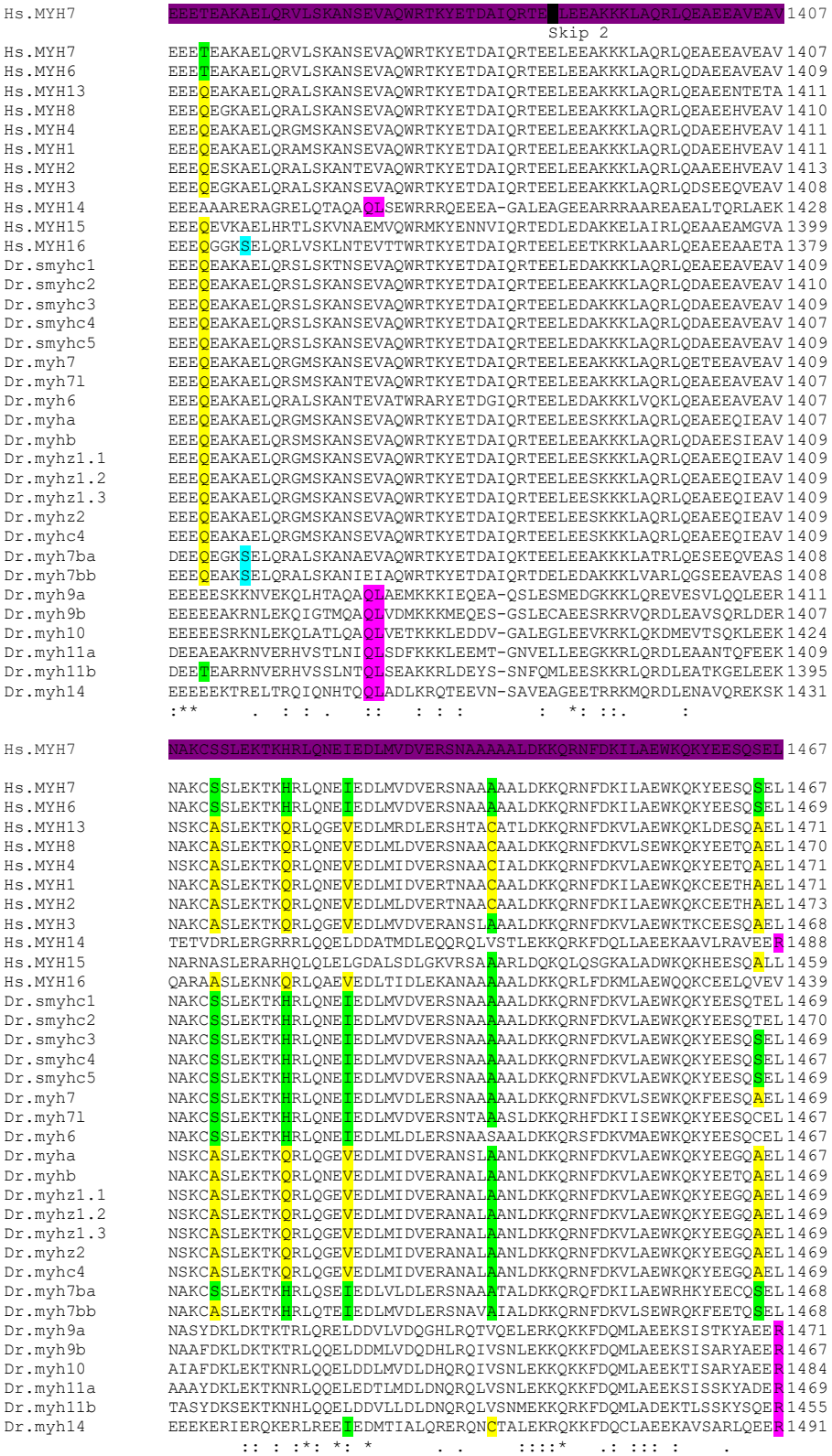
```
Hs.MYH7          SLDIDHNQYKFGHTKVFFKAGLLGLLEEMRDERLSRIITRIQAQSRGVLARMEYKKLLER 807
                           Actin binding domain  MYPBC binding site

Hs.MYH7          SLDIDHNQYKFGHTKVFFKAGLLGLLEEMRDERLSRIITRIQAQSRGVLARMEYKKLLER 807
Hs.MYH6          SLDIDHNQYKFGHTKVFFKAGLLGLLEEMRDERLSRIITRMQAQARGQLMRIEFKKIVER 809
Hs.MYH13         SIDVDREQFRFGNTKVFFKAGLLGLLEEMRDEKLVTLMTSTQAVCRGYLMRVEFKKMMER 811
Hs.MYH8          SIDIDHTQYKFGHTKVFFKAGLLGLLEEMRDEKLAQIITRTQAVCRGFLMRVEYQKMLQR 810
Hs.MYH4          SIEIDHTQYKFGHTKVFFKAGLLGTLEEMRDEKLAQLITRTQAICRGFLMRVEFRKMMER 811
Hs.MYH1          SIDIDHTQYKFGHTKVFFKAGLLGLLEEMRDEKLAQLITRTQAMCRGFLARVEYQKMVER 811
Hs.MYH2          SIDIDHTQYKFGHTKVFFKAGLLGLLEEMRDDKLAQLITRTQARCRGFLARVEYQRMVER 813
Hs.MYH3          SIDIDHTQYKFGHTKVFFKAGLLGTLEEMRDDRLAKLITRTQAVCRGFLMRVEFQKMVQR 808
Hs.MYH14         ALELDPNLYRVGQSKIFFRAGVLAQLEEERDLKVTDIIVSFQAAARGYLARRAFQKRQQQ 829
Hs.MYH15         SLEIDHTQYRFGITKVFFKAGFLGQLEAIRDERLSKVFTLFQARAQGKLMRIKFQKILEE 799
Hs.MYH16         AIDLDVNEYKIGHTKVFFRAGILARLEDMRDERLAKIMTMLQCRLRGFLMRVEFKKMLER 779
Dr.smyhc1        SLDIDHNQYKLGHTKVFFKAGLLGTLEEMRDDRLALIITGIQARARGILSRLEFQKIVER 809
Dr.smyhc2        SLDIDHNQYKLGHTKVFFKAGLLGTLEEMRDDRLALIITGIQARARGILSRIEFQKIVER 810
Dr.smyhc3        SLDIDHNQYKFGHTKVFFKAGLLGTLEEMRDDRLALIITNIQARARGLLSRVEFQKIVDR 809
Dr.smyhc4        SLDIDHNQYRFGHTKVFFKAGLLGTLEEMRDDRLALIITGIQARARGILSRIEFQKIVDR 807
Dr.smyhc5        SIDIDHNQYKFGHTKVFFKAGLLGTLEEMRDDRLALIITNIQARSRGLLSRIEFQKLVDR 809
Dr.myh7          SLDIDHNQYKFGHTKVFFKAGLLGQLEEMRDDRLSLIISGIQARSRGLLARVEFQKIVER 809
Dr.myh7l         SLDIDHQQYRFGHTKVFFKAGLLGQLEEMRDERLSKIITGIQAKSRGLLSRAEYIKMVER 807
Dr.myh6          SLDIDHTQYKFGHTKVFFKAGLLGTLEEMRDDQLARILTGIQAFARGLLMRVEYQKLVER 807
Dr.myha          SIDVNHDEYRFGHTKVFFKAGLLGTLEEMRDEKLASLVTMTQALCRAYLMRREFVKMMER 807
Dr.myhb          SIDVDHTQYKFGHTKVFFKAGLLGTLEEMRDEKLVILVTMTQAVCRGYVMRKEFVKMMAR 809
Dr.myhz1.1       SIDVNHDEYRFGHTKVFFKAGLLGTLEEMRDEKLASLVTMTQALCRAYLMRREFVKMMER 809
Dr.myhz1.2       SIDVNHDEYRFGHTKVFFKAGLLGTLEEMRDEKLASLVTMTQALCRAYLMRREFVKMTER 809
Dr.myhz1.3       SIDVNHDEYRFGHTKVFFKAGLLGTLEEMRDEKLASLVTMTQALCRAYLMRREFVKMMER 809
Dr.myhz2         SIDVNHDEYRFGHTKVFFKAGLLGTLEEMRDEKLATLVTMTQALCRAYLMRREFVKMMER 809
Dr.myhc4         SIDVNHDEYRFGHTKVFFKAGLLGTLEEMRDEKLATLVTMTQALCRAYLMRREFVKMMER 809
Dr.myh7ba        SLDIDHNQYRFGHTKVFFKAGLLGQLEEMRDERLAKILTLLQAASRGKIMRMELNKMTQR 808
Dr.myh7bb        SLDIDHTQYKFGHSKVFFKAGLLGQLEDMRDERLSKVLTLLQAFCRGKLMRMERKRMMKE 808
Dr.myh9a         ALELDSNLYRIGQSKVFFRAGVLAHLEEERDMKITDVIINFQAWCRGYVARRAFAKRQQQ 812
Dr.myh9b         ALELDPNLYRIGQSKVFFRAGVLAHLEEERDMKITDVIINFQAWCRGYVARKAFAKRQQQ 808
Dr.myh10         ALELDPNLYRIGQSKIFFRTGVLAHLEEERDLKITDIIIYFQSVCRGYLARKAFAKKQQQ 825
Dr.myh11a        HLDIDPNLYRIGQSKIFFRTGVLAQLEEERDLKITVIIIAFQSQARGFLARKAFAKRQQQ 810
Dr.myh11b        HLDLDPNLYRIGLSKIFFRTGVLAQLEEERDLKLTDIIIAFQAQARGFLGRKAFSNKQKQ 796
Dr.myh14         ALELDKNLFRVGQSKVFFRAGVLAHLEEERDLKITDTIIRFQSAARGYLARKAFHKKQQQ 832
                  ::::   ::.* :*:**::*.*. **  ** ::   .   *.  :. : *    .   .

Hs.MYH7          RDSLLVIQWNIRAFMGVKNWPWMKLYFKIKPLLKSAEREKEMASMKEEFTRLKEALEKSE 867
                 MyHC binding site
Hs.MYH7          RDSLLVIQWNIRAFMGVKNWPWMKLYFKIKPLLKSAEREKEMASMKEEFTRLKEALEKSE 867
Hs.MYH6          RDALLVIQWNIRAFMGVKNWPWMKLYFKIKPLLKSAETEKEMATMKEEFGRIKETLEKSE 869
Hs.MYH13         RDSIFCIQYNIRSFMNVKHWPWMNLFFKIKPLLKSAEAEKEMATMKEDFERTKEELARSE 871
Hs.MYH8          REALFCIQYNVRAFMNVKHWPWMKLFFKIKPLLKSAETEKEMATMKEEFQKTKDELAKSE 870
Hs.MYH4          RESIFCIQYNIRAFMNVKHWPWMKLYFKIKPLLKSAETEKEMANMKEEFEKTKEELAKTE 871
Hs.MYH1          RESIFCIQYNVRAFMNVKHWPWMKLYFKIKPLLKSAETEKEMANMKEEFEKTKEELAKTE 871
Hs.MYH2          REAIFCIQYNIRSFMNVKHWPWMKLFFKIKPLLKSAETEKEMATMKEEFQKIKDELAKSE 873
Hs.MYH3          RESIFCIQYNIRSFMNVKHWPWMKLFFKIKPLLKSAETEKEMATMKEEFQKTKDELAKSE 868
Hs.MYH14         QSALRVMQRNCAAYLKLRHWQWWRLFTKVKPLLQVTRQDEVLQARAQELQKVQELQQQSA 889
Hs.MYH15         RDALILIQWNIRAFMAVKNWPWMRLFFKIKPLVKSSEVGEEVAGLKEECAQLQKALEKSE 859
Hs.MYH16         RMGLKVIQQNVHKFLQLRFWGWWKLYNKVKPLLNVARQEEEMKAKEEELRKAMAQTQELV 839
Dr.smyhc1        RDSLLVIQWNVRAFMGVKNWPWMKLYFKIKPLLKTAETEKEMANMKEEFTKLKEAYAKSE 869
Dr.smyhc2        RDSLLVIQWNVRAFMGVKNWPWMKLYFKIKPLLKTAETEKEMANMKEEFTKLKEAYAKSE 870
Dr.smyhc3        RDALLVIQWNVRAFMGVKNWPWMKLYFKIKPLLRSAEAEKEMANMKEEFLKLKEAYAKSE 869
Dr.smyhc4        RDALLVIQWNVRAFMGVKNWPWMKLYFKIKPLLRSAEAEKEMANMKEEFLKLKEAYAKSE 867
Dr.smyhc5        RDALLVIQWNVRAFMGVKNWPWMKLYFKIKPLLRSAEAEKEMANMKEEFLKLKEAYAKSE 869
Dr.myh7          RDALLVIQWNVRAFMGVKNWPWMKLFFKIKPLLKSAEAEKEMANMKDEFAKLKEAYAKSE 869
Dr.myh7l         RDALLVIQWNVRAFMAVKNWPWMKLFFKIKPLLRSAEAEKEMANMKEEFLKLKEAYAKSE 867
Dr.myh6          RDALMVVQWNLRSFLGVKNWPWMKLFFKIKPLLKSAESEKEMANMKDEFNKLKEALEKSD 867
Dr.myha          RESIYTIQYNIRSFMNVKHWPWMKVYYKIKPLLKSAETEKELANMKEDFVKCKEDLVKAE 867
Dr.myhb          RESIYSIQYNIRSFMNVKHWPWMKLYFKIKPLLKSAETEKEMAAMKENFEKMKEDLAKAL 869
Dr.myhz1.1       RDAIYTIQYNVRSFMNVKHWPWMKVYYKIKPLLKSAETEKELATMKEDFVKCKEDLAKAE 869
Dr.myhz1.2       RDAIYTIQYNVRSFMNVKHWPWMKVYYKIKPLLKSAETEKELATMKEDFVKCKEDLAKAE 869
Dr.myhz1.3       RESIYTIQYNIRSFMNVKHWPWMKVYYKIKPLLKSAETEKELATMKEDFVKCKEALAKAE 869
Dr.myhz2         RESIYTIQYNIRSFMNVKHWPWMKVYYKIKPLLKSAETEKELATMKEDFVKCKEALAKAE 869
Dr.myhc4         RESIYTIQYNIRSFMNVKHWPWMKVYYKIKPLLKSAETEKELATMKEDFVKCKEDLVKAE 869
Dr.myh7ba        KEALMIIQWNIRAFNIVKNWPWMKLFFKIKPLLRSAATEKELAALKEEFLKLKEALEKSE 868
Dr.myh7bb        KEALMVIQWNIRAFYAVKNWPWMCLFFKIKPLLRSAATEKELATLKEEFQKLKEALERSE 868
Dr.myh9a         LTAMRVIQRNCAAYLKLRNWQWWRLFTKVKPLLQVTRQEEEMVAKEEELVKMKERQQQAE 872
Dr.myh9b         LTAMRVIQRNCAAYLKLRNWQWWRLFTKVKPLLQVSRQEEEMQAKEEELSKVREKQQVAE 868
Dr.myh10         LSALKVLQRNCAAYLKLRHWQWWRLFTKVKPLLQVTRQEEEMQAKDEELIKVKERQVKVE 885
Dr.myh11a        LTAMKVIQRNCAAYLKLRNWQWWRLFTKVKPLLQVTRQEEEMSLKEEELQKAKESAQKFE 870
Dr.myh11b        LTAMKVLQRNCAVYLTLRNWQWWRLFTKVKPLLQVTRQEEEMNQIEEELQKAKEIAQKSE 856
Dr.myh14         LSALRVMQRNCAAYLKLRNWQWWRLFTKVKPLLQVTRQDEEIQAREAQLQKAKDKLSKLE 892
                   .:  :* *   :  :: * *  :: *:***:. :   : :     :  :
```

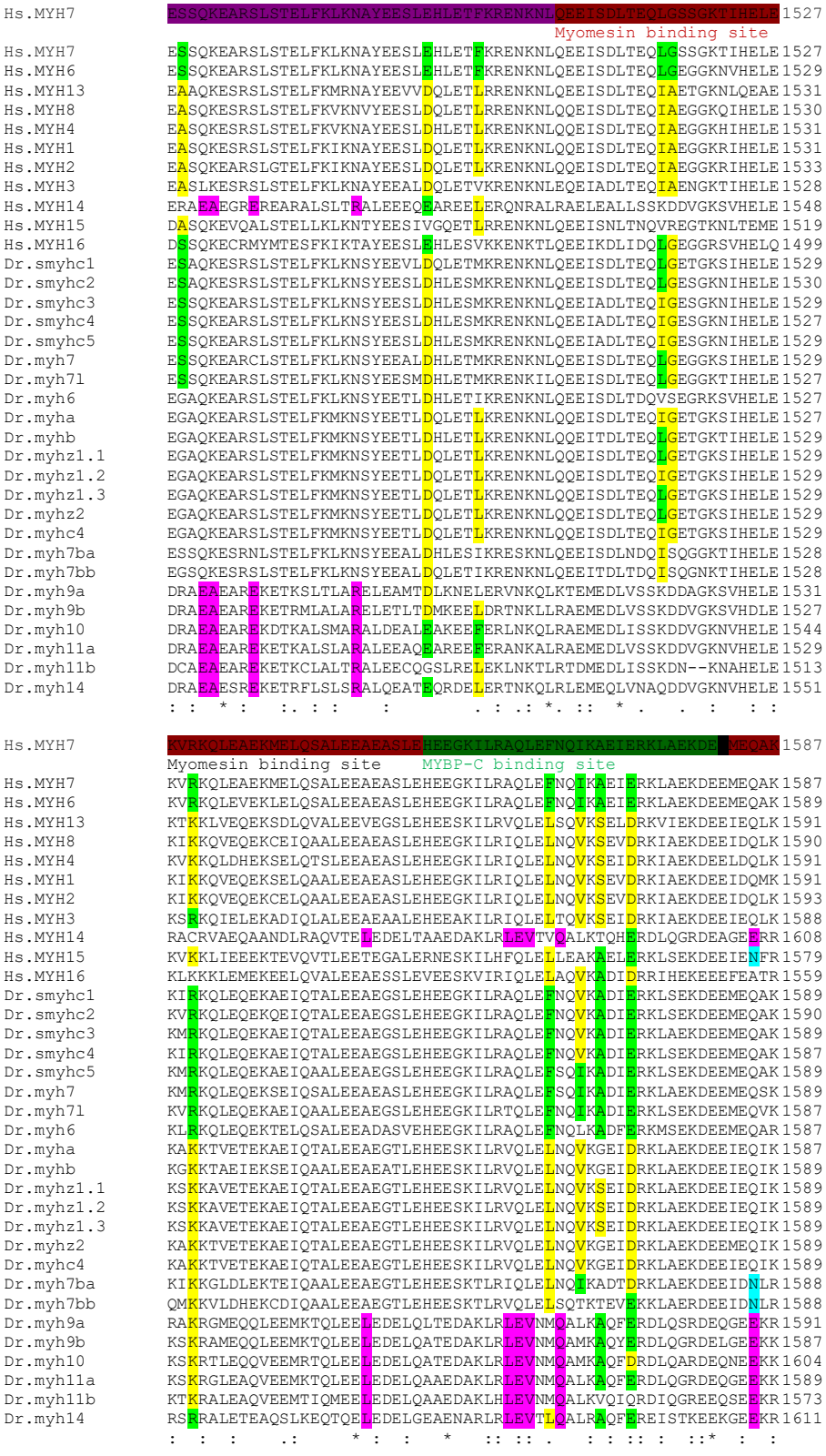
```
Hs.MYH7          ARRKELEEKMVSLLQEKNDLQLQVQAEQDNLADAEERCDQLIKNKIQLEAKVKEMNERLE 927

Hs.MYH7          ARRKELEEKMVSLLQEKNDLQLQVQAEQDNLADAEERCDQLIKNKIQLEAKVKEMNERLE 927
Hs.MYH6          ARRKELEEKMVSLLQEKNDLQLQVQAEQDNLNDAEERCDQLIKNKIQLEAKVKEMNERLE 929
Hs.MYH13         ARRKELEEKMVSLLQEKNDLQLQVQSETENLMDAEERCEGLIKSKILLEAKVKELTERLE 931
Hs.MYH8          AKRKELEEKMVTLLKEKNDLQLQVQSEADSLADAEERCEQLIKNKIQLEAKIKEVTERAE 930
Hs.MYH4          AKRKELEEKMVTLMQEKNDLQLQVQAEADALADAEERCDQLIKTKIQLEAKIKEVTERAE 931
Hs.MYH1          AKRKELEEKMVTLMQEKNDLQLQVQAEADSLADAEERCDQLIKTKIQLEAKIKEVTERAE 931
Hs.MYH2          AKRKELEEKMVTLLKEKNDLQLQVQAEAEGLADAEERCDQLIKTKIQLEAKIKEVTERAE 933
Hs.MYH3          AKRKELEEKLVTLVQEKNDLQLQVQAESENLLDAEERCDQLIKAKFQLEAKIKEVTERAE 928
Hs.MYH14         REVGELQGRVAQLEEERARLAEQLRAEAELCAEAEETRGRLAARKQELELVVSELEARVG 949
Hs.MYH15         FQREELKAKQVSLTQEKNDLILQLQAEQETLANVEEQCEWLIKSKIQLEARVKELSERVE 919
Hs.MYH16         NKVKELEEKTATLSQEKNDLTIQLQAEQENLMDAEERLTWMMKTKMDLESQISDMRERLE 899
Dr.smyhc1        ARKKELEEKMVSLLQEKNDLQLAVQSEQDNLVDAEERCEGLIKSKIQLEAKAKELTERLE 929
Dr.smyhc2        ARKKELEEKMVSLLQEKNDLQLAVQSEQDNLADAEERCEGLIKSKIQFEAKVKELTERLE 930
Dr.smyhc3        ARRKELEEKMVSLLQEKNDLQLAVQAEQDNLCDAEERCEGLIKNKIQLEAKAKELTERLE 929
Dr.smyhc4        ARRKELEEKMVSLLQEKNDLQLAVQSEQDNLCDAEERCEGLIKNKIQLEAKAKELTERLE 927
Dr.smyhc5        ARRKELEEKMVSLLQEKNDLQLAVQAEQDNLCDAEERCEGLIKNKIQLEAKAKELTERLE 929
Dr.myh7          ARRKELEEKMVSLLQEKNDLQLQVQAEQDNLCDAEERCDQLIKNKIQLEAKAKELTERLE 929
Dr.myh7l         ARRKELEEKMVTLLQEKNDLQLQVQAEQDNLCDAEERCEGLIKNKIQMEAKAKELTERLE 927
Dr.myh6          ARRKELEEKMVSLLQEKNDLLLQVQSEQDTLTDAEERCEQLIKSKIQLEAKVKELSERIE 927
Dr.myha          AKKKELEEKMVALLQEKNDLQLAVASEAENLSDAEERCEGLIKSKIQLEAKLKETTERLE 927
Dr.myhb          AKKKELEEKMVTLLQEKNDLQLQVASETENLSDAEERCEGLIKSKIQLEAKLKEATERLE 929
Dr.myhz1.1       AKKKELEEKMVALLQEKNDLQLAVASESENLSDAEERCEGLIKSKIQLEAKLKETTERLE 929
Dr.myhz1.2       AKKKELEEKMVALLQEKNDLQLAVASESENLSDAEERCEGLIKSKIQLEAKLKETTERLE 929
Dr.myhz1.3       AKKKELEEKMVALLQEKNDLQLAVASESENLSDAEERCEGLIKSKIQLEAKLKETTERLE 929
Dr.myhz2         AKKKELEEKMVALLQEKNDLQLAVASESENLSDAEERCEGLIKSKIQLEAKLKETTERLE 929
Dr.myhc4         AKKKELEEKMVALLQEKNDLQLAVASEAENLSDAEERCEGLIKSKIQLEAKLKETTERLE 929
Dr.myh7ba        AKRKELEEKQVSLIQEKNDLSLQLQAEQDNLADAEDRCDLLIKTKIQLEAKVKEMTERLE 928
Dr.myh7bb        VKRKELEEKQVSLVQEKNDLSLQLQAEQDNLADAEDRCNLLIKAKIQMEGKIKELMERLE 928
Dr.myh9a         DQLKESEAKQKQLNAEKLALQEQLQAETELCQEAEEMRSRLTARMQEMEEVLHELESRLE 932
Dr.myh9b         QQLVEMEVKQQQLNAEKMALQEQLQAEMDLCAEADEMRNRLVAKKQELEEILHDLEARVE 928
Dr.myh10         NELVEMERKHQQLLEEKNILAEQLQAETELFAEAEEMRARLVAKKQELEEILHDLESRVE 945
Dr.myh11a        IELKDIALKHTQLMDERNQLQEKLQAETELYAEAEEMRVRLASKKQELEEILHEMEARLE 930
Dr.myh11b        TELKEITQKHDQVVEERNKLQAKLQEEAELYAESEEVRIRLETKKQELEEVLHEMEARLE 916
Dr.myh14         LDFTELDKKNQQLIEEKSVLTDQLQAEAELFAEAEEMRARLANRKQELEDVLGELESRLE 952
                     :   :   :  *:  *   :  * :   : ::    :      :*    :   *

Hs.MYH7          DEEEMNAELTAKKRKLEDECSELKRDIDDLELTLAKVEKEKHATENKVKNLTEEMAGLDE 987

Hs.MYH7          DEEEMNAELTAKKRKLEDECSELKRDIDDLELTLAKVEKEKHATENKVKNLTEEMAGLDE 987
Hs.MYH6          DEEEMNAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAGLDE 989
Hs.MYH13         EEEEMNSELVAKKRNLEDKCSSLKRDIDDLELTLTKVEKEKHATENKVKNLSEEMTALEE 991
Hs.MYH8          EEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAGLDE 990
Hs.MYH4          DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAGLDE 991
Hs.MYH1          DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAGLDE 991
Hs.MYH2          DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAGLDE 993
Hs.MYH3          DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEELSGLDE 988
Hs.MYH14         EEEECSRQMQTEKKRLQQHIQELEAHLEAEEGARQKLQLEKVTTEAKMKKFEEDLLLLED 1009
Hs.MYH15         EEEEINSELTARGRKLEDECFELKKEIDDLETMLVKSEKEKRTTEHKVKNLTEEVEFLNE 979
Hs.MYH16         EEEGMAASLSAAKRKLEGELSDLKRDLEGLETTLAKTEKEKQALDHKVRTLTGDLSLRED 959
Dr.smyhc1        DEEEMNAELVAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAALDE 989
Dr.smyhc2        DEEEMNAELVAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAALDE 990
Dr.smyhc3        DEEEMNAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAALDE 989
Dr.smyhc4        DEEEMNAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAALDE 987
Dr.smyhc5        DEEEMNAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAALDE 989
Dr.myh7          DEEEMNAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAALDD 989
Dr.myh7l         DEEEMNAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAALDD 987
Dr.myh6          DEEEINADLTAKRRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNITEEMASLDE 987
Dr.myha          DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMASQDE 987
Dr.myhb          DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMTSQDE 989
Dr.myhz1.1       DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAAQDE 989
Dr.myhz1.2       DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAAQDE 989
Dr.myhz1.3       DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMAAQDE 989
Dr.myhz2         DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMASQDE 989
Dr.myhc4         DEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATENKVKNLTEEMASQDE 989
Dr.myh7ba        DEEEMNATVLAKKRKLEDECAELKKDIDDLEITLAKVEKEKHATENKVKNLIEEMAALDE 988
Dr.myh7bb        DEEEMNATILAKKRKLEDECIELKKDLDDLEITLAKVEKEKHATENKVKNLVEEMAALDE 988
Dr.myh9a         EEEERVAQFQSEKKKMQQNIGDLEQQLDEEEAARQKLQLEKVTMDAKLKKIEEDLMVIED 992
Dr.myh9b         EEEERANHLQAEKKKMQQNIADLEQQLDEEEAARQKLQLEKVTMEAKLKKTEEEVMVLDD 988
Dr.myh10         EEEERNQSLQNEKKKMQSHIQDLEEQLDEEEAARQKLQLEKVTAEAKIKKMEEDILLLED 1005
Dr.myh11a        EEEDRGAALQMEKKKMHEQIKDLEEHLEEEEDARQKLQLEKVTCDAKIKKLEDDILIMDD 990
Dr.myh11b        EEEERSQAFQQEKGNLQQKLKELENHLAEQDATRQKLQLESGAADGKIKKLEEDVLIMED 976
Dr.myh14         EEEERTVQLTNEKKRIQQHVQDLEEQLEEEEGTRQRLQLEKVTLESKVKSLEAETLTLAE 1012
                 :**     .     .:. .  .*: .:   :    : : *. : : *::.   :     :
```

```
Hs.MYH7          IIAKLTKEKKALQEAHQQALDDLQAEEDKVNTLTKAKVKLEQQVDDLEGSLEQEKKVRMD 1047

Hs.MYH7          IIAKLTKEKKALQEAHQQALDDLQAEEDKVNTLTKAKVKLEQQVDDLEGSLEQEKKVRMD 1047
Hs.MYH6          IIAKLTKEKKALQEAHQQALDDLQVEEDKVNSLSKSKVKLEQQVDDLEGSLEQEKKVRMD 1049
Hs.MYH13         NISKLTKEKKSLQEAHQQTLDDLQVEEDKVNGLIKINAKLEQQTDDLEGSLEQEKKLRAD 1051
Hs.MYH8          TIAKLSKEKKALQETHQQTLDDLQAEEDKVNILTKAKTKLEQQVDDLEGSLEQEKKLRMD 1050
Hs.MYH4          TIAKLTKEKKALQEAHQQTLDDLQMEEDKVNTLTKAKTKLEQQVDDLEGSLEQEKKLCMD 1051
Hs.MYH1          TIAKLTKEKKALQEAHQQTLDDLQAEEDKVNTLTKAKIKLEQQVDDLEGSLEQEKKIRMD 1051
Hs.MYH2          TIAKLTKEKKALQEAHQQTLDDLQAEEDKVNTLTKAKIKLEQQVDDLEGSLEQEKKLRMD 1053
Hs.MYH3          TIAKLTREKKALQEAHQQALDDLQAEEDKVNSLNKTKSKLEQQVEDLESSLEQEKKLRVD 1048
Hs.MYH14         QNSKLSKERKLLEDRLAEFSSQAAEEEEKVKSLNKLRLKYEATIADMEDRLRKEEKGRQE 1069
Hs.MYH15         DISKLNRAAKVVQEAHQQTLDDLHMEEEKLSSLSKANLKLEQQVDELEGALEQERKARMN 1039
Hs.MYH16         SITKLQKEKRALEELHQKTLDDLQAEEDKVNHLTKNNSKLSTQIHELEDNWEQEKKIRAE 1019
Dr.smyhc1        IIAKLTKEKKALQEAHQQTLDDLQSEEDKVNTLTKAKAKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.smyhc2        IIAKLTKEKKALQEAHQQTLDDLQSEEDKVNTLTKAKAKLEQQVDDLEGSLEQEKKIRMD 1050
Dr.smyhc3        IIAKLTKEKKALQEAHQQTLDDLQSEEDKVNTLTKAKAKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.smyhc4        IIAKLTKEKKALQEAHQQTLDDLQSEEDKVNTLTKAKAKLEQQVDDLEGSLEQEKKLRMD 1047
Dr.smyhc5        IIARLTKEKKALQEAHQQTLDDLQSEEDKVNTLTKAKAKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.myh7          IIAKLTKEKKALQEAHQQTLDDLQSEEDKVNTLTKAKAKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.myh7l         IIAKLTKEKKALQEAHQQTLDDLQSEEDKVNTLTKAKVKLEQQVDDLEGSLEQEKKIRMD 1047
Dr.myh6          NIMKLTKEKKALQEAHQQTLDDLQSEEDKVNTLTKAKVKLEQQVDDLEGSLEQEKKVRMD 1047
Dr.myha          SIAKLTKEKKALQEAHQQTLDDLQAEEDKVNTLTKSKTKLEQQVDDLEGSLEQEKKLRMD 1047
Dr.myhb          VIVKLTKEKKALQEAHQQTLDDLQAEEDKVNTLTKAKAKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.myhz1.1       SIGKLTKEKKALQEAHQQTLDDLQAEEDKVNTLTKSKTKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.myhz1.2       SIGKLTKEKKALQEAHQQTLDDLQAEEDKVNTLTKSKTKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.myhz1.3       SIGKLTKEKKALQEAHQQTLDDLQAEEDKVNTLTKSKTKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.myhz2         SIAKLTKEKKALQEAHQQTLDDLQAEEDKVNTLTKSKTKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.myhc4         SIAKLTKEKKALQEAHQQTLDDLQAEEDKVNTLTKSKTKLEQQVDDLEGSLEQEKKLRMD 1049
Dr.myh7ba        TILKLTKEKKALQESHQQTLDDLQTEEDKVNTLTKAKAKLEQQVDDLEGSLEQEKKLRMD 1048
Dr.myh7bb        TISRLSKEKKALQDAHQQALEDLQSEENKVNMLSKAKIKLEQQVDDLEGSLEQEKKVRMD 1048
Dr.myh9a         QNAKLSKEKKQMEERISEFTTNLAEEEEKSKSLQKLKTKHETMITDLEDRLRKEEKMRQE 1052
Dr.myh9b         QNNKLSKEKKLMEERIAEFTTNLAEEEEKSKSLQKLKNKHEAMITDLEDRLRREEKQRQE 1048
Dr.myh10         QNSKFLKEKKLLEDRVGEMTSQLAEEEEKAKNLGKVKNKQEMMMVDLEERLKKEEKTRQE 1065
Dr.myh11a        QNNKLQKERKILEERIADFSSNLAEEEEKSKNLTKLKNKHESMISELEVRLKKEEKTRQE 1050
Dr.myh11b        QKNKLQKEKQQLEERLADFSSNLAEEEEKSKNLTKLKAKHESMISDLEVRMKKEEKSRQD 1036
Dr.myh14         QRDRLSKEKKQLEERLNEVTDQLTEEEEKVKSLNKLKNKQEAVIADIEERLKREEQGRLE 1072
                    :: :  : :::   .   :   **:* . * * . * .    ::*   .:*.:   :

Hs.MYH7          LERAKRKLEGDLKLTQESIMDLENDKQQLDERLKKKDFELNALNARIEDEQALGSQLQKK 1107

Hs.MYH7          LERAKRKLEGDLKLTQESIMDLENDKQQLDERLKKKDFELNALNARIEDEQALGSQLQKK 1107
Hs.MYH6          LERAKRKLEGDLKLTQESIMDLENDKLQLEEKLKKKEFDINQQNSKIEDEQVLALQLQKK 1109
Hs.MYH13         LERAKRKLEGDLKMSQESIMDLENDKQQIEEKLKKKEFELSQLQAKIDDEQVHSLQFQKK 1111
Hs.MYH8          LERAKRKLEGDLKLAQESTMDMENDKQQLDEKLEKKEFEISNLISKIEDEQAVEIQLQKK 1110
Hs.MYH4          LERAKRKLEGDLKLAQESTMDTENDKQQLNEKLKKKEFEMSNLQGKIEDEQALAIQLQKK 1111
Hs.MYH1          LERAKRKLEGDLKLAQESTMDIENDKQQLDEKLKKKEFEMSGLQSKIEDEQALGMQLQKK 1111
Hs.MYH2          LERAKRKLEGDLKLAQESIMDIENEKQQLDEKLKKKEFEISNLQSKIEDEQALGIQLQKK 1113
Hs.MYH3          LERNKRKLEGDLKLAQESILDLENDKQQLDERLKKKDFEYCQLQSKVEDEQTLGLQFQKK 1108
Hs.MYH14         LEKLKRRLDGESSELQEQMVEQQQRAEELRAQLGRKEEELQAALARAEDEGGARAQLLKS 1129
Hs.MYH15         CERELHKLEGNLKLNRESMENLESSQRHLAEELRKKELELSQMNSKVENEKGLVAQLQKT 1099
Hs.MYH16         VEKARRKAESDLKMTIDNLNEMERSKLDLEEVVKKRDLEINSVNSKYEDEQSLNSTLQRK 1079
Dr.smyhc1        LERAKRKLEGDLKLTQESIMDLENDKQQMEEKLKKKDFEISQLNSKIEDEQALGAQLQKK 1109
Dr.smyhc2        LERAKRKLEGDLKLTQENLMDLENDKQQMEERLKKKDFEISQLNSKIEDEQALEAQLQKK 1110
Dr.smyhc3        LERAKRKLEGDLKLTQESVMDLENDKQQLEERLKKKDFEISQLSSKIEDEQAMAAQLQKK 1109
Dr.smyhc4        LERAKRKLEGDLKLTQESVMDLENDKQQLEERLKKKDFEISQLSSRIEDEQAMAAQLQKK 1107
Dr.smyhc5        LERVKRKLEGDLKLTQESVMDLENDKQQLEERLKKKDFEISQLTSRIEDEQAMATQLQKK 1109
Dr.myh7          LERAKRKLEGDLKLTQESLMDLENDKQQLEERLKKKDFEISQLNGKIEDEQTICIQLQKK 1109
Dr.myh7l         LERAKRKLEGDLKLTQESIMDLENDKQQLEEKLKKKDFEISQLNSKIDDEQSIIIQLQKK 1107
Dr.myh6          LERSKRKLEGDVKLTQENVMDLENDKQQLEDKLKKKDFEINQLNQRIEDEQMASVQLQKK 1107
Dr.myha          LERAKRKLEGDLKLAQESIMDLENDKQQSEEKIKKKDFETSQLLSKIEDEQSLGAQLQKK 1107
Dr.myhb          LERAKRKLEGDLKLAQESLMDLENDKQQLEEKLKKKDFETSQFLSRIEDEQSLGIQFQKK 1109
Dr.myhz1.1       LERAKRKLEGDLKLAQESIMDLENDKQQSEEKLKKKDFETSQLLSKIEDEQSLGAQLQKK 1109
Dr.myhz1.2       LERAKRKLEGDLKLAQESIMDLENDKQQSEEKLKKKDFETSQLLSKIEDEQSLGAQLQKK 1109
Dr.myhz1.3       LERAKRKLEGDLKLAQESIMDLENDKQQSEEKLKKKDFETSQLLSKIEDEQSLGAQLQKK 1109
Dr.myhz2         LERAKRKLEGDLKLAQESIMDLENDKQQSEEKIKKKDFETAQLLSKIEDEQSLGAQLQKK 1109
Dr.myhc4         LERAKRKLEGDLKLAQESIMDLENDKQQSEEKIKKKDFETAQLLSKIEDEQSLGAQLQKK 1109
Dr.myh7ba        LERSKRKLEGDLKLSMESVMDLENDKQQLEEKLKKKDFEMNQISSKIEDDQVLIIQLQKK 1108
Dr.myh7bb        LERVKRKLEGDLKLSVESSMDLDNSKQQLEERLKKKDHEMVQIGAKIEEEQALVIQLHKK 1108
Dr.myh9a         LEKNRRKLEGDSTELHDQIAELQAQIAELRAQLAKKEEELQAALARIEEEAALKNAAQKS 1112
Dr.myh9b         LEKNRRKLEGDSTELHDQIAELQAQIAELKAQLAKKEEELQEALARIEEEAAQKNLAQKK 1108
Dr.myh10         LEKAKRKLDAETTDLQDQIAELQAQIDELKIQLAKKEEELQAVLARGDEEVAQKNNALKQ 1125
Dr.myh11a        LDKAKRKLEAESNDLQEQIADLQAQIADLKAQLAKKEEELQAALARLEDETGQKNNALKK 1110
Dr.myh11b        VEKAKRKLETEHSDLQEQMKNLQTLIAELKTQQARSEMEIQELQASVEAESTQKSNALKK 1096
Dr.myh14         QEKWKRRMEGEAVEAQEQLSDMSLLVTELRGSLSQREKEITTLQTRLEEEGARRTEAQRA 1132
                  ::  :: : :     :.  : .    .      : : :        : :        :
```

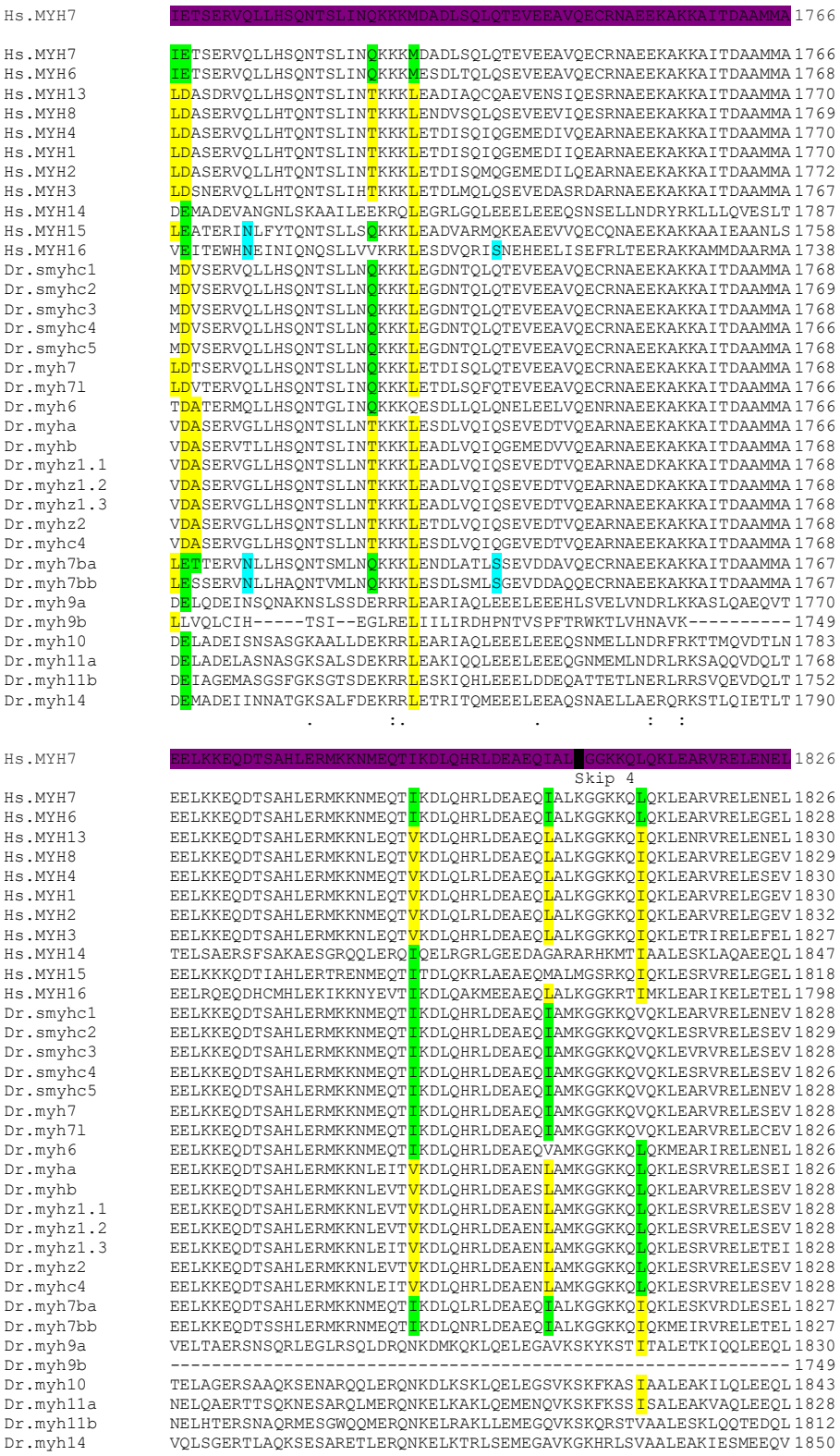
```
Hs.MYH7          LKELQARIEELEEELEAERTARAKVEKLRSDLSRELEEISERLEEAGGATSVQIEMNKKR 1167

Hs.MYH7          LKELQARIEELEELEAERTARAKVEKLRSDLSRELEEISERLEEAGGATSVQIEMNKKR 1167
Hs.MYH6          LKENQARIEELEELEAERTARAKVEKLRSDLSRELEEISERLEEAGGATSVQIEMNKKR 1169
Hs.MYH13         IKELQARIEELEEEIEAEHTLRAKIEKQRSDLARELEEISERLEEASGATSAQIEMNKKR 1171
Hs.MYH8          IKELQARIEELGEEIEAERASRAKAEKQRSDLSRELEEISERLEEAGGATSAQVELNKKR 1170
Hs.MYH4          IKELQARIEELEEEIEAERASRAKAEKQRSDLSRELEEISERLEEAGGATSAQIEMNKKR 1171
Hs.MYH1          IKELQARIEELEEEIEAERASRAKAEKQRSDLSRELEEISERLEEAGGATSAQIEMNKKR 1171
Hs.MYH2          IKELQARIEELEEEIEAERASRAKAEKQRSDLSRELEEISERLEEAGGATSAQIEMNKKR 1173
Hs.MYH3          IKELQARIEELEEEIEAERATRAKTEKQRSDYARELEELSERLEEAGGVTSTQIELNKKR 1168
Hs.MYH14         LREAQAALAEAQEDLESERVARTKAEKQRRDLGEELEALRGELEDTLDSTNAQQELRSKR 1189
Hs.MYH15         VKELQTQIKDLKEKLEAERTTRAKMERERADLTQDLADLNERLEEVGGSSLAQLEITKKQ 1159
Hs.MYH16         LKEHQDRIEELEEELEAERAMRAKVEKQRSDLSRDLEDLSDRLEEAGGATSAQIEQNRKR 1139
Dr.smyhc1        LKELQARIEELEEELEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1169
Dr.smyhc2        LKELQARIEELEEELEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1170
Dr.smyhc3        LKELQARIEELEEELEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1169
Dr.smyhc4        LKELQARIEELEEELEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1167
Dr.smyhc5        LKELQARIEELEEELEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1169
Dr.myh7          LKELQARIEELEEELEAERAARAKVEKQRADLARELEEISERLEEAGGATAAQIEMNKKR 1169
Dr.myh7l         LKELQARVEELEEELEAERAARAKVEKQRADLARELEEISERLEEAGGATAAQIEMNKKR 1167
Dr.myh6          LKENQARIEELEEELDAERAARAKVEKQRSDISRELEDISERLEEAGGATSAQVELNKKR 1167
Dr.myha          IKELQARIEELEEEIEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1167
Dr.myhb          IKELQARIEELEEEIEAERAARAKVEKQRSDLARELEEISERLEEAGGATSAQIEMNKKR 1169
Dr.myhz1.1       IKELQARIEELEEEIEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1169
Dr.myhz1.2       IKELQARIEELEEEIEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1169
Dr.myhz1.3       IKELQARIEELEEEIEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1169
Dr.myhz2         IKELQARIEELEEEIEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1169
Dr.myhc4         IKELQARIEELEEEIEAERAARAKVEKQRADLSRELEEISERLEEAGGATAAQIEMNKKR 1169
Dr.myh7ba        IKELQARIEELEEEMEAERSTRAKMEKHRSDSSKELEELSERLEEAGGATSAQIEMNKKR 1168
Dr.myh7bb        IKELQTRIEELEEELEAERAARSKSEKQRSDVSRELEELSERLEEAGGATTAQIEMNKKR 1168
Dr.myh9a         IREMEAQISELQEDLELEKAARNKAEKQRRDLGEELEALKTELEDTLDSTAAQQELRAKR 1172
Dr.myh9b         IRELESQLSELQEDLELERAARTKAEKHRRDLGEELEALKTELEDTLDSTAAQQELRTKR 1168
Dr.myh10         LRELQAQLAELQEDLESEKAARNKAEKLKRDLSEELEALKTELEDTLDTTAAQQELRSKR 1185
Dr.myh11a        IRELEGHISDLQEDLESERAARNKAEKTKRDLGEELEALKSELEDTLDTTATQQELRAKR 1170
Dr.myh11b        IHEMEGLLSELQDELEAEQGAGRKSEKARKELEEELSALRTELEDSLDTTAVQQELRAKR 1156
Dr.myh14         LREAMSQVSELKEEVENERGMRERAEKQRRDLGEELEALRTELEDTLDTTAAQQELRSRR 1192
                 ::*    : :  :.:: *:    : *: : :  .:*  :  .**:  . : .* *   ::

Hs.MYH7          EAEFQKMRRDLEEATLQHEATAAALRKKHADSVAELGEQIDNLQRVKQKLEKEKSEFKLE 1227
                                    Skip 1
Hs.MYH7          EAEFQKMRRDLEEATLQHEATAAALRKKHADSVAELGEQIDNLQRVKQKLEKEKSEFKLE 1227
Hs.MYH6          EAEFQKMRRDLEEATLQHEATAAALRKKHADSVAELGEQIDNLQRVKQKLEKEKSEFKLE 1229
Hs.MYH13         EAEFQKMRRDLEEATLQHEATAATLRKKQADSVAELGEQIDNLQRVKQKLEKEKSELKME 1231
Hs.MYH8          EAEFQKLRRDLEEATLQHEAMVAALRKKHADSMAELGEQIDNLQRVKQKLEKEKSELKME 1230
Hs.MYH4          EAEFQKMRRDLEESTLQHEATAAALRKKHADSVAELGEQIDSLQRVKQKLEKEKSELKME 1231
Hs.MYH1          EAEFQKMRRDLEEATLQHEATAATLRKKHADSVAELGEQIDNLQRVKQKLEKEKSEMKME 1231
Hs.MYH2          EAEFQKMRRDLEEATLQHEATAATLRKKHADSVAELGEQIDNLQRVKQKLEKEKSEMKME 1233
Hs.MYH3          EAEFLKLRRDLEEATLQHEAMVAALRKKHADSVAELGEQIDNLQRVKQKLEKEKSEFKLE 1228
Hs.MYH14         EQEVTELKKTLEEETRIHEAAVQELRQRHGQALGELAEQLEQARRGKGAWEKTRLALEAE 1249
Hs.MYH15         ETKFQKLHRDMEEATLHFETTSASLKKRHADSLAELEGQVENLQQVKQKLEKDKSDLQLE 1219
Hs.MYH16         EAELLKLRRELEEAALQSEATASTLRKKHVDSMAELTEHVESLQRVKSKLEKDKQVMKAE 1199
Dr.smyhc1        EAEFQKLRRDLEEATLQHEATAATLRKKHADSVADLGEQIDNLQRVKQKLEKEKSELKLE 1229
Dr.smyhc2        EAEFQKLRRDLEEATLQHEATASTLRKKHADSVSDLGEQIDNLQRVKQKLEKEKSELRLE 1230
Dr.smyhc3        EAEFQKLRRDLEEATLQHEATASTLRKKHADSVSDLGEQIDNLQRVKQKLEKEKSELRLE 1229
Dr.smyhc4        EAEFQKLRRDLEEATLQHEATASTLRKKHADSVSDLGEQIDNLQRVKQKLEKEKSELRLE 1227
Dr.smyhc5        EAEFQKLRRDLEEATLQHEATASTLRKKHADSVSDLGEQIDNLQRVKQKLEKEKSELRLE 1229
Dr.myh7          EAEFQKLRRDLEEATLQHEATAATLRKKQADSVAELGEQIDNLQRVKQKLEKEKSELRLE 1229
Dr.myh7l         EAEFQKLRRDLEEATLQHEATAATLRKKQADSVAELGEQIDNLQRVKQKLEKEKSELRLE 1227
Dr.myh6          DAEFQKIRRDLEESTLQHEATTASLRKKHADSVAELGEQIDNLQRVKQKLEKEKVELKLE 1227
Dr.myha          EAEFQKLRRDLEESTLQHEATAAALRKKQADSVAELGEQIDNLQRVKQKLEKEKSEYKME 1227
Dr.myhb          EAEFQKLRRDLEESTLQHEATAAALRKKQADSVAELGEQIDNLQRVKQKLEKEKSELKME 1229
Dr.myhz1.1       EAEFQKLRRDLEESTLQHEATAAALRKKQADSVAELGEQIDNLQRVKQKLEKEKSEYKME 1229
Dr.myhz1.2       EAEFQKLRRDLEESTLQHEATAAALRKKQADSVAELGEQIDNLQRVKQKLEKEKSEYKME 1229
Dr.myhz1.3       EAEFQKLRRDLEESTLQHEATAAALRKKQADSVAELGEQIDNLQRVKQKLEKEKSEYKME 1229
Dr.myhz2         EAEFQKLRRDLEESTLQHEATAAALRKKQADSVAELGEQIDNLQRVKQKLEKEKSEYKME 1229
Dr.myhc4         EAEFQKLRRDLEESTLQHEATAAALRKKQADSVAELGEQIDNLQRVKQKLEKEKSEYKME 1229
Dr.myh7ba        EADFLKLRRDLEEAILHHEAMTAALRKKHADSVVELSEQIDSLQRVKQKLEKERSEAKME 1228
Dr.myh7bb        EADFIKLRRDLEEASLHHETTIAMLRRKHADTVAEMGEQLDNLQRVKQKLEKEKAETRME 1228
Dr.myh9a         ETEVTQLKKTLEDEARAHEQMLSEVRQKHNQAFEELNEQLEQSKRSKASVDKAKQALESE 1232
Dr.myh9b         ETEVAQLKKALEEDAKVHEQVMAEIRQKHSQAFDELNEQLEQVKRNKVSVEKSKQALESE 1228
Dr.myh10         EQEVAELKKAIDDETRNHESQIQEMRQRHGTALEEISEQLEQAKRVKGNLEKNKQTLESD 1245
Dr.myh11a        EQEVTLLKRAIEDESRVHEAQVQEMRQKHTQALEELTEQLEQSKRVKVNLEKAKQALEKE 1230
Dr.myh11b        EQEVAMLKKLIEDEGRSHEAQVHELKQKHAQAVDELSQQLDQSKRAKATLEKAKQALEKE 1216
Dr.myh14         EAELGELQRCLEEETRRHEAQLSELRIKHTAAIDSLQEQLDNAKRSRQSLEKAKAVLEEE 1252
                 : ..  ::: :::     *     :: ::  :. .:  :::. :: :   :* :   .
```

```
Hs.MYH7          LDDVTSNMEQIIKAKANLEKMCRTLEDQMNEHRSKAEETQRSVNDLTSQRAKLQTENGEL 1287

Hs.MYH7          LDDVTSNMEQIIKAKANLEKMCRTLEDQMNEHRSKAEETQRSVNDLTSQRAKLQTENGEL 1287
Hs.MYH6          LDDVTSNMEQIIKAKANLEKVSRTLEDQANEYRVKLEEAQRSLNDFTTQRAKLQTENGEL 1289
Hs.MYH13         IDDMASNIEALSKSKSNIERTCRTVEDQFSEIKAKDEQQTQLIHDLNMQKARLQTQNGEL 1291
Hs.MYH8          TDDLSSNAEAISKAKGNLEKMCRSLEDQVSELKTKEEEQQRLINDLTAQRARLQTEAGEY 1290
Hs.MYH4          INDLASNMETVSKAKANFEKMCRTLEDQLSEIKTKEEEQQRLINELSAQKARLHTESGEF 1291
Hs.MYH1          IDDLASNMETVSKAKGNLEKMCRALEDQLSEIKTKEEEQQRLINDLTAQRARLQTESGEY 1291
Hs.MYH2          IDDLASNVETVSKAKGNLEKMCRTLEDQLSELKSKEEEQQRLINDLTAQRGRLQTESGEF 1293
Hs.MYH3          IDDLSSSMESVSKSKANLEKICRTLEDQLSEARGKNEEIQRSLSELTTQKSRLQTEAGEL 1288
Hs.MYH14         VSELRAELSSLQTARQEGEQRRRRLELQLQEVQGRAGDGERARAEAAEKLQRAQAELENV 1309
Hs.MYH15         VDDLLTRVEQMTRAKANAEKLCTLYEERLHEATAKLDKVTQLANDLAAQKTKLWSESGEF 1279
Hs.MYH16         IDDLNASMETIQKSKMNAEAHVRKLEDSLSEANAKVAELERNQAEINAIRTRLQAENSEL 1259
Dr.smyhc1        LDDVVSNMEQIVKSKSNLEKMCRTLEDQMSEYRTKAEEGQRTINDFTMQKAKLQTENGEL 1289
Dr.smyhc2        LDDVVSNMEQIVKAKANLEKMCRTLEDQMSEYRTKAEEGQRTINDFTMQKAKLQTENGEL 1290
Dr.smyhc3        LDDVVSNMEQIAKAKANLEKMCRTLEDQMSEYRTKYEEGQRSINDFTMKKAKLQTENGEL 1289
Dr.smyhc4        LDDVVSNMEQIVKAKANLEKMCRTLEDQMSEYRTKYEEAQRSINDFTMQKAKLQTENGEL 1287
Dr.smyhc5        LDDVVSNMEQLAKAKANLEKICRTLEDQMSEYRTKYEEGQRSINDFTMQKARLQTENGEL 1289
Dr.myh7          LDDVVSNMEHVVKTKANLEKMTRSLEDQMNEYKTKYEEGQRCINDFTMQKSKLQSENGEL 1289
Dr.myh7l         LDDVASSMEHIVKSKTNMEKVNRTLEDQMNEYRNKCEEYQRSLNDFTTQKAKLQAENDEF 1287
Dr.myh6          LDDLASNMESIVKAKVNLEKMCRSLEDQMNEHRSKAEEAQRALNDVSTQKAKLLTENGEL 1287
Dr.myha          IDDLSSNMEAVAKAKANLEKMCRTLEDQLSEIKSKSDENLRQINDLSAQRARLQTENGEF 1287
Dr.myhb          VDDVSSSMEAVAKSKTNLEKMCRTLEDQLSEFKSKHDEHVRHINDLSAQKARLQTENGEM 1289
Dr.myhz1.1       IDDLSSNMEAVAKAKANLEKMCRTLEDQLSEIKSKNDENLRQINDLSAQRARLQTENGEF 1289
Dr.myhz1.2       IDDLSSNMEAVAKAKANLEKMCRTLEDQLSEIKSKNDENLRQLNDLSAQRARLQTENGEF 1289
Dr.myhz1.3       IDDLSSNMEAVAKAKANLEKMCRTLEDQLSEIKSKNDENIRQINDLSAQRARLQTENGEF 1289
Dr.myhz2         IDDLSSNMEAVAKAKANLEKMCRTVEDQLSEIKSKNDENLRQINDLSAQRARLQTENGEF 1289
Dr.myhc4         IDDLSSNMEAVAKAKANLEKMCRTVEDQLSEIKSKNDENLRQINDLSAQRARLQTENGEF 1289
Dr.myh7ba        ADDLTSNLEQLAKGKATAEKTCRLYEDQMNESKAKVEELQRQLSDTNTQRARAQAESAEL 1288
Dr.myh7bb        CEDLASNVEHLSRAKTTTEKMCRMYEDQLNESKTKIEELQRQLMDVTSQKARAQTESAEV 1288
Dr.myh9a         RNELQIELKSLSQSKNDSENRRKKAESQLQELQVKHTESERQKHELLDKVSKMQAELESL 1292
Dr.myh9b         RNELQIELQTLMQGKGESEHRRKKAEAQLQELQVKHTESERQRIELAERLTKMQAELDNV 1288
Dr.myh10         NKELTNEVKSLQQAKSESEHKRKKLEAQLQEVMARFSEGEKVKGELADRTHKIQTELDNV 1305
Dr.myh11a        TSELHVELRSLTQGKQDVEHKKKKLEGQLADLQSRFNDSERHKAELGDRVSKITVELESV 1290
Dr.myh11b        VGDLNGNLRSLGNAKQDLEQKKKKVETQLADLQTRFNESERKREELGDAVSKLNTEYNNV 1276
Dr.myh14         RLNLSAELKTLQGGKMESERGRKRAEGQLQELNARLSQAERERE-EREERLGKLQSELESL 1312
                   ::      :   :   *       *    :   :  .  :   :       :   :  .

Hs.MYH7          SRQLDEKEALISQLTRGKLTYTQQLEDLKRQLEEEVKAKNALAHALQSARHDCDLLREQY 1347

Hs.MYH7          SRQLDEKEALISQLTRGKLTYTQQLEDLKRQLEEEVKAKNALAHALQSARHDCDLLREQY 1347
Hs.MYH6          ARQLEEKEALISQLTRGKLSYTQQMEDLKRQLEEEGKAKNALAHALQSARHDCDLLREQY 1349
Hs.MYH13         SHRVEEKESLISQLTKSKQALTQQLEELKRQMEEETKAKNAMAHALQSSRHDCDLLREQY 1351
Hs.MYH8          SRQLDEKDALVSQLSRSKQASTQQIEELKHQLEEETKAKNALAHALQSSRHDCDLLREQY 1350
Hs.MYH4          SRQLDEKDAMVSQLSRGKQAFTQQIEELKRQLEEETKAKSTLAHALQSARHDCDLLREQY 1351
Hs.MYH1          SRQLDEKDTLVSQLSRGKQAFTQQIEELKRQLEEEIKAKSALAHALQSSRHDCDLLREQY 1351
Hs.MYH2          SRQLDEKEALVSQLSRGKQAFTQQIEELKRQLEEEIKAKNALAHALQSSRHDCDLLREQY 1353
Hs.MYH3          SRQLEEKESIVSQLSRSKQAFTQQTEELKRQLEEENKAKNALAHALQSSRHDCDLLREQY 1348
Hs.MYH14         SGALNEAESKTIRLSKELSSTEAQLHDAQELLQEETRAKLALGSRVRAMEAEAAGLREQL 1369
Hs.MYH15         LRRLEEKEALINQLSREKSNFTRQIEDLRGQLEKETKSQSALAHALQKAQRDCDLLREQY 1339
Hs.MYH16         SREYEESQSRLNQILRIKTSLTSQVDDYKRQLDEESKSRSTAVVSLANTKHDLDLVKEQL 1319
Dr.smyhc1        SRQLEEKDSLVSQLTRGKQSYTQQIEDLKRQLEEEVKAKNALAHAVQSARHDSDLLREQF 1349
Dr.smyhc2        SRQLEEKDSLVSQLTRGKQSYTQQIEDLKRQLEEEVKAKNALAHAVQSARHDAELLREQY 1350
Dr.smyhc3        SRQLEEKDSLVSQLTRGKQSYTQQIEDLKRQLEEEVKAKNALAHAVQSARHDAELLREQY 1349
Dr.smyhc4        SRQLEEKDSLVSQLTRGKQSYTQQIEDLRRQLEEEVKAKNALAHAVQSARHDAELLREQY 1347
Dr.smyhc5        TRQLEEKDSLVSQLTRSKQSYTQQIEDLKRQLEEEVKAKNALAHAVQSARHDSDLLREQY 1349
Dr.myh7          SRQLEEKDSLVSQLTRSKMSYTQQIEDLKRQLEEETKAKSALAHAVQSARHDTDLLREQY 1349
Dr.myh7l         SRQLEEKESLVSQLTRGKNSFSQQLEDLKRQLDEEIKAKNALAHALQSARHDTDLLREQY 1347
Dr.myh6          GRQLEEKECLISQLTRGKTSYTQQLEDLRRQLEEEVKAKNALAHAVQSARHDCDLLREQF 1347
Dr.myha          GRQLEEKEALVSQLTRGKQAFTQQIEELKRQIEEEVKAKNALAHAVQSARHDCDLLREQF 1347
Dr.myhb          GRQLEEKESLVSQLTRSKQAYTQQIEELKRQIEEEVKAKNSLAHAVQSSRHDCDLLREQY 1349
Dr.myhz1.1       GRQLEEKEALVSQLTRGKQAFTQQIEELKRQIEEEVKAKNALAHAVQSARHDCDLLREQF 1349
Dr.myhz1.2       GRQLEEKEALVSQLTRGKQAFTQQIEELKRQIEEEVKAKNALAHAVQSARHDCDLLREQF 1349
Dr.myhz1.3       GRQLEEKEALVSQLTRGKQAFTQQIEELKRQIEEEVKAKNALAHAVQSARHDCDLLREQF 1349
Dr.myhz2         GRQLEEKEALVSQLTRGKQAFTQQIEELKRQIEEEVKAKNALAHAVQSARHDCDLLREQF 1349
Dr.myhc4         GRQLEEKEALVSQLTRGKQAFTQQIEELKRQIEEEVKAKNALAHAVQSARHDCDLLREQF 1349
Dr.myh7ba        GRKLEEREALVSQLQRSKNTFSQNIEELKKQLEEESKSKNALTHALQSSRHDCDLLREQY 1348
Dr.myh7bb        SRRLEEKEVLVMQLQRTKIAFSQTVEELKKQLEEESKAKNSLAHAVQSSRHDCDLLREQF 1348
Dr.myh9a         QGTVTKVESKSIKAAKDCSAVESQLKDAQALLEEETRQKLAISTRLRQLEDEQNNLKEML 1352
Dr.myh9b         NTLLSDAEGKSIKASKDCSTVESQLQDVQEVLQEETRQKLALNTRLRQLEDEQHSLREQL 1348
Dr.myh10         SCLLEDAEKKGIKLTKDVSSLESQLQDTQELLQEETRQKLNLSSRIRQLEEEKNNLLEQQ 1365
Dr.myh11a        TNLLNEAEGKNIKLSKDVASLSSQVQDTQELLAEETRQKLQLSTKLRQIEDDRNALQEQL 1350
Dr.myh11b        NSILNEAESKNIKLSKDVVSLNSQLQDAQELLAEETRQKLNFSTRLRQMEDERNGLLEQI 1336
Dr.myh14         SSSLSSSDSKSHRLHKEVSSLESQLHDVQELLQEETRQKLALGSRVRALEEEKAGLMERL 1372
                       . :    :  :         .: :  : :* : :      :   . :   : *
```


Skip 2


Myomesin binding site
Myomesin binding site
MYBP-C binding site

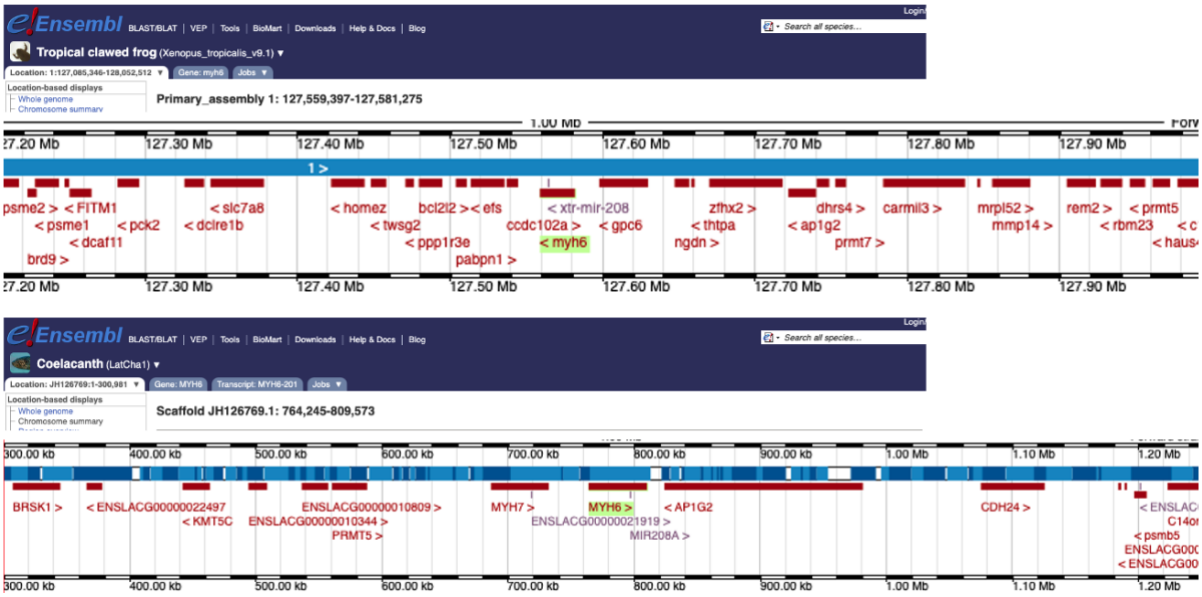
```
Hs.MYH7          RNHLRVVDSLQTSLDAETRSRNEALRVKKKMEGDLNEMEIQLSHANRMAAEAQKQVKSLQ 1647
                 Myomesin binding site
Hs.MYH7          RNHLRVVDSLQTSLDAETRSRNEALRVKKKMEGDLNEMEIQLSHANRMAAEAQKQVKSLQ 1647
Hs.MYH6          RNHQRVVDSLQTSLDAETRSRNEVLRVKKKMEGDLNEMEIQLSHANRMAAEAQKQVKSLQ 1649
Hs.MYH13         RNSQRAAEALQSVLDAEIRSRNDALRLKKKMEGDLNEMEIQLGHSNRQMAETQKHLRTVQ 1651
Hs.MYH8          RNHTRVVETMQSTLDAEIRSRNDALRVKKKMEGDLNEMEIQLNHANRLAAESLRNYRNTQ 1650
Hs.MYH4          RNHLRVVESMQSTLDAEIRSRNDALRIKKKMEGDLNEMEIQLNHANRQAAEALRNLRNTQ 1651
Hs.MYH1          RNHIRIVESMQSTLDAEIRSRNDAIRLKKKMEGDLNEMEIQLNHANRMAAEALRNYRNTQ 1651
Hs.MYH2          RNHIRIVESMQSTLDAEIRSRNDAIRLKKKMEGDLNEMEIQLNHANRMAAEALRNYRNTQ 1653
Hs.MYH3          RNYQRTVETMQSALDAEVRSRNEAIRLKKKMEGDLNEIEIQLSHANRQAAETLKHLRSVQ 1648
Hs.MYH14         RQLAKQLRDAEVERDEERKQRTLAVAARKKLEGELEELKAQMASAGQGKEEAVKQLRKMQ 1668
Hs.MYH15         RKQQCTIDSLQSSLDSEAKSRIEVTRLKKKMEEDLNEMELQLSCANRQVSEATKSLGQLQ 1639
Hs.MYH16         KNHQRAIESLQASLEAEAKGRAEALRLKKKMETDLNEMEIQLDHANKNNSELVKTLKRLQ 1619
Dr.smyhc1        RNQQRVVDTLQSSLESETRSRNEALRLKKKMEGDLNEMEIQLSQANRQASEAQKQLKGLH 1649
Dr.smyhc2        RNQQRVVDTLQSSLESETRSRNEALRLKKKMEGDLNEMEIQLSQANRQASEAQKQLKGLH 1650
Dr.smyhc3        RNQQRMIDTLQSSLESETRSRNEALRLKKKMEGDLNEMEIQLSQANRQASEAQKQLKGLH 1649
Dr.smyhc4        RNQQRMIDTLQSSLESETRSRNEALRLKKKMEGDLNEMEIQLSQANRQASEAQKQLKGLH 1647
Dr.smyhc5        RNQQRMIDTLQSSLESETRSRNEALRLKKKMEGDLNEMEIQLSQANRQASEAQKQLKSLQ 1649
Dr.myh7          RNLQRTIDTLQSSLESETRSRNEALRIKKKMEGDLNEMEIQLSQANRQAAEAQKQLKSVH 1649
Dr.myh7l         RNQQRTIDTLQSALESETRSRNEALRIKKKMEGDLNEMEIQLSQANRQAAEAQKQLKSVQ 1647
Dr.myh6          RNYQRMIESLQASLEAETRSRNEALRVKKKMEGDLNEMEIQLSQANRQAADAQKQLKMVQ 1647
Dr.myha          RNSQRVTEAMQSTLDSEVRSRNDALRIKKKMEGDLNEMEIQLSHANRQAAEAQKQLRNVQ 1647
Dr.myhb          RNSQRIIDSMQSTLDAEVRSRNDALRIKKKMEGDLNEMEIQLSHANRQAAEAQKQLRNVQ 1649
Dr.myhz1.1       RNSQRITDSMQSTLDSEVRSRNDALRIKKKMEGDLNEMEIQLSHANRQAAEAQKQLRNVQ 1649
Dr.myhz1.2       RNSQRITDSMQSTLDSEVRSRNDALRIKKKMEGDLNEMEIQLSHANRQAAEAQKQLRNVQ 1649
Dr.myhz1.3       RNSQRVTESMQSTLDSEVRSRNDALRIKKKMEGDLNEMEIQLSHANRQAAEAQKQLRNVQ 1649
Dr.myhz2         RNSQRVTEAMQSTLDSEVRSRNDALRIKKKMEGDLNEMEIQLSHANRQAAEAQKQLRNVQ 1649
Dr.myhc4         RNSQRVTEAMQSTLDSEVRSRNDALRIKKKMEGDLNEMEIQLSHANRQAAEAQKQLRNVQ 1649
Dr.myh7ba        RNHQRALESMQATLDAEAKSRSEAIRVKKKMENDLNEMEVQLNHANWLATESQKMVRNLQ 1648
Dr.myh7bb        RNHQRTLEGMQTTLDAETRARNEAIRVKKKMENDMNEMEIHLNHANRQAVESQKMVRNLQ 1648
Dr.myh9a         KQLVKQVREMEMELEDERKQRAQAVSVRKKLELDLSELAAQIDLANKARDEALKQLKKLQ 1651
Dr.myh9b         RQLLKQVREMEMELEDERKQRTLAMAARKKMELDLKELEAAIDQANKNRDEALKQLKKVQ 1647
Dr.myh10         RALVKQVREMEAELEDERKQRALAVAAKKKLEMDLKDVEAQIEAANKARDEAIKQLRKLQ 1664
Dr.myh11a        RQLVKQVRELETELEDERKQRTALAASKKKLEGDLKDLEGQIETSNKGRDEAIKQLRKLQ 1649
Dr.myh11b        KQLLKQVRELEAELEDEQKMRTSLAAAKKKLEGDLQDLEDQVDVNSRARDEAVKQLRKIQ 1633
Dr.myh14         RALNKQVRELETMLEEEKTQRAQALTVKKQLETELQEAEAQVEAANRGREEAFRQMKRLQ 1671
                 :         :   : *   *      :*::* ::.:    :   .    :  :     :

Hs.MYH7          SLLKDTQIQLDDAVRANDDLKENIAIVERRNNLLQAELEELRAVVEQTERSR-KLAEQEL 1706
                 Myomesin binding site
Hs.MYH7          SLLKDTQIQLDDAVRANDDLKENIAIVERRNNLLQAELEELRAVVEQTERSR-KLAEQEL 1706
Hs.MYH6          SLLKDTQIQLDDAVRANDDLKENIAIVERRNNLLQAELEELRAVVEQTERSR-KLAEQEL 1708
Hs.MYH13         GQLKDSQLHLDDALRSNEDLKEQLAIVERRNGLLLEELEEMKVALEQTERTR-RLSEQEL 1710
Hs.MYH8          GILKETQLHLDDALRGQEDLKEQLAIVERRANLLQAEIEELWATLEQTERSR-KIAEQEL 1709
Hs.MYH4          GILKDTQLHLDDAIRGQDDLKEQLAMVERRANLMQAEVEELRASLERTERGR-KMAEQEL 1710
Hs.MYH1          AILKDTQLHLDDALRSQEDLKEQLAMVERRANLLQAEIEELRATLEQTERSR-KIAEQEL 1710
Hs.MYH2          GILKDTQIHLDDALRSQEDLKEQLAMVERRANLLQAEIEELRATLEQTERSR-KIAEQEL 1712
Hs.MYH3          GQLKDTQLHLDDALRGQEDLKEQLAIVERRANLLQAEVEELRATLEQTERAR-KLAEQEL 1707
Hs.MYH14         AQMKELWREVEETRTSREEIFSQNRESEKRLKGLEAEVLRLQEELAASDRAR-RQAQQDR 1727
Hs.MYH15         IQIKDLQMQLDDSTQLNSDLKEQVAVAERRNSLLQSELEDLRSLQEQTERGR-RLSEEEL 1698
Hs.MYH16         QQIKDLQVQMDEDARQHEELRKQYNLQERRLSLLQTELEEVRSALEGSERSR-KLLEQEV 1678
Dr.smyhc1        GHLKDAQLQLDDALRGNDDLKENIAIVERRNNLLQAELDELRSLVEQTERGR-KLAEQEL 1708
Dr.smyhc2        GHLKDAQLQLDDALRGNDDLKENIAIVERRNNLLQAELDELRSLVEQTERGR-KLAEQEL 1709
Dr.smyhc3        GHLKDAQLQLDDALRGNDDLKENIAIVERRNNLLQAELDELRSLVEQTERGR-KLAEQEL 1708
Dr.smyhc4        GHLKDAQLQLDDALRGNDDLKENIAIVERRNNLLQAELDELRSLVEQTERGR-KLAEQEL 1706
Dr.smyhc5        GHLKDAQMQLDDALRANDDLKENIAIVERRNNLLQAELDELRSLVEQTERGR-KLAEQEL 1708
Dr.myh7          AHMKDAQLQLDDSLRTNEDLKENTAIVERRNNLLQAELEELRAALEQTERGR-KLAEQEL 1708
Dr.myh7l         AHLKDSQLQLDDSLRSNDDLKENTAIVERRNALLQAELEELRAVLEQTERGR-KLAEQEL 1706
Dr.myh6          SCLKETQLQMDDTLHSNDDLKENTTLLERRNNLMQTELEELRGILEQTERVR-KLAEQEL 1706
Dr.myha          SQLKDAQLHLDDAVRAQEDMKEQVAMVERRNTLMQSEIEELRAALEQTERGR-KVAEQEL 1706
Dr.myhb          GQLKDAQLHLDEALRAQEDMKEQVAMVERRNNLMQAEIEELRVALEQTERGR-KVAEQEL 1708
Dr.myhz1.1       AQLKDAQLHLDDAVRGQEDMKEQVAMVERRNTLMQSEIEELRAALEQTERGR-KVAEQEL 1708
Dr.myhz1.2       AQLKDAQLHLDDAVRGQEDMKEQVAMVERRNTLMQSEIEELRAALEQTERGR-KVAEQEL 1708
Dr.myhz1.3       AQLKDAQLHLDDAVRGQEDMKEQVAMVERRNTLMQSEIEELRAALEQTERGR-KVAEQEL 1708
Dr.myhz2         AQLKDAQLHLDDAVRGQEDMKEQVAMVERRNTLMQSEIEELRAALEQTERGR-KVAEQEL 1708
Dr.myhc4         AQLKDAQLHLDDAVRGQEDMKEQVAMVERRNTLMQSEIEELRAALEQTERGR-KVAEQEL 1708
Dr.myh7ba        TQIKDLQMELDETVHQNEELKEQSAVTERRNNLLTAEVEELRCQLEQNDRAR-KLAEFEL 1707
Dr.myh7bb        LQIKDLQVELDESMHHCEELKEQVAVTERRNTLLSAELEELRGVAEQTDRMR-KVAEHEL 1707
Dr.myh9a         AQMKEQMREFEDLRLSRDESLNQAKENERKIKSMEAEIMQLHEDLAAADRAK-RQIQQER 1710
Dr.myh9b         AQMKDLLRELEDTRLSREEILAQSKENEKKVKSMEAEMIQMQEGKLAAVDSLHSNLHMH- 1706
Dr.myh10         AQMKDYQRELEEARTSRDEIFTQSKENEKKLKSLEAEILQLQEDLASSERAR-RHAEQER 1723
Dr.myh11a        AQMKDFQRELDDAHAAREEVLSSAKENERKAKTLEAELLQLQEDLAAAERAK-KQVEAER 1708
Dr.myh11b        TQMKDYQRELEDARASHKEVLSDARESERKARAMEAEILHLHEELASAEKAR-KHAERER 1692
Dr.myh14         TQMKELIRELDETKLARDEIVAQSKDSEKRLQTLEAELLQLTEDLSVSERQK-RQAQQER 1730
                   :*:   ..::     .:   .    *::   :  *:  :               . .
```


Hs.MYH7 IETSERVQLLHSQNTSLINQKKKMDADLSQLQTEVEEAVQECRNAEEKAKKAITDAAMMA 1766
Hs.MYH7 IETSERVQLLHSQNTSLINQKKKMDADLSQLQTEVEEAVQECRNAEEKAKKAITDAAMMA 1766
Hs.MYH6 IETSERVQLLHSQNTSLINQKKKMESDLTQLQSEVEEAVQECRNAEEKAKKAITDAAMMA 1768
Hs.MYH13 LDASDRVQLLHSQNTSLINTKKKLEADIAQCQAEVENSIQESRNAEEKAKKAITDAAMMA 1770
Hs.MYH8 LDASERVQLLHTQNTSLINTKKKLENDVSQLQSEVEEVIQESRNAEEKAKKAITDAAMMA 1769
Hs.MYH4 LDASERVQLLHTQNTSLINTKKKLETDISQIQGEMEDIVQEARNAEEKAKKAITDAAMMA 1770
Hs.MYH1 LDASERVQLLHTQNTSLINTKKKLETDISQIQGEMEDIIQEARNAEEKAKKAITDAAMMA 1770
Hs.MYH2 LDASERVQLLHTQNTSLINTKKKLETDISQMQGEMEDILQEARNAEEKAKKAITDAAMMA 1772
Hs.MYH3 LDSNERVQLLHTQNTSLIHTKKKLETDLMQLQSEVEDASRDARNAEEKAKKAITDAAMMA 1767
Hs.MYH14 DEMADEVANGNLSKAAILEEKRQLEGRLGQLEEELEEEQSNSELLNDRYRKLLLQVESLT 1787
Hs.MYH15 LEATERINLFYTQNTSLLSQKKKLEADVARMQKEAEEVVQECQNAEEKAKKAAIEAANLS 1758
Hs.MYH16 VEITEWHNEINIQNQSLLVVKRKLESDVQRISNEHEELISEFRLTEERAKKAMMDAARMA 1738
Dr.smyhc1 MDVSERVQLLHSQNTSLLNQKKKLEGDNTQLQTEVEEAVQECRNAEEKAKKAITDAAMMA 1768
Dr.smyhc2 MDVSERVQLLHSQNTSLLNQKKKLEGDNTQLQTEVEEAVQECRNAEEKAKKAITDAAMMA 1769
Dr.smyhc3 MDVSERVQLLHSQNTSLLNQKKKLEGDNTQLQTEVEEAVQECRNAEEKAKKAITDAAMMA 1768
Dr.smyhc4 MDVSERVQLLHSQNTSLLNQKKKLEGDNTQLQTEVEEAVQECRNAEEKAKKAITDAAMMA 1766
Dr.smyhc5 MDVSERVQLLHSQNTSLLNQKKKLEGDNTQLQTEVEEAVQECRNAEEKAKKAITDAAMMA 1768
Dr.myh7 LDTSERVQLLHSQNTSLLNQKKKLETDISQLQTEVEEAVQECRNAEEKAKKAITDAAMMA 1768
Dr.myh7l LDVTERVQLLHSQNTSLINQKKKLETDLSQFQTEVEEAVQECRNAEEKAKKAITDAAMMA 1766
Dr.myh6 TDATERMQLLHSQNTGLINQKKKQESDLLQLQNELEELVQENRNAEEKAKKAITDAAMMA 1766
Dr.myha VDASERVGLLHSQNTSLLNTKKKLESDLVQISEVEDTVQEARNAEEKAKKAITDAAMMA 1766
Dr.myhb VDASERVTLLHSQNTSLINTKKKLEADLVQIQGEMEDVVQEARNAEEKAKKAITDAAMMA 1768
Dr.myhz1.1 VDASERVGLLHSQNTSLLNTKKKLEADLVQIQSEVEDTVQEARNAEDKAKKAITDAAMMA 1768
Dr.myhz1.2 VDASERVGLLHSQNTSLLNTKKKLEADLVQIQSEVEDTVQEARNAEDKAKKAITDAAMMA 1768
Dr.myhz1.3 VDASERVGLLHSQNTSLLNTKKKLEADLVQIQSEVEDTVQEARNAEEKAKKAITDAAMMA 1768
Dr.myhz2 VDASERVGLLHSQNTSLLNTKKKLETDLVQIQSEVEDTVQEARNAEEKAKKAITDAAMMA 1768
Dr.myhc4 VDASERVGLLHSQNTSLLNTKKKLESDLVQIQGEVEDTVQEARNAEEKAKKAITDAAMMA 1768
Dr.myh7ba LETTERVNLLHSQNTSMLNQKKKLENDLATLSSEVDDAVQECRNAEEKAKKAITDAAMMA 1767
Dr.myh7bb LESSERVNLLHAQNTVMLNQKKKLESDLSMLSGEVDDAQQECRNAEEKAKKAITDAAMMA 1767
Dr.myh9a DELQDEINSQNAKNSLSSDERRRLEARIAQLEEELEEEHLSVELVNDRLKKASLQAEQVT 1770
Dr.myh9b LLVQLCIH-----TSI--EGLRELIILIRDHPNTVSPFTRWKTLVHNAVK---------- 1749
Dr.myh10 DELADEISNSASGKAALLDEKRRLEARIAQLEEELEEEQSNMELLNDRFRKTTMQVDTLN 1783
Dr.myh11a DELADELASNASGKSALSDEKRRLEAKIQQLEEELEEEQGNMEMLNDRLRKSAQQVDQLT 1768
Dr.myh11b DEIAGEMASGSFGKSGTSDEKRRLESKIQHLEEELDDEQATTETLNERLRRSVQEVDQLT 1752
Dr.myh14 DEMADEIINNATGKSALFDEKRRLETRITQMEEELEEAQSNAELLAERQRKSTLQIETLT 1790
EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIALKGGKKQLQKLEARVRELENEL 1826
Skip 4
Hs.MYH7 EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIALKGGKKQLQKLEARVRELENEL 1826
Hs.MYH6 EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIALKGGKKQLQKLEARVRELEGEL 1828
Hs.MYH13 EELKKEQDTSAHLERMKKNLEQTVKDLQHRLDEAEQLALKGGKKQIQKLENRVRELENEL 1830
Hs.MYH8 EELKKEQDTSAHLERMKKNLEQTVKDLQHRLDEAEQLALKGGKKQIQKLEARVRELEGEV 1829
Hs.MYH4 EELKKEQDTSAHLERMKKNMEQTVKDLQLRLDEAEQLALKGGKKQIQKLEARVRELESEV 1830
Hs.MYH1 EELKKEQDTSAHLERMKKNLEQTVKDLQHRLDEAEQLALKGGKKQIQKLEARVRELEGEV 1830
Hs.MYH2 EELKKEQDTSAHLERMKKNMEQTVKDLQLRLDEAEQLALKGGKKQIQKLEARVRELEGEV 1832
Hs.MYH3 EELKKEQDTSAHLERMKKNLEQTVKDLQHRLDEAEQLALKGGKKQIQKLETRIRELEFEL 1827
Hs.MYH14 TELSAERSFSAKAESGRQQLERQIQELRGRLGEEDAGARARHKMTIAALESKLAQAEEQL 1847
Hs.MYH15 EELKKKQDTIAHLERTRENMEQTITDLQKRLAEAEQMALMGSRKQIQKLESRVRELEGEL 1818
Hs.MYH16 EELRQEQDHCMHLEKIKKNYEVTIKDLQAKMEEAEQLALKGGKRTIMKLEARIKELETEL 1798
Dr.smyhc1 EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIAMKGGKKQVQKLEARVRELENEV 1828
Dr.smyhc2 EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIAMKGGKKQVQKLESRVRELESEV 1829
Dr.smyhc3 EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIAMKGGKKQVQKLEVRVRELESEV 1828
Dr.smyhc4 EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIAMKGGKKQVQKLESRVRELESEV 1826
Dr.smyhc5 EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIAMKGGKKQVQKLEARVRELENEV 1828
Dr.myh7 EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIAMKGGKKQVQKLEARVRELESEV 1828
Dr.myh7l EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQIAMKGGKKQVQKLEARVRELECEV 1826
Dr.myh6 EELKKEQDTSAHLERMKKNMEQTIKDLQHRLDEAEQVAMKGGKKQLQKMEARIRELENEL 1826
Dr.myha EELKKEQDTSAHLERMKKNLEITVKDLQHRLDEAENLAMKGGKKQLQKLESRVRELESEI 1826
Dr.myhb EELKKEQDTSAHLERMKKNLEVTVKDLQHRLDEAESLAMKGGKKQLQKLEARVRELESEV 1828
Dr.myhz1.1 EELKKEQDTSAHLERMKKNLEVTVKDLQHRLDEAENLAMKGGKKQLQKLESRVRELESEV 1828
Dr.myhz1.2 EELKKEQDTSAHLERMKKNLEVTVKDLQHRLDEAENLAMKGGKKQLQKLESRVRELESEV 1828
Dr.myhz1.3 EELKKEQDTSAHLERMKKNLEITVKDLQHRLDEAENLAMKGGKKQLQKLESRVRELETEI 1828
Dr.myhz2 EELKKEQDTSAHLERMKKNLEVTVKDLQHRLDEAENLAMKGGKKQLQKLESRVRELESEV 1828
Dr.myhc4 EELKKEQDTSAHLERMKKNLEITVKDLQHRLDEAENLAMKGGKKQLQKLESRVRELESEV 1828
Dr.myh7ba EELKKEQDTSAHLERMKKNMEQTIKDLQLRLDEAEQIALKGGKKQIQKLESKVRDLESEL 1827
Dr.myh7bb EELKKEQDTSSHLERMKRNMEQTIKDLQNRLDEAEQIALKGGKKQIQKMEIRVRELETEL 1827
Dr.myh9a VELTAERSNSQRLEGLRSQLDRQNKDMKQKLQELEGAVKSKYKSTITALETKIQQLEEQL 1830
Dr.myh9b ---------- 1749
Dr.myh10 TELAGERSAAQKSENARQQLERQNKDLKSKLQELEGSVKSKFKASIAALEAKILQLEEQL 1843
Dr.myh11a NELQAERTTSQKNESARQLMERQNKELKAKLQEMENQVKSKFKSSISALEAKVAQLEEQL 1828
Dr.myh11b NELHTERSNAQRMESGWQQMERQNKELRAKLLEMEGQVKSKQRSTVAALESKLQQTEDQL 1812
Dr.myh14 VQLSGERTLAQKSESARETLERQNKELKTRLSEMEGAVKGKHRLSVAALEAKIESMEEQV 1850

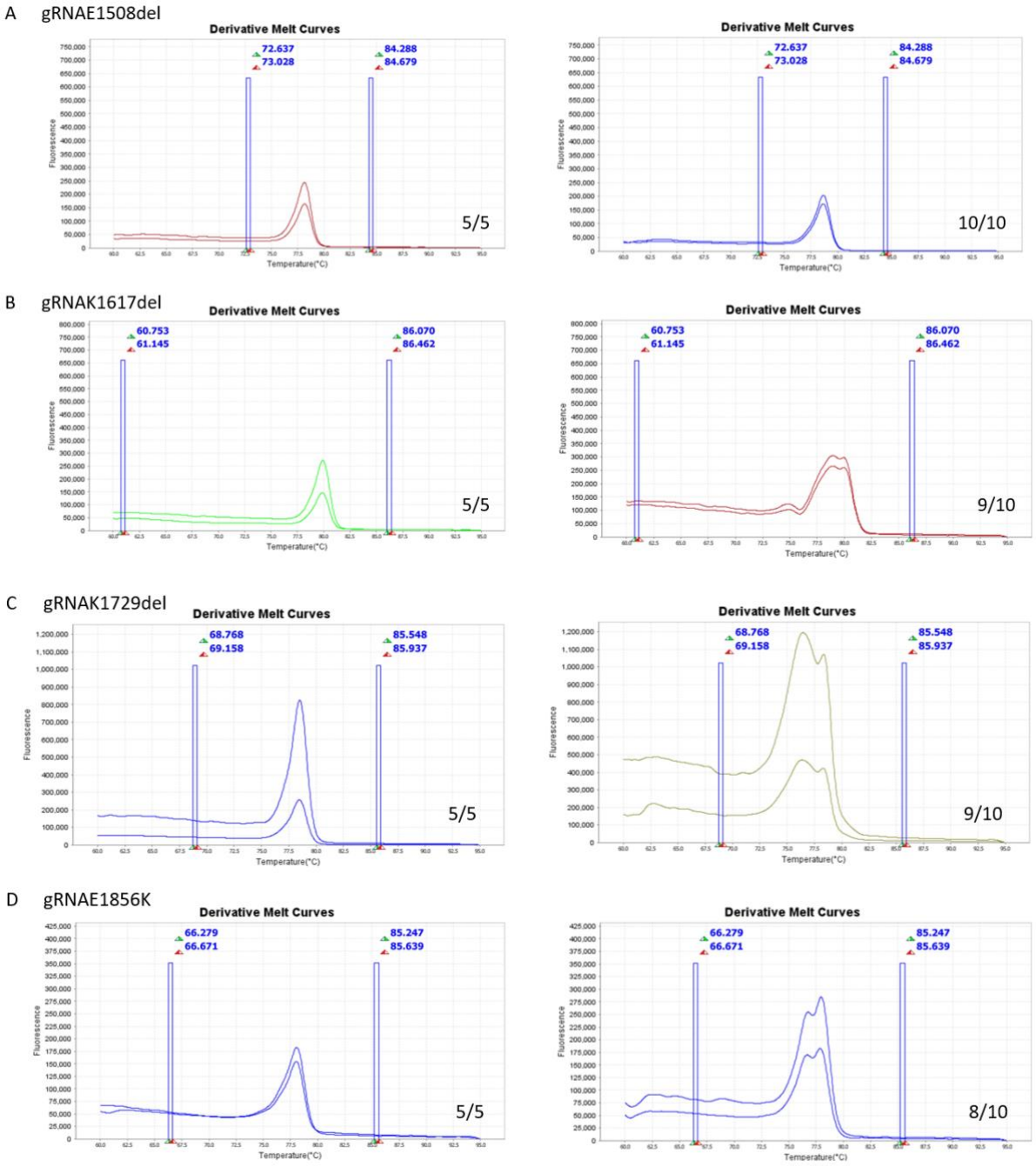
```
Hs.MYH7          EAEQKRNAESVKGMRKSERRIKELTYQTEEDRKNLLRLQDLVDKLQLKVKAYKRQAEEAE 1886

Hs.MYH7          EAEQKRNAESVKGMRKSERRIKELTYQTEEDRKNLLRLQDLVDKLQLKVKAYKRQAEEAE 1886
Hs.MYH6          EAEQKRNAESVKGMRKSERRIKELTYQTEEDKKNLLRLQDLVDKLQLKVKAYKRQAEEAE 1888
Hs.MYH13         DVEQKRGAEALKGAHKYERKVKEMTYQAEEDHKNILRLQDLVDKLQAKVKSYKRQAEEAE 1890
Hs.MYH8          ENEQKRNAEAVKGLRKHERRVKELTYQTEEDRKNVLRLQDLVDKLQAKVKSYKRQAEEAE 1889
Hs.MYH4          ESEQKHNVEAVKGLRKHERRVKELTYQTEEDRKNILRLQDLVDKLQTKVKAYKRQAEEAE 1890
Hs.MYH1          ESEQKRNVEAVKGLRKHERKVKELTYQTEEDRKNILRLQDLVDKLQAKVKSYKRQAEEAE 1890
Hs.MYH2          ESEQKRNAEAVKGLRKHERRVKELTYQTEEDRKNILRLQDLVDKLQAKVKSYKRQAEEAE 1892
Hs.MYH3          EGEQKKNTESVKGLRKYERRVKELTYQSEEDRKNVLRLQDLVDKLQVKVKSYKRQAEEAD 1887
Hs.MYH14         EQETRERILSGKLVRRAEKRLKEVVLQVEEERRVADQLRDQLEKGNLRVKQLKRQLEEAE 1907
Hs.MYH15         EGEIRRSAEAQRGARRLERCIKELTYQAEEDKKNLSRMQTQMDKLQLKVQNYKQQVEVAE 1878
Hs.MYH16         DGEQKQHVETVKTLCKNERRLKELVFQTEEDHKTNQRMQALVEKLQNKLKVYKRQIEEAE 1858
Dr.smyhc1        ELEQRKASESVKGVRKYERRIKELTYQTEEDRKNLARLQDLVDKLQLKVKSYKRAAEEAE 1888
Dr.smyhc2        EMEQRKASDSVKGVRKYERRIKELTYQTEEDRKNLARLQDLVDKLQLKVKSYKRTAEEAE 1889
Dr.smyhc3        EMEQRKASESVKGVRKYERRIKELTYQTEEDRKNLARLQDLVDKLQLKVKSYKRAAEEAE 1888
Dr.smyhc4        EMEQRKASESVKGVRKYERRIKELTYQTEEDRKNLARLQDLVDKLQLKVKSYKRAAEEAE 1886
Dr.smyhc5        ELEQKKASESVKGIRKYERRIKELTYQTEEDRKNLARLQDLVDKLQLKVKSYKRAAEEAE 1888
Dr.myh7          ESEQKKSSEAVKGIRKYERRIKELTYQTEEDRKNLARLQDLVDKLQLKVKAYKRAAEEAE 1888
Dr.myh7l         EAEQKRSSESVKGIRKYERRIKELTYQTEEDRKNIARLQDLVDKLQLKVKAYKRAAEESE 1886
Dr.myh6          DAEQKRGSESVKGVRKFERRIKELTYQTDEDRKNLARLQDLVDKLQLKVKSYKRSAEEAE 1886
Dr.myha          DAEQRRGADAVKGVRKYERRVKELTYQTEEDKKNINRLQDLVDKLQLKVKAYKRQSEEAE 1886
Dr.myhb          EAEQRRGADAVKGVRKYERRVKELTYQTEEDKKNIIRLQDLVDKLQLKVKAYKRQSEDAE 1888
Dr.myhz1.1       EAEQRRGADAVKGVRKYERRVKELTYQTEEDKKNVNRLQDLVDKLQLKVKAYKRQSEEAE 1888
Dr.myhz1.2       EAEQRRGADAVKGVRKYERRVKELTYQTEEDKKNVNRLQDLVDKLQLKVKAYKRQSEEAE 1888
Dr.myhz1.3       EAEQRRGADAVKGVRKYERRVKELTYQTEEDKKNVNRLQDLVDKLQLKVKAYKRQSEEAE 1888
Dr.myhz2         EAEQRRGADAVKGVRKYERRVKELTYQTEEDKKNVNRLQDLVDKLQLKVKAYKRQSEEAE 1888
Dr.myhc4         EAEQRRGADAVKGVRKYERRVKELTYQTEEDKKNINRLQDLVDKLQLKVKAYKRQSEEAE 1888
Dr.myh7ba        ESEQKKSNELQKGIRKYERRIKELTYQTEEDRKNIARLQELIDKLQAKVKTYKRQAEDAE 1887
Dr.myh7bb        DCEQKKSAEFQKGIRKYERRIKELTYQTEEDRKTLLRMQDLIDKLQAKVKSFKRQAEDAE 1887
Dr.myh9a         DSEMKERQQSTKQVRRVEKKLKEVLLQVEDERRNADQSKTETEKANIRLKQMKRQLEETE 1890
Dr.myh9b         ------------------------------------------------------------ 1749
Dr.myh10         EQEAKERAAANKIVRRTEKKLKEVFMQVEDERRHADQYKEQMEKANSRMKQLKRQLEEAE 1903
Dr.myh11a        EQESRDKQNTAKAVRQKDKKLKEMMTQVEDERKQAEQYKDQADKATARVKQLKRQLEESE 1888
Dr.myh11b        DHESRERQAIAKAMRQKDKKLKELMNQVEDERKQMEQYKDQAEKANLRARQLKMQMEESD 1872
Dr.myh14         EQERQERAMANKLVRKTEKKLKEVLMQVEDERRHADQYREQLDKSMGRLRQLKRQLEEVE 1910

Hs.MYH7          EQANTNLSKFRKVQHELDEAEERADIAESQVNKLRAKSRDIGTKGLNEE---------- 1935

Hs.MYH7          EQANTNLSKFRKVQHELDEAEERADIAESQVNKLRAKSRDIGTKGLNEE---------- 1935
Hs.MYH6          EQANTNLSKFRKVQHELDEAEERADIAESQVNKLRAKSRDIGAKQKMHDEE--------- 1939
Hs.MYH13         EQANTQLSRCRRVQHELEEAAERADIAESQVNKLRAKSRDVGSQKMEE------------ 1938
Hs.MYH8          EQSNANLSKFRKLQHELEEAEERADIAESQVNKLRVKSREVHTKISAE------------ 1937
Hs.MYH4          EQSNVNLAKFRKLQHELEEAKERADIAESQVNKLRVKSREVHTKVISEE----------- 1939
Hs.MYH1          EQSNVNLSKFRRIQHELEEAEERADIAESQVNKLRVKSREVHTKIISEE----------- 1939
Hs.MYH2          EQSNTNLAKFRKLQHELEEAEERADIAESQVNKLRVKSREVHTKVISEE----------- 1941
Hs.MYH3          EQANAHLTKFRKAQHELEEAEERADIAESQVNKLRAKTRDFTSSRMVVHESEE------- 1940
Hs.MYH14         EEASRAQAGRRRLQRELEDVTESAESMNREVTTLRNRLRRGPLTFTTRTVRQV----FRL 1963
Hs.MYH15         TQANQYLSKYKKQQHELNEVKERAEVAESQVNKLKIKAREFGKKVQEE----------- 1926
Hs.MYH16         DQANQTLARYRKTVHELDDAEDRAGMAETALNKLRTRHRVAGKGITSVEIIQVSKTGTSK 1918
Dr.smyhc1        EQANSNLGKFRKLQHELDEAEERADIAESQVNKMRAKSRDSGPKKGHDEE---------- 1938
Dr.smyhc2        EQANSNLGKFRKLQHELDEAEERADIAESQVNKLRAKSRDSGSKKGADEE---------- 1939
Dr.smyhc3        EQANSNLGKFRKLQHELDEAEERADIAESQVNKLRAKSRDTGSKKGHDEE---------- 1938
Dr.smyhc4        EQANSNLGKFRKLQHELDEAEERADIAESQVNKLRAKSRDTGSKKGHDEE---------- 1936
Dr.smyhc5        EQANSNLTKFRKLQHELDEAEERADIAESQVNKLRAKSRDTGSKKGQEEE---------- 1938
Dr.myh7          EQANTNLSKFRKIQHELDEAEERADIAESQVNKLRAKSRDVSSKKGHDQE---------- 1938
Dr.myh7l         EQANVHLGKFRKLQHELDEAEERADIAESQVNKLRAKSRDVGPKKAFDEE---------- 1936
Dr.myh6          ELANANTAKLRKLQHELEEAEERADIAESQVNKLRAKTRDGAPGKKTQDE---------- 1936
Dr.myha          EQANSHLSKLRKVQHELEEAEERADISESQVNKLRAKSRDAGKAKEE------------- 1933
Dr.myhb          EQANTHLTRFRKVQHELEEAQERADIAESQVNKLRAKSREFGKGKEAEE----------- 1937
Dr.myhz1.1       EQANSHLSKLRKVQHELEEAEERADIAESQVNKLRAKSRDSGKGKDAAE----------- 1937
Dr.myhz1.2       EQANSHLSKLRKVQHELEEAEERADIAESQVNKLRAKSRDSGKGKEAAE----------- 1937
Dr.myhz1.3       EQANSHLSKLRKVQHELEEAEERADIAESQVNKLRAKSRDAGKGKESAE----------- 1937
Dr.myhz2         EQANSHLSKLRKVQHELEEAEERADIAESQVNKLRAKSRDAGKAKEE------------- 1935
Dr.myhc4         EQANSHLSKLRKVQHELEEAEERADISESQVNKLRAKSRDAGKAKEE------------- 1935
Dr.myh7ba        EQANSNLTKYKKIQHELDDAEERADIAESQVNKLRTRSRDSGSKTTPKLAE---------- 1938
Dr.myh7bb        EQVNCNMTRFRKIQHDLDEAEERADMAESQVNKLRVRTRETHVVKVVE------------ 1935
Dr.myh9a         EEAARANASCRKLRRELEDATESASAMNREVSTLKNKLRRGDFTGNVRRAI------GRT 1944
Dr.myh9b         ------------------------------------------------------------ 1749
Dr.myh10         EEATRANASRRKLQRELDDATEASEGLSREVNTLKNRLRRGGPVSFSSSRS------GRR 1957
Dr.myh11a        EESQRITAARRKLQRELDEATETNDAMSREVNSLKSKLRRGNETSSFSSTPRRTGGGGRR 1948
Dr.myh11b        EESQRATAARRKLQKELNEAIEANDALSREVSSLRSKLRRGSDSSFSTPLRRSAGGSFRQ 1932
Dr.myh14         EENSRSNAQRRKLQRELEEMSDSMQSMNRELNTLRSQLRRAPLPLSMRAGRRA------- 1963
```

```
Hs.MYH7         ------------------------------------------------------------1935

Hs.MYH7         ------------------------------------------------------------1935
Hs.MYH6         ------------------------------------------------------------1939
Hs.MYH13        ------------------------------------------------------------1938
Hs.MYH8         ------------------------------------------------------------1937
Hs.MYH4         ------------------------------------------------------------1939
Hs.MYH1         ------------------------------------------------------------1939
Hs.MYH2         ------------------------------------------------------------1941
Hs.MYH3         ------------------------------------------------------------1940
Hs.MYH14        ------EEGVASDEEAEEAQPGSGPSPE------P----EGS-PPAHPQ-----------1995
Hs.MYH15        ------------------------------------------------------------1926
Hs.MYH16        T---L------SEE-------------------------------------------------1923
Dr.smyhc1       ------------------------------------------------------------1938
Dr.smyhc2       ------------------------------------------------------------1939
Dr.smyhc3       ------------------------------------------------------------1938
Dr.smyhc4       ------------------------------------------------------------1936
Dr.smyhc5       ------------------------------------------------------------1938
Dr.myh7         ------------------------------------------------------------1938
Dr.myh7l        ------------------------------------------------------------1936
Dr.myh6         ------------------------------------------------------------1936
Dr.myha         ------------------------------------------------------------1933
Dr.myhb         ------------------------------------------------------------1937
Dr.myhz1.1      ------------------------------------------------------------1937
Dr.myhz1.2      ------------------------------------------------------------1937
Dr.myhz1.3      ------------------------------------------------------------1937
Dr.myhz2        ------------------------------------------------------------1935
Dr.myhc4        ------------------------------------------------------------1935
Dr.myh7ba       ------------------------------------------------------------1938
Dr.myh7bb       ------------------------------------------------------------1935
Dr.myh9a        G---V-----ESDDENDLK------S-------------EGS-EPTPE------------1964
Dr.myh9b        ------------------------------------------------------------1749
Dr.myh10        Q---LQMEGDFSDDDADSK------ASD------L----NEN-QPPPAE-----------1986
Dr.myh11a       G---IIDSSDAAEDDADMQ------S-D------Y----NGT-KSNE-------------1974
Dr.myh11b       ARRTMMETSEIPEDGGPSA------T-------------SVC-QPGELQMESISNTQDNN1972
Dr.myh14        ----LVD--DLSQENSDSEDPGASPTPSSGPPGTPTPSDNALGPPPPYSL---TDAE---2011
```

Appendix 4.3 – Tropical Clawed Frog and Coelocanth MYH6/7 neighbour genes



Appendix 5.1 HRM derivative melt curves results showing non-injected siblings vs CRISPR/Cas9 injected embryos



A) *smyhc1* gRNAE1508del, uninjected controls show single peak with melting temperature at 78 °C, no shift in melting temperature peak of injected embryos **B)** *smyhc1* gRNAK1617del uninjected controls show single peak with melting temperature at 80 °C, injected embryos show derivative melt curve show shifted double peak **C)** *smyhc1* gRNAK1617del uninjected controls show single peak with melting temperature at 78 °C, injected embryos show derivative melt curve show shifted double peak **D)** *smyhc1* gRNAE1856K uninjected controls show single peak with melting temperature at 78 °C, injected embryos show derivative melt curve show shifted double peak.

Appendix 5.1 – *smyhc1* F3 generation genotyping, length, and weight measurements

5.1.1. - Smyhc1^{kg179}

Fish no.	Gender	Length (mm)	Mass (mg)	Genotype (By HRM)	Genotype by Seq
1	m	26	370	Wild Type	Wild Type
2	f	25	390	Wild Type	Wild Type
5	m	24	340	Wild Type	
8	m	26	310	Wild Type	Wild Type
12	m	26	320	?	Wild Type
13	m	25	400	Wild Type	
17	m	23	230	Wild Type	
19	m	26	340	Wild Type	
20	m	25	310	Wild Type	
25	m	27	370	Wild Type	
29	m	27	340	Wild Type	
30	m	25	340	Wild Type	
40	m	26	310	Wild Type	
3	m	25	360	Heterozygote	Heterozygote
4	f	26	420	Heterozygote	Heterozygote
7	m	28	440	Heterozygote	
9	m	25	290	Heterozygote	
10	m	27	330	Heterozygote	
11	m	25	280	Heterozygote	
15	m	26	280	Heterozygote	
16	m	26	280	Heterozygote	
21	m	25	240	Heterozygote	
22	m	27	310	Heterozygote	
24	m	29	440	Heterozygote	
27	f	26	360	Heterozygote	
28	f	25	340	Heterozygote	
33	m	25	250	Heterozygote	
34	m	28	340	Heterozygote	
35	m	26	340	Heterozygote	
37	m	25	360	Heterozygote	
38	m	29	410	Heterozygote	
39	m	28	360	Heterozygote	
6	m	23	370	Mutant	Mutant
14	m	26	340	Mutant	Mutant
18	m	29	390	Mutant	
23	m	26	360	Mutant	
26	m	28	340	Mutant	
31	m	26	310	Mutant	
32	m	29	390	Mutant	
36	m	23	240	Mutant	

5.1.2. - *Smyhc1*^{kg180}

Fish no.	Gender	Length (mm)	Mass (mg)	Genotype (By HRM)	Genotype by Seq
13	m	22	160	Wild Type	
15	m	23	250	Wild Type	
16	f	22	220	Wild Type	
18	m	24	210	Wild Type	
24	mm	23	220	Wild Type	
25	m	22	200	Wild Type	
28	m	23	180	Wild Type	
29	m	22	180	Wild Type	
30	f	21	180	Wild Type	
32	m	24	240	Wild Type	
33	m	23	220	Wild Type	
36	m	20	200	Wild Type	
43	f			Wild Type	Wild Type
48	f			Wild Type	Wild Type
51	m			Wild Type	
54	m			Wild Type	
56	m			Wild Type	
3	m	24	220	Heterozygote	Heterozygote
6	m	24	210	Heterozygote	Heterozygote
8	m	21	210	Heterozygote	
9	f	23	180	Heterozygote	
10	m	24	240	Heterozygote	
11	m	24	240	Heterozygote	
14	m	22	200	Heterozygote	
17	f	21	180	Heterozygote	
31	m	21	160	Heterozygote	
34	f	23	230	Heterozygote	
37	f	22	210	Heterozygote	
39	f	25	250	Heterozygote	
41	f			Heterozygote	Heterozygote
42	f			Heterozygote	Heterozygote
44	f			Heterozygote	
45	f			Heterozygote	
47	ff			Heterozygote	
49	m			Heterozygote	
52	m			Heterozygote	
53	m			Heterozygote	
55	m			Heterozygote	
1	m	23	220	Mutant	Mutant
2	f	25	270	Mutant	Mutant
4	m	24	220	Mutant	Mutant
5	f	22	180	Mutant	Mutant
7	f	21	170	Mutant	Mutant
12	m	24	290	Mutant	Mutant
19	m	22	200	Mutant	Mutant
20	f	25	280	Mutant	Mutant
21	f	22	220	Mutant	Mutant
22	m	25	210	Mutant	Mutant
23	m	22	250	Mutant	Mutant
26	m	19	100	Mutant	Mutant
27	m	23	200	Mutant	Mutant
38	f	22	210	Mutant	Mutant
40	m	22	190	Mutant	Mutant
6	f			Mutant	Mutant
10	m			Mutant	Mutant

5.1.3. - Summary of smyhc1^{kg179} and smyhc1^{kg180} mendalian ratio and Chi squared test

smyhc1^{kg179}

Summary of fish numbers	expected	observed		chi squared
wt	10	13	33%	0.53661755
het	20	19	48%	>0.05 therefore observed values are the predicted ratio
mut	10	8	20%	
total no.	40	40		

smyhc1^{kg180}

Summary of fish numbers	expected	observed		chi squared
wt	13.75	17	31%	0.196464031
het	27.5	21	38%	>0.05 therefore observed values are the predicted ratio
mut	13.75	17	31%	
total no.	55	55		

5.1.4. - Genotype of dead fish from F3 generation of smyhc1 heterozygous in-crosses from 5 dpf to 4 mpf

Fish death from F3 *smyhc1$^{kg179/+}$* and *smyhc1$^{kg180/+}$* in-cross

smyhc1^{kg179}

Fish no.	Genotype by Seq
1	Heterozygote
2	Heterozygote
3	Wild Type
4	Heterozygote
5	Mutant
6	Heterozygote
7	Heterozygote
8	Wild Type
9	Heterozygote
10	Mutant
11	Heterozygote
12	Mutant
13	Heterozygote
14	Heterozygote
15	Heterozygote
16	Mutant
17	Heterozygote
18	Wild Type
19	Heterozygote
20	Wild Type
21	Heterozygote
22	Heterozygote
23	Heterozygote
24	Mutant

smyhc1^{kg180}

Fish no.	Genotype by Seq
1	Heterozygote
2	Heterozygote
3	Wild Type
4	Heterozygote
5	Mutant
6	Heterozygote

Appendix 5.2 – Zebrafish swimming velocity 2-30 dpf

2 dpf	Lay	Date/Info	Genotype	Swimming Velocity				count
				BTS- (mm/s)	SD	BTS+ (mm/s)	SD	
	1	29/09/2019	Wild Type	327.08	86.11	8.99	4.25	4
			Heterozygote	259.06	99.80	8.20	3.73	8
			Mutant	102.24	32.06	0.49	0.28	5
	2	11/03/2020	Wild Type	278.47	134.59	7.17	3.17	6
			Heterozygote	238.93	174.56	6.66	4.84	10
			Mutant	131.72	21.61	0.66	0.21	5
	3	11/03/2020	Wild Type	311.32	184.05	6.17	3.99	6
			Heterozygote	258.96	119.74	7.12	6.08	10
			Mutant	130.50	27.23	0.41	0.33	6
	4	11/03/2020	Wild Type	369.12	297.27	11.38	7.51	5
			Heterozygote	233.24	141.47	12.69	8.02	11
			Mutant	180.93	99.43	0.75	0.79	5
	Average (1-4)		**Wild Type**	**321.50**	**175.51**	**8.43**	**4.73**	
			Heterozygote	**247.55**	**133.89**	**8.66**	**5.67**	
			Mutant	**136.35**	**45.08**	**0.58**	**0.40**	

5 dpf	Lay	Date/Info	Genotype	Swimming Velocity				count
				BTS- (mm/s)	SD	BTS+ (mm/s)	SD	
	1	29/09/2019	Wild Type	573.06	125.65	58.58	24.47	6
			Heterozygote	544.61	219.44	55.25	19.24	9
			Mutant	453.88	153.73	3.29	2.39	5
	2	11/03/2020	Wild Type	579.10	291.34	93.96	19.32	6
			Heterozygote	620.23	333.73	55.63	28.71	10
			Mutant	377.15	442.20	1.36	0.68	5
	3	11/03/2020	Wild Type	548.59	390.91	44.38	60.51	5
			Heterozygote	690.93	366.66	30.28	16.17	11
			Mutant	293.42	160.62	0.77	0.55	5
	Average (1-3)		**Wild Type**	**566.92**	**269.30**	**65.64**	**34.77**	
			Heterozygote	**618.59**	**306.61**	**47.05**	**21.37**	
			Mutant	**374.82**	**252.18**	**1.81**	**1.21**	

17 dpf	Lay	Date/Info	Genotype	Swimming Velocity				count
				BTS- (mm/s)	SD	BTS+ (mm/s)	SD	
	1	11/12/2020	Wild Type	15.97	7.68	3.93	2.06	7
			Heterozygote	15.25	4.93	3.39	0.81	7
			Mutant	23.82	20.01	0.35	0.31	5
	2	11/12/2020	Wild Type	12.80	1.71	4.09	0.66	4
			Heterozygote	15.46	7.33	4.10	1.47	7
			Mutant	10.10	4.84	0.62	0.32	4
	3	11/12/2020	Wild Type	11.74	2.69	4.02	0.34	5
			Heterozygote	12.42	5.39	3.93	1.01	9
			Mutant	14.66	5.05	0.18	0.06	4
	Average (1-3)		**Wild Type**	**13.50**	**4.02**	**4.01**	**1.02**	
			Heterozygote	**14.38**	**5.88**	**3.80**	**1.09**	
			Mutant	**16.19**	**9.97**	**0.38**	**0.23**	

20 dpf	Lay	Date/Info	Genotype	Swimming Velocity				count
				BTS- (mm/s)	SD	BTS+ (mm/s)	SD	
	1	11/12/2020	Wild Type	25.46	6.91	4.67	1.61	6
			Heterozygote	22.85	8.90	4.21	2.65	8
			Mutant	24.11	6.27	0.46	0.21	4
	2	11/12/2020	Wild Type	34.24	19.19	4.91	1.77	6
			Heterozygote	40.95	17.35	4.93	1.78	13
			Mutant	33.26	7.73	0.42	0.00	3
	3	11/12/2020	Wild Type	36.35	29.32	3.64	0.63	3
			Heterozygote	24.72	20.73	5.90	2.40	8
			Mutant	37.38	28.80	0.34	0.16	4
	Average (1-3)		**Wild Type**	**32.02**	**18.48**	**4.41**	**1.34**	
			Heterozygote	**29.51**	**15.66**	**5.01**	**2.28**	
			Mutant	**31.59**	**14.27**	**0.41**	**0.12**	

30 dpf	Lay	Date/Info	Genotype	Swimming Velocity				count
				BTS- (mm/s)	SD	BTS+ (mm/s)	SD	
	1	11/12/2020	Wild Type	50.65	26.85	4.12	2.77	5
			Heterozygote	59.89	40.38	6.49	2.63	7
			Mutant	75.45	7.71	6.17	2.51	3
	2	11/12/2020	Wild Type	60.56	27.75	6.15	1.09	4
			Heterozygote	41.56	21.25	5.67	1.09	7
			Mutant	32.53	17.45	5.60	0.78	4
	3	11/12/2020	Wild Type	33.72	15.71	5.86	0.60	4
			Heterozygote	38.38	8.45	5.28	1.74	7
			Mutant	51.37	14.16	6.16	2.13	4
	Average (1-3)		**Wild Type**	**48.31**	**23.44**	**5.38**	**1.49**	
			Heterozygote	**46.61**	**23.36**	**5.81**	**1.82**	
			Mutant	**53.12**	**13.11**	**5.97**	**1.81**	

Appendix 5.3 – BLAST search of gRNA to zebrafish genome

***smyhc1* gRNAKO1**										
5′-CATGTCAAAAATACGAGTTTGGG-3′										
Genomic Location	**Overlapping Gene(s**	**Orientation**	**Query name**	**Query start**	**Query end**	**Query ori**	**Length**	**Score**	**E-val**	**%ID**
24:40667704-40667724	smyhc1	Reverse	Query_1	1	21	Forward	21	42	0.024	100
***smyhc1* gRNAKO2**										
5′-ACCACAGAGGAATCGTACACTGG-3′										
Genomic Location	**Overlapping Gene(s**	**Orientation**	**Query name**	**Query start**	**Query end**	**Query ori**	**Length**	**Score**	**E-val**	**%ID**
24:40665790-40665812	smyhc1	Forward	Query_1	1	23	Forward	23	46	0.002	100
***smyhc1* gRNA - (Li et al, 2020)**										
5′-GGCTGACAGCATGTACTGGTAGG-3′										
Genomic Location	**Overlapping Gene(s**	**Orientation**	**Query name**	**Query start**	**Query end**	**Query ori**	**Length**	**Score**	**E-val**	**%ID**
24:40665704-40665724	smyhc1	Forward	Query_1	3	23	Forward	21	42	0.035	100
24:40698260-40698280	smyhc2	Forward	Query_1	3	23	Forward	21	42	0.035	100
24:40723771-40723791	smyhc3	Forward	Query_1	3	23	Forward	21	42	0.035	100
24:40744293-40744313	CU633479.1	Forward	Query_1	3	23	Forward	21	42	0.035	100
24:40772959-40772979	CU633479.2	Forward	Query_1	3	23	Forward	21	42	0.035	100
***smyhc1* exon 16 (Whittle et al, 2020)**										
Genomic Location	**Overlapping Gene(s**	**Orientation**	**Query name**	**Query start**	**Query end**	**Query ori**	**Length**	**Score**	**E-val**	**%ID**
24:40662386-40662462	smyhc1	Reverse	Query_1	1	77	Forward	77	152	3.00E-34	100
2:24264943-24264998	myh7	Reverse	Query_1	12	64	Forward	56	61.8	7.00E-07	89.29
24:40688153-40688209	smyhc2	Reverse	Query_1	21	77	Forward	57	57.8	1.00E-05	87.72

Appendix 5.4 – early STOP codon in exon 16 of smyhc1 from Whittle et al, 2020