

Novel functional domains in the muscle disease associated protein KY

Oscar Harrad

Master of Science by Research

University of York

Biology

December 2020

Abstract

Kyphoscoliosis Peptidase (KY) is a skeletal muscle Z-disk protein that has been implicated in several inherited myopathies. The mechanism underpinning KY's function has remained elusive, but previous research suggests that KY could have protease activity over generic substrates. However, protease activity over endogenous substrates have not been proven. The highly conserved transglutaminase/protease (TGN/PROT) domain is the only distinguishable feature within KY. This domain includes conservation of the three catalytic residues which have been shown to be required for enzymatic activity in cysteine proteases. Here, biochemical assays have shown that this domain is not enzymatically active. In addition, electroporations of the *ky/ky* mouse muscle with a KY version lacking the TGN/PROT domain indicates that this domain is not required for Z-disk localisation but is required for the rescue of atrophic muscle fibres.

Here, *in-silico* analysis has shown that human KY (hKY) protein has gained a nuclear localisation signal (NLS) located upstream of the putative start site of the mouse and other vertebrate homologues. This NLS appears to be functional, as in contrast to the mouse KY protein, hKY shows nuclear localisation in C2C12 myoblasts and COS-7 fibroblasts. Within electroporated adult mouse muscle hKY localises both to the Z-disk and the nucleus. The Z-disc/nuclear localization of hKY is highly reminiscent of IGFN1, a known protein partner of mouse KY. Co-transfections of hKY and IGFN1 in C2C12 myoblasts leads to strong accumulation of both protein in nuclear aggregates, while IGFN1 is not able to change localization of mouse KY to the nucleus. Thus, the mouse and human KY orthologues show different expression patterns. Further analysis shows that this NLS is conserved within primates and therefore may have evolved to facilitate the distinct biomechanics of primate muscles compared to other placental mammals.

Table of Contents

Abstract	2
Table of Figures	6
Acknowledgements	7
Declaration	7
Chapter 1: Introduction	8
1.1 Skeletal muscle structure	9
1.1.1 Skeletal muscle structure.....	9
1.1.2 Muscle Fibers.	11
1.2 Skeletal muscle hypertrophy	12
1.3 Skeletal muscle atrophy	15
1.4 Kyphoscoliosis peptidase.....	18
1.4.1 The <i>ky/ky</i> Mouse.....	18
1.4.2 The <i>Ky</i> Gene.....	19
1.4.3 KY's Interacting protein partners.....	20
1.4.4 Human cases of <i>KY</i> mutations.....	21
1.4.5 Myofibrillar myopathies.....	23
1.5 Immunoglobulin-like and fibronectin type III domain containing-1 (IGFN1).24	
1.5.1 Overview of IGFN1	24
1.6 Hypotheses.....	26
1.6.1 Does KY contain any enzymatic activity with the TGN/PROT domain?..26	
1.6.2 Does KY's TGN/PROT domain control Z-disk localization?	26
1.6.3 Can the Human KY localize to the Z-disk?	26
1.7 Objectives and Aims.	27
Chapter 2: Methods	28
2.1 Table of Reagents	29
2.1.1 Solutions and Reagents	29
2.1.2 Antibodies	29
2.2 Heteroduplex analysis.....	30
2.2.1 PCR Amplification.....	30
2.2.2 Polyacrylamide Gel Electrophoresis	31
2.3 DNA Cloning	32

2.3.1 Transformation	32
2.3.2 DNA-preps	32
2.4 Cell culture.....	33
2.4.1 Proliferation.....	33
2.4.2 Freezing	33
2.4.3 Transfections	33
2.4.4 Fixation and Immunofluorescence.	34
2.5 In-Vivo Experiments.....	34
2.5.1 In vivo electroporation	34
2.5.2 Animal research ethics	36
2.5.3 Muscle sectioning.....	36
2.5.4 Immunofluorescence of mouse muscle sections	37
2.6 Histology.....	38
2.7 Enzymatic Tests	38
2.7.1 Transglutaminase Enzymatic test.....	38
2.7.2 Protease Enzymatic test.....	38
2.8 KY Protein Purification	39
2.8.1 Sample Prep.....	39
2.8.2 Column Prep.....	39
2.8.3 Analytical Size Exclusion	40
2.8.4 Size Exclusion	40
2.9 Statistical analysis and Figure preparation.	41
2.9.1 Quantification of C2C12 speckled phenotype	41
2.9.3 Statistical tests performed on Enzymatic Assays.....	42
2.9.4 Quantification of colocalization	43
2.9.5 Complementation quantification.	44
2.10 <i>In-Silico</i> Analysis	45
2.10.1 Sequencing and Alignment	45
2.10.2 Phylogenetic Analysis	46
Chapter 3: <i>In-vitro</i> analysis of the mouse KY	47
3.1 Overexpression of <i>Ky</i> variants within C2C12 mouse myoblasts.....	48
3.2 Purification of <i>Ky</i> Protein constructs.....	50

3.3 Transglutaminase Assay's on Purified KY protein.	51
3.4 Protease Assay's on Purified KY protein.	52
Chapter 4: <i>In-vivo</i> rescue tests with KY TGN/PROT domain variants	54
4.1 Histological analysis of the <i>Ky</i> deficiency seen within the C3 strain of mice. .	55
4.2 In-vivo localisation of recombinant KY proteins.	57
4.3 In vivo complementation of a <i>ky/ky</i> deficiency	60
Chapter 5: The human KY	64
5.1 <i>In-silico</i> human <i>KY</i> analysis	65
5.2 In vitro overexpression of human KY	69
5.3 Co-localisation of IGFN1_V1 and the human KY	71
5.4 Co-Transfection of IGFN1_V1 and the mouse KY	74
5.5 In Vivo localisation of human KY	76
Chapter 6: Discussion and Future work	80
6.1.1 KY does not contain any enzymatic activity within the conserved TGN/PROT domain	81
6.1.2 KY's TGN/PROT domain is not required for Z-disk localisation	82
6.1.3 The human KY can localise to the Z-disk	82
6.2 Overview	83
6.3 KY's TGN/PROT domain	84
6.4 KY's Z-disk Localisation.....	86
6.5 The human <i>KY</i> gene	86
6.6 The Nuclear localisation signal conserved within primates	87
6.7 IGFN1 and KY's Interaction	89
6.8 <i>Ky</i> as a regulator of protein turnover	90
6.9 Future Work.....	91
6.9.1 <i>In-Silico</i> Analysis	91
6.9.2 <i>In-vitro</i> Experiments	91
6.9.3 <i>In-vivo</i> Experiments	92
7. Reference list	93
8. Appendix	103

Table of Figures

Figure 1. Schematic diagram of skeletal muscle structure	10
Figure 2. The hypertrophic PI3K/Akt signalling pathway	13
Figure 3. The Myostatin atrophy signalling pathway	16
Figure 4. Polyacrylamide gels loaded with PCR amplified Ky.....	31
Figure 5. Mouse electroporation on a wildtype mouse. A) DNA microinjection	35
Figure 6. Whole Mount Images of successfully electroporated muscle fibres	36
Figure 7. Counting of cells and cells showing a speckled phenotype.	42
Figure 8. Co-localisation measurements between two proteins	43
Figure 9. Quantification of cross-sectional area for transfected and untransfected fibers.	44
Figure 10. The parameters used within MEGA-X statistical tests	46
Figure 11. Mutagenised KY Constructs	48
Figure 12. Overexpression of KY constructs within C2C12 myoblasts.	50
Figure 13. Schematic diagrams of purified KY constructs.....	51
Figure 14. KY's TGN/PROT domain holds no transglutaminase enzymatic activity	52
Figure 15. KY's TGN/PROT domain does not contain any protease activity.	53
Figure 16. Histological analysis of the soleus of Wildtype and ky/ky mice. A) Haematoxylin and eosin-stained sections from wildtype and ky/ky mice soleus.....	56
Figure 17. Z-disk localization of mutagenised KY constructs.....	59
Figure 18. Enzymatically dead versions of Ky can induce a hypertrophic response within the muscles of Ky deficient mice.....	63
Figure 19 Protein sequences of Homo sapiens hKY-X5 against the hKY.....	66
Figure 20. Conserved nuclear localization signal within primates shows high genetic divergence within other species.....	68
Figure 21. Cloned human KY isoform X5 (hKy-eGFP).	70
Figure 22. Cytoplasmic and nuclear localization of pcDNA3.1-hKy-eGFP. A-B) pcDNA3.1-hKy-eGFP transfected into C2C12 myoblasts	71
Figure 23. Nuclear localization pattern of IGFNI_V1	72
Figure 24. Colocalization of pcDNA3.1-hKy-eGFP and pDEST47-IGFNI_V1-tdTomato.....	73
Figure 25. Co-expression of pDEST47-mKy-tdTomato and pDEST47-IGFNI_V1-eGFP.....	75
Figure 26. pcDNA3.1-hKy-eGFP localizes to the Z-disk and the nucleus in vivo.....	78
Figure 27. IGFNI_V1 localises to the Z-disk.....	79

Acknowledgements

I would like to express my sincerest appreciation to Dr Gonzalo Blanco for giving me the opportunity to be involved with this research. His counsel, expertise and unwavering support have been invaluable and for that I am deeply indebted. I have thoroughly enjoyed the past two years working in the Blanco Genetics Laboratory and will miss working alongside you.

In addition, I would like to thank Amy Brown, my dear friend, colleague, and mentor. Under her guidance I continued to develop the necessary skills to complete my research. Her sense of humour and incredible work ethic have seen me through many trying situations.

I would also like to thank my TAP member Dr Sean Sweeney, who has patiently sat through TAP meetings and provided me with support and useful advice! Furthermore, I would like to thank the York Protein Production Laboratory for all their hard work, the Imaging and teaching staff for being extremely accommodating and helpful, and all of the technicians and store teams who keep us up and running.

Finally, I would like to thank my Family, especially my parents, without whom I would not have been in a position to complete my research or thesis. My parents have been a constant pillar of support and guidance, and without them I certainly would not be in the position I am today. I cannot thank you both enough!

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. Any information which has been derived from other sources has been explicitly referenced in the text. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

Chapter 1: Introduction

1.1 Skeletal muscle structure

1.1.1 Skeletal muscle structure

Skeletal muscle is built up from bundles of muscle fibers, more commonly known as fascicles. Each fascicle is encased by the perimysium; connective tissue which provides support and integrity to the fascicle. Further stability and structural integrity are provided by another layer of connective tissue, known as the epimysium which encases the perimysium. Each muscle fiber is composed of bundles of myofibrils, muscle fibers are encased by the sarcolemma; a cell membrane which provides a scaffold to which myofibrils adhere (Figure 1.A). The basic units of myofibrils are sarcomeres whereby the normal organisation of sarcomeric proteins gives muscles their distinctive pattern. The sarcomeres are defined by their Z disks (Clark, *et.al.*, 2002), which form the boundaries of the sarcomere. The width of the Z-disk varies depending on fiber type, fast fibers having narrow Z-disks (~30-50nm), whilst slow fibers have wider Z-disks (~100nm). Z-disks are essential for contraction, the transfer of tension, and signalling (Luther, *et.al.*, 2003). Thin actin filaments are tethered to the Z-disk via the positive end of their filament (barbed end). Thick myosin filaments are contained between these actin filaments which are extruding from the Z disk (Figure 1.B) (Clark, *et.al.*, 2002). Skeletal muscle is adhered to the bone via tendons, typically occurring at each end of the muscle. Tendons allow for the conversion of contractile forces into movement.

The Z-disk is arguably the most important structure within the context of skeletal muscle, providing an anchor for proteins and a structural linkage along the muscle fiber (Luther, 2009). Thus, allowing for the transmission of tension and contractile forces along the fiber. The Z-disks are extremely strong structures, able to withstand large forces applied through them, throughout muscle contraction. Furthermore, an extremely large repertoire of proteins and their interacting partners localize to the Z-disk, allowing many functions to occur including mechanosensing and signalling (Luther, 2009). Mutations within many of these proteins have a profound effect, causing a range of cardiac and skeletal muscle diseases (Frank and Frey, 2011). The identity of the mechanism regulating stretch sensors is still elusive, but it is widely

believed these sensory proteins are located at the Z-disk (Knöll, *et.al.*, 2011). In order for a transcriptional response to occur these sensors would have to communicate with the nucleus, either through the action of signalling proteins, or through protein translocation to the nucleus.

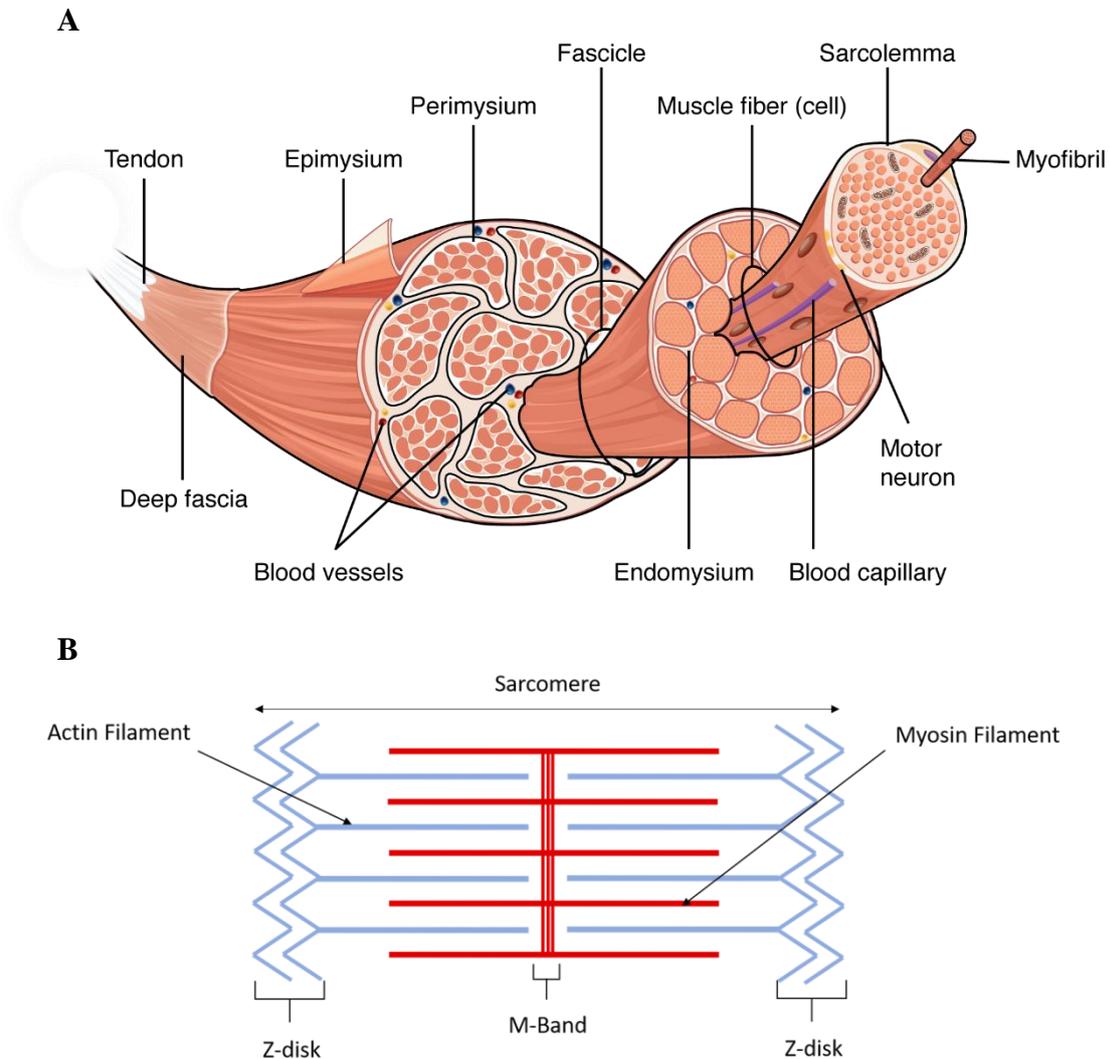


Figure 1. Schematic diagram of skeletal muscle structure. A) Muscle structure from bone attachment to myofibrils. Tendons are used to adhere muscles to bone. The muscle is encased by the epimysium. The muscle itself splits into different fascicles, which are encased by the perimysium. Each fascicle is built up of multiple muscle fibers which are encased by the sarcolemma. This image is available at <https://open.oregonstate.edu/aandp/chapter/10-2-skeletal-muscle/>. Muscle fibers are formed through a multitude of myofibrils **B) Sarcomere structure.** Myofibrils are developed from horizontal and linear repeating sarcomeres. The Sarcomere is the basic contractile unit, shown is the key structural components.

1.1.2 Muscle Fibers.

Muscle Fibers are divided into multiple fiber types, the proportion of these fibers varies significantly between individuals and species (Simonea and Bouchard, 1989; Klont, *et.al.*, 1998). The muscle fiber type is determined by the differing isoforms of myosin heavy chain (MHC) that are expressed, alongside myosin light chain (MLC), and the myosin binding protein C (MBC), together these proteins form the thick myosin filament within the fiber. Originally muscle fibers were divided into two classes, type I (slow) and type II (fast). Type I fibers are oxidative, producing slow sustained contractions and are able to resist fatigue, these fibers are known for maintaining correct posture (Pocock, *et.al.*, 2013, p.125-126). Conversely, type II fibers are divided into two distinct groups, type IIa fibers which are an intermediate being oxidative and glycolytic. Type IIb are glycolytic, producing powerful short contractions required for dynamic movement (Pocock, *et.al.*, 2013, p.125-126). More recently, other intermediate fibers have been identified and can be characterised through myosin ATPase histochemical staining (Scott, *et.al.*, 2001). Most skeletal muscles contain a mixture of both type I and type II fibers, although some muscles predominantly express one type. The soleus muscle is primarily composed of type I fibers, while the extensor digitorum longus muscle is primarily type II fibers.

Muscle fibers are formed through the fusion of multiple myoblasts. Cell fusion occurs when two Z bodies (small complexes of alpha actinin and associated proteins) grow, fuse, and align to form the Z disk, allowing development of the muscle fiber (Sanger, *et.al.*, 2006). This process requires dramatic remodelling of the cytoskeleton allowing formation of a large, interconnected network between sarcomeres (Wang, *et.al.*, 2005). Muscle fibers are multinucleated and post-mitotic. It has been theorised that a single nucleus has a limited synthetic capacity and can only regulate the gene expression within a select region (Myonuclear domain (MND)) within the myofiber (Hall and Ralston, 1989; Pavlath, *et.al.*, 1989). Although a single myonuclei is thought to control the gene expression within the MND, a uniformity of gene expression is seen across the muscle fiber, suggesting some cross talk between these regions. Conversely, it has been reported that different MND's show differing levels of gene expression across

the muscle fiber (Wilkins, *et.al.*, 2001). Myonuclei located at neuromuscular junctions and myotendinous junctions, express a specialized subset of genes relevant to these specialised regions of the muscle. This suggests that each MND is under the control of a single myonuclei, and expresses a subset of specialized genes needed for that region. A more recent study has suggested that not all myonuclei are positioned equally, and that there are multiple subsets of myonuclei, each having specialized rules that control myonuclei spacing (Perillo and Folker, 2018). This may explain the differences in gene expression which are observed along the myofiber.

1.2 Skeletal muscle hypertrophy

Muscular cells are postmitotic, they are long lived and require constant maintenance and upkeep. The active nature of muscle cells (myocytes) allows the generation of mechanical forces, which in turn produces a large variety of stressors, such as increased temperature, membrane lesions, and cellular tension (Morton, *et.al.*, 2009). Skeletal muscle is extremely plastic and can go through vast changes, these changes are dependent on the external and internal stimuli muscles receive. Muscles adapt to the functional requirements put upon them, adapting through changes in muscle mass, fiber size, and the type of fibers expressed.

Muscular hypertrophy is one form of adaptation, with prolonged resistance training muscle hypertrophy is prominent, leading to an increase in muscle size and thus strength. In untrained individuals muscle hypertrophy is nearly non-existent. Muscle hypertrophy increases muscle fiber size through the addition of contractile elements, such as actin and myosin, which leads to a larger cross-sectional area of the muscle fiber. Contractile hypertrophy occurs through the addition of new sarcomeres in series or in parallel with the fiber (Vierck, *et.al.*, 2000). Further to this increase in fiber cross-sectional area, expansion of the extracellular matrix (ECM) is observed which provides further support to these larger fibers (White, *et.al.*, 2009). Conversely to this, muscle growth can occur through the addition of new muscles fibers (hyperplasia).

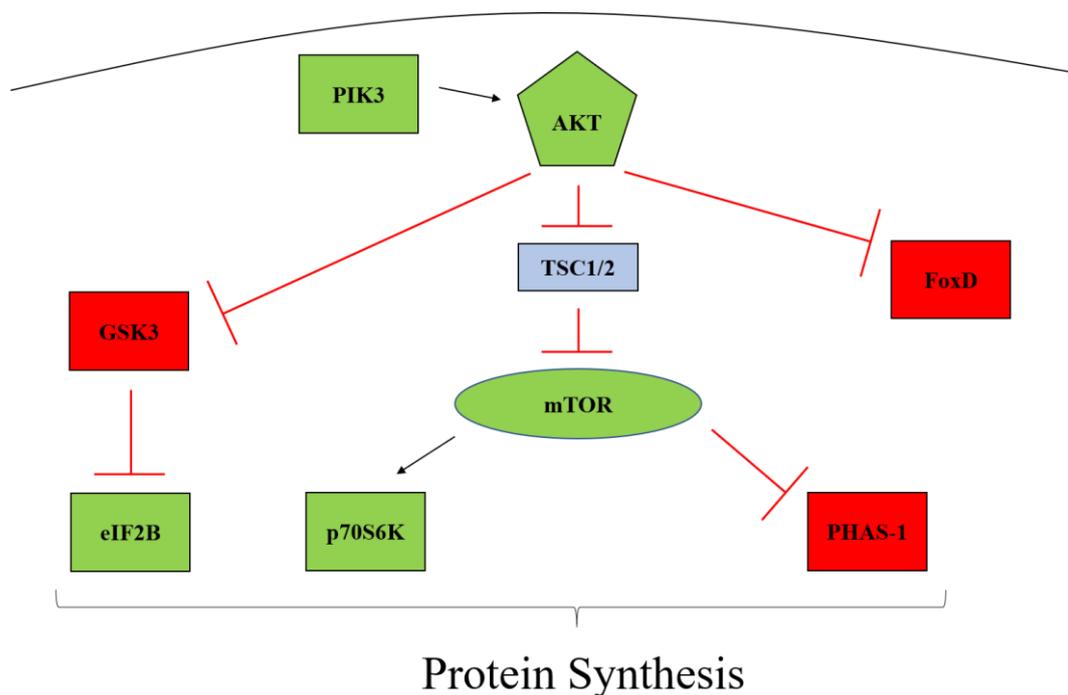


Figure 2. The hypertrophic PI3K/Akt signalling pathway. A simplified schematic of the phosphatidylinositol-3 kinase (PI3K) / Akt pathway. Up-regulators of muscle hypertrophy are shown in green whilst down regulators are shown in red. Intermediates are shown in blue. The arrows signify protein activation, while the red lines with a horizontal line signify protein inhibition. The curved line represents the myonuclei membrane. This image has been modified from (Glass, 2005), <https://www.sciencedirect.com/science/article/pii/S1357272505001317?via%3Dihub>.

The protein derived growth factor ‘insulin-like growth factor 1’ (IGF-1) has been shown to induce hypertrophy in muscle fibers (Musrò, *et.al.*, 1999), this occurs through the stimulation of the phosphatidylinositol-3 kinase (PI3K) / Akt pathway. Alongside inducing hypertrophy, IGF-1 can also block transcriptional upregulation of key mediators of muscle atrophy (protein degradative pathway). IGF-1 binds to the tyrosine receptor IGFR1 causing the phosphorylation of Insulin Receptor Substrate 1 (IRS-1), which in turn activates PI3K (Rommel, *et al.*, 2001). Subsequently PI3K activates Akt through phosphorylation. Akt has a large plethora of functions, this includes phosphorylation of tuberous sclerosis complex-1 and -2 (TSC1/2) proteins, leading to their inhibition. Inhibition of TSC1/2 releases the inhibition upon mTOR,

which has been shown to have an imperative role in integrating a variety of growth signals, such as nutritional stimulation and growth factors (Panzhinskiy, *et.al.*, 2013).

Activation of mTOR leads to protein synthesis, allowing growth of the fiber. Akt further aids the production of proteins utilising a distinct pathway, which is separate to the mTOR pathway. Akt inhibits GSK3, releasing the inhibition on the protein initiation factor eIF2B, allowing for an increase in protein synthesis, and aiding muscle hypertrophy. Furthermore, Akt activation leads to the inhibition of the Forkhead box transcription factors (FOXO) family. The FOXO family has a significant role in inducing atrophy (Sandri, *et.al.*, 2004). This inhibition suggests, alongside inducing hypertrophy, that the PI3K/Akt pathway can significantly inhibit atrophy initiation.

The initiation of exercise induced hypertrophy remains elusive, it is hypothesised that there are three main factors responsible for this initiation: mechanical tension, metabolic stress, and muscle damage (Schoenfeld, 2010). Mechanical tension is produced through force production and stretch. Both of these stimuli have a large effect on increasing muscle growth. To date, no definitive mechanosensory proteins have been identified, although a few candidates have been suggested such as Titin and FLNC (Wackerhage, *et.al.*, 2018), both of which are Z-disk localizing proteins. Nucleocytoplasmic shuttling proteins such as Muscle LIM protein (MLP) which is encoded by Cystine and Glycine-rich protein 3 (Csrp3), have also been implicated as potential stretch sensors (Boateng, *et.al.*, 2009; Vafidaki, *et.al.*, 2015). A multitude of studies support metabolic stress as having a more important role in determining the hypertrophic response (Schoenfeld, 2013). The accumulation of metabolites such as lactate, initiates a stress induced mechanism which has been theorised to induce a hypertrophic response. Finally, muscular damage which can result from resistance training is thought to cause hypertrophy. This damage occurs in multiple forms; tears in the sarcolemma, basal lamina and supportive tissue are all seen. Alongside injury to contractile elements, and the cytoskeleton (Schoenfeld, 2010). This damage is believed to induce an inflammatory response, in which growth factors are released and in turn induce hypertrophy.

1.3 Skeletal muscle atrophy

Muscular atrophy is a loss of muscle mass through active protein degradation. Muscle atrophy is prevalent in a multitude of diseases, such as Chronic kidney failure and heart disease. The most prevalent form of muscle atrophy is seen within sarcopenia where muscle loss is observed through the process of ageing. Atrophy is also seen in response to certain stimuli, such as starvation or immobilisation of a limb, both of these can cause a varying range of muscular atrophy.

Myostatin is a member of the transforming growth factor beta (TGF β) superfamily. Overexpression of myostatin has been shown to induce muscle atrophy, myostatin null mice show vastly increased muscle mass (McPherron, *et.al.*, 1997). Myostatin is therefore a key factor which promotes muscular atrophy. Myostatin has been shown to bind to the membrane bound activin type II receptor (ActRII), which in turn leads to the phosphorylation of Smad2/3 (Rebbapragada *et.al.*, 2003). Also, phosphorylated Smad2/3 can form a heteromeric complex with Smad 4 which is able to translocate to the nucleus and inhibit gene transcription (Goodman and Hornberger, 2014). Furthermore, phosphorylated Smad2/3 can inhibit Akt activity, thereby preventing Akt induced hypertrophy and protein synthesis. The inhibition upon Akt releases inhibition upon the FOXO family. The FOXO transcription factors induce expression of MuRF1 and MAFbx (Bodine *et.al.*, 2001), genes involved within the upregulation of the ubiquitin-proteasome system. The ubiquitin-proteasome pathway is the primary pathway utilised within muscle atrophy, alongside other degradative pathways. Myostatin has also been shown to activate the Autophagy-lysosome system (Wang, *et.al.*, 2015), another distinct protein degradation pathway utilised within muscle atrophy. Myostatin has also been linked to the inhibition of PI3, the upstream activator of Akt, placing more inhibition upon the IGF-1 hypertrophy pathway (Morissette *et.al.*, 2009).

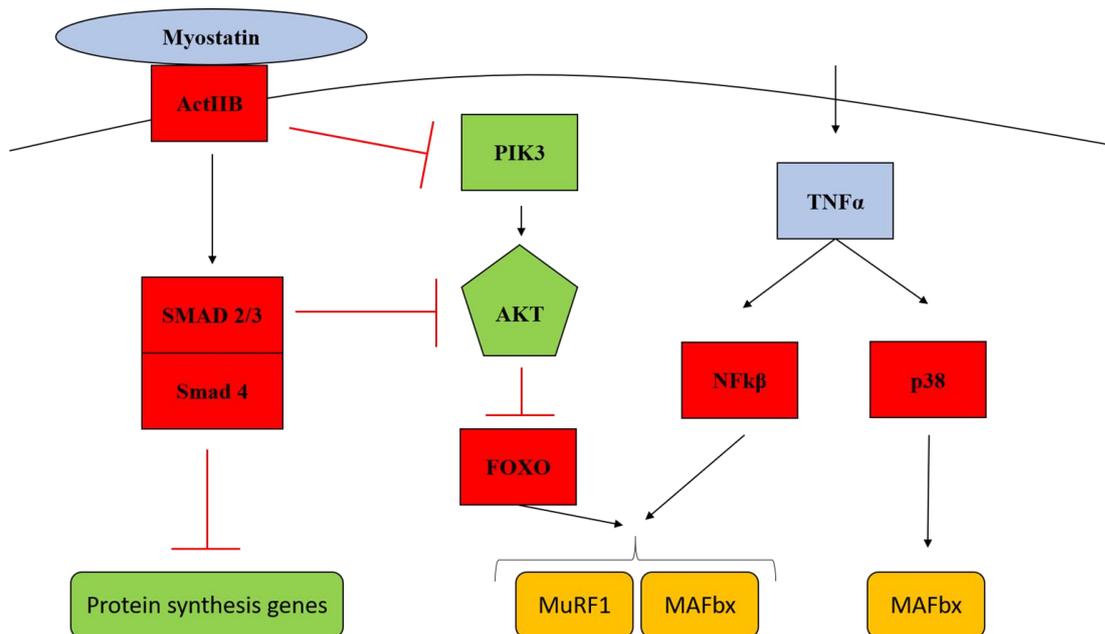


Figure 3. The Myostatin atrophy signalling pathway. A simplified schematic of the atrophy promoting pathways. Upregulators of atrophy are shown in red, whilst downregulators of atrophy are shown in green. Intermediates are shown in blue. The arrows signify protein activation, while the red lines with a horizontal line signify protein inhibition. The curved line represents the myonuclei membrane. This image has been modified from an original image (Glass, 2005) and is available at <https://www.sciencedirect.com/science/article/pii/S1357272505001317?via%3Dihub>

Muscular myopathies and a multitude of diseases are known to lead to a loss of skeletal muscle, within these cases pro-inflammatory cytokines are known to rise. TNF- α is a polypeptide cytokine which is linked to antitumor immune responses. TNF- α has been shown to trigger multiple cell responses and has been associated with muscular pathology (De Larichaudy, *et.al.*, 2012). Importantly, TNF- α plays a significant role in promoting muscle atrophy. TNF- α binds to its receptor inducing the activation of I κ B Kinase complex (IKK) through phosphorylation, this in turn leads to ubiquitination and degradation of I κ B through the proteasome. I κ B degradation leads to activation of NF- κ B (Yaron *et.al.*, 1998). Mice with constitutively active I κ B have shown an upregulation of MuRF1, and when these mice are crossed into a MuRF1-/- background, there was a significant reduction in muscle atrophy. Thus, showing that

the removal of MuRF1 inhibits the atrophic ability of I κ B signalling (Cai *et.al.*, 2004). NF- κ B also upregulates MAFbx through the p38 MAPK pathway (Ho, *et.al.*, 2005; Zhang and Li, 2012).

The Ubiquitin-proteasome system (UPS) plays an important role in muscle atrophy which has been thoroughly established (Bilodeau, *et.al.*, 2016). Protein ubiquitination occurs throughout the body and is used as a signal for protein degradation. The 26S proteasome complex recognises ubiquitinated proteins and in an ATP dependent manner, leads to the proteolytic degradation of a variety of proteins. Ubiquitin chains are added to proteins via three enzymes: ubiquitin activating E1, ubiquitin conjugating E2, and ubiquitin ligase E3. Ubiquitin activating E1 binds with a ubiquitin molecule, passing the ubiquitin to ubiquitin conjugating E2. Ubiquitin conjugating E3 allows the linking of the Ubiquitin to a lysine residue of the target protein. The FOXO transcription factors promote expression of ubiquitin ligases. FOXO's have also been shown to induce the autophagy/lysosome system (Bilodeau, *et.al.*, 2016).

Autophagic/lysosomal proteolysis has been shown to play a significant role in protein degradation within muscles. Inhibition of autophagic proteins leads to a dystrophic muscle through the build of protein aggregates and damaged organelles (Masiero and Sandri, 2010). Autophagy is a mechanism that takes cytoplasmic material and organelles and delivers them to a lysosome for degradation, damaged mitochondria and protein aggregates are cleared through autophagy. Multiple UPS genes are also able to upregulate genes involved within autophagy. Chaperone assisted selective autophagy (CASA) is a selective type of autophagy which is a tension induced degradative pathway essential for muscle maintenance, and protein turnover (Arndt, *et.al.*, 2010). Instead of protecting and refolding damaged Z-disk proteins such as FLNC, these proteins are targeted for degradation while inducing the expression of new FLNC. CASA also facilitates the degradation of these damaged proteins. The chaperone associated ubiquitin ligase CHIP and p62 initiate CASA by ubiquitinating the target protein. BAG3 has been identified as being a key regulator of the CASA system, specifically coordinating the activity of the small heat shock proteins HspB8 and HspB8 during the degradation of proteins post ubiquitination (Guilbert, *et.al.*,

2018). Further to this, mutations within BAG3 cause a severe form of myofibrillar myopathy, characterised by protein aggregation, and muscle weakness. BAG3 has been shown to localise to the Z-disk, where it is imperative for FLNC turnover. BAG3 also interacts with factors of the hippo system in order to induce FLNC transcription, thus aiding the replenishment of FLNC, further aiding Z-disk stabilisation (Jokl, *et.al.*, 2018). This mechanism suggests BAG3 is therefore a critical factor underpinning normal FLNC turnover in tension bearing cells and thus sarcomeric maintenance.

1.4 Kyphoscoliosis peptidase

1.4.1 The *ky/ky* Mouse

Kyphoscoliosis peptidase (*Ky*) is imperative for normal muscular function. The first reported *Ky* deficiency (*ky/ky*) developed through a sporadic mutation within the BDL strain of mice. This *ky/ky* mouse showed the development of kyphoscoliosis which could be detected quickly post weaning (Dickinson and Meikle, 1973). The *Ky* mutation was identified through positional cloning, revealing a premature STOP codon caused by a CG deletion (Blanco, *et.al.*, 2001). All *ky/ky* muscles are smaller and weaker than controls. This inherent muscular weakness likely contributes to the spinal abnormalities observed. Weakening of the paraspinal muscles contribute to this development, alongside preventing mice from reaching towards a ledge during the placing response test (Blanco, *et.al.*, 2001) (Appendix). To date no phenotype has been observed throughout the embryonic development stage, highlighting the need for physical stressors and mechanical load to reveal muscular abnormalities, which are quickly observed post weaning.

On average *ky/ky* mice have a body weight that is ~80% of the weight of the wildtype mouse. During postnatal muscular growth, a cycle of fiber death and regeneration is extremely prevalent in the postural muscles, such as the soleus and paraspinal muscles. This is evident due to the vast number of internalised nuclei present within *ky/ky* fibers, which is a hallmark of fiber regeneration. Intriguingly, these muscles show an adaptive shift of the contractile protein isoforms, typical to that of type I slow muscles (Marechal, *et.al.*, 1996). This adaption likely reflects an inability to cope with the

increased demand put upon muscles, associated with a higher body weight. At an ultrastructural level, the soleus presented with Z-disk thickening, A-line streaming, and overlapping of thick filaments (Beatham, *et.al.*, 2004). The thickened Z-disk potentially reflects an inherent instability within the Z-disk, leading to an adaptive over-production of Z-disk proteins to aid stabilisation. Additionally, accumulation of sarcomeric proteins such as Filamin C (FLNC) and Xin aggregates are commonly seen within *ky/ky* myofibers, reflecting a reduced ability to degrade proteins. Furthermore, *ky/ky* muscles are unable to induce hypertrophy in response to mechanical overload (Blanco, *et.al.*, 2001).

The Z-disk co-chaperone BAG3 has been shown to be upregulated in *Ky* deficient mice and Zebrafish models (Jokl, *et.al.*, 2018). As previously mentioned BAG3 is part of the CASA pathway, aiding the degradation and potential replenishment of FLNC. The authors also suggested that the BAG3 dysregulation seen could be the causative reason behind FLNC aggregation seen in *ky/ky* models. RNA profiling of the *ky/ky* mouse has shown an upregulation of stretch response proteins Ankyrin repeat domain 2 (Ankrd2) and Cystine rich protein 3 (Csrp3). Further to this, an upregulation of the energy transducing proteins Uncoupling protein 1 and 2 (Ucp1/2) is observed (Blanco, *et.al.*, 2004). The upregulation of BAG3 seems to be a significant hallmark of the *ky/ky* phenotype, likely contributing to the large protein aggregates seen within these models. Again, the upregulation of the stretch receptors Ankrd2 and Csrp3 show an inability of the muscles to respond to mechanical overload, highlighting *Ky*'s critical function in adapting and responding to mechanical stresses.

1.4.2 The *Ky* Gene

The mouse *Ky* gene encodes a 72 kDa protein which is expressed within skeletal muscle, the heart, and at a much lower level within the CNS (Blanco, *et.al.*, 2001). The Kyphoscoliosis peptidase protein (KY) contains a predicted transglutaminase/protease (TGN/PROT) domain between residues 171 and 282, identified through a PSI-BLAST database search (Blanco, *et.al.*, 2001). Thus, placing KY in the family of transglutaminase-like proteins. To this day there is no sufficient evidence to claim that this region holds any enzymatic activity. Transglutaminases are conserved across

many species, most of the proteins within this superfamily are predicted to be enzymatically active, utilising a catalytic triad. This triad consists of a cysteine, a histidine, and an aspartate (Makarova, *et.al.*, 1999), highly resembling the catalytic site of papain-like proteases (Anantharaman, *et.al.*, 2001). Transglutaminase activity enables the catalysation of proteinase resistant isopeptide bonds, leading to the cross linking of proteins. These enzymes require the binding of Ca^{2+} in order to initiate their activity (Griffin, *et.al.*, 2002). KY has homology to multiple proteins from the transglutaminase family, all of which contain a domain homologue to the human transglutaminase (Blanco, *et.al.*, 2001). Many of the eukaryotic transglutaminases are thought to have been derived from ancestral proteases (Makarova, *et.al.*, 1999), providing evidence that KY may contain protease activity. Conversely, some conflicting research has suggested conserved transglutaminase domains are not enzymatically active but have been re-assigned for a different role, such as mediating protein-protein interactions (Anantharaman, *et.al.*, 2001). In-vivo KY is known to localise to the Z-disk (Beatham, *et.al.*, 2004), likely contributing to the Z-disk abnormalities observed within the *ky/ky* mouse.

1.4.3 KY's Interacting protein partners

KY has been shown to interact with several sarcomeric proteins, typically immunoglobulin rich proteins, which include FLNC, Immunoglobulin-like and fibronectin type III domain containing 1 (IGFN1), and the myosin binding protein C (Beatham, *et.al.*, 2004). These proteins all have a significant role within maintaining muscle integrity, allowing for normal structural organisation and function. KY has previously been implicated as a structural protein, forming a large protein complex with IGFN1 and FLNC which was shown to localise to the Z-disk (Baker, *et.al.*, 2010). The interaction between KY and FLNC has been extensively explored. *Ky* deficient mice and humans both show abnormal FLNC distribution, alongside large FLNC protein aggregates when compared to the wildtype muscles (Hedberg-Oldfors, *et.al.*, 2016). This highlights KY's importance in maintaining normal FLNC distribution and potentially in regulating FLNC protein degradation through BAG3. A series of co-

transfections performed in C2C12 myoblasts also showed KY may be able to modulate the levels of FLNC (Beatham, *et.al.*, 2004).

1.4.4 Human cases of KY mutations

In more recent years mutations of KY have been identified within patients presenting with novel myopathies. These patients all show a remarkable similarity to the *ky/ky* mouse, suggesting KY's role is highly conserved between species. *Ky* deficient muscles within humans and mice both show Xin and FLNC protein aggregates. Histological studies within patients also revealed a huge diversity in the size of muscle fibers, many of which were very small with internalised nuclei (Appendix). These features are commonly seen within myofibrillar myopathies which are characterised by the formation of aggregates, and the dysregulation of Z-disk proteins (Palmio and Udd, 2016).

The first reported case of a KY deficiency was seen in a seven-year-old girl. A one base deletion in KY resulted in a frame shift and led to a premature STOP codon. This KY variant was shown to be expressed within skeletal muscle; therefore, this variant must have encoded a non-functional truncated protein. The girl presented with walking difficulties, generalised muscle weakness and mild contractures within the hips, shoulder, and feet. She had undergone Achilles tendon elongation surgery. Tissue biopsies revealed muscle atrophy and extreme fiber size variability (Appendix). Furthermore, she had internalised myonuclei, and further studies also revealed a thickened Z-disk's alongside sarcomeric disorganisation. Large protein aggregates of FLNC and Xin were seen, again suggesting dysregulation of protein degradation (Hedberg-Oldfors, *et.al.*, 2016)

A KY deficiency was next seen to cause a congenital myopathy within two Arab Israeli brothers from a first cousin marriage. A homozygous variant was shared between these brothers, where a premature STOP codon in exon 6 results. Both patients presented with muscle weakness and mild atrophy of the lower limbs. The younger brother was affected more severely than the older brother. The younger brother showed muscle weakness and atrophy which extended into his upper limbs, further to this Kyphosis

was shown to develop. He also showed intellectual impairment, anxiety, and impulsive behaviour. Cranial and spinal MRI's revealed normal cardiological functions. Pathological studies of the quadriceps at 17 years of age, showed split fibers, fiber size variability and internalised nuclei. Interestingly, type II fiber uniformity was seen, conversely to type I fiber uniformity in mice. FLNC aggregation was again seen within these biopsies. Electron microscopy further revealed Z-disk thickening as well as an enlarged endoplasmic reticulum (Straussberg, *et.al.*, 2016).

The next reported cases of a *KY* mutation were identified within twelve Bedouin individuals, these individuals were part of the same highly inbred tribe, presenting with hereditary spastic paraplegia. A *KY* homozygous variant was identified within this family. This variant was caused by a disruptive single base pair insertion within the first exon of *KY*. These patients showed spasticity of the lower limbs, showing toe walking and equinus deformity. Within the older patient's kyphoscoliosis was shown to develop, most patients showed atrophy of the tongue. Alongside a few patients who also showed an intellectual disability. Pathological studies revealed fiber size variability and centralised nuclei. High expression of this *KY* variant was seen within the muscles and to a lower level in the central nervous system (CNS). Again, a clear muscular pathology is observed within these cases, with patients showing muscular weakness and spasticity. Furthermore, most patients developed kyphosis and or scoliosis, further implicating muscular weakness as the primary cause of this pathology (Yogev, *et.al.*, 2017).

The final case was found within an Iranian patient who presented with myofibrillar myopathy. A new nonsense mutation was identified within *KY*. The patient was a 29-year-old male, symptoms of muscular weakness presented at the age of 3. His lower limbs showed muscle weakness and atrophy, alongside having equinovarus foot deformity. Mild scoliosis and joint contracture were also identified. Bloods were taken, his serum showed abnormally high creatine phosphokinase (CPK) and lactate dehydrogenase (LDH). CPK is an enzyme found in the brain, heart, and skeletal muscles. Serum CPK can be elevated as a normal response after exercise, other causes include muscular inflammation and muscular dystrophy (Lasich, 2014). LDH is found

across your whole body, and normally rises after physical exertion. Serum LDH has been shown to rise in response to muscle damage and muscular myopathies (Haller and DiMauro, 2012). His sister also presents with generalised muscle weakness, alongside an aunt who had passed away.

The cases described above all show remarkable similarities and some subtle differences. In all cases vast fiber size variability is seen alongside internalised nuclei which is a marker of regeneration. *KY* deficient muscles all showed a thickened Z-disk (Beatham, *et.al.*, 2004), further implicating a primary involvement of the Z-disk within the *ky/ky* pathology. Further to this, patients deficient in *KY* show aggregation of FLNC and Xin. FLNC is a structural cross-linker critical for normal muscle pathology; mutations and dysregulation of FLNC cause severe myopathies (Leber, *et.al.*, 2016; Chen, *et.al.*, 2019). It is interesting to note the *KY* variant seen in the 12 Bedouin individuals lead to spasticity, and a lack of control of the lower limbs. The authors also argued that the lower expression of *KY* within the CNS explained the neurological and cognitive aspects within this pathology. It is clear from the presented cases *KY* has a distinct role in maintaining a normal muscular pathology. These cases also indicate *KY* has a significant role in maintaining normal protein regulation, shown by the dysregulation of protein turnover and build-up of protein aggregates. Importantly, one case here reported that *KY* localised to the sarcolemma of the muscle fiber, not the Z-disk (Straussberg, *et.al.*, 2016). This seems unlikely considering the high homology of *Ky* observed between species.

1.4.5 Myofibrillar myopathies

The phenotype seen in *KY* deficient patients are indicative of myofibrillar myopathies (MFMs), which generally are autosomal dominant myopathies with a late onset of progressive muscle weakness (Ruparelia, *et.al.*, 2012). A wide variety of proteins can be mutated within MFMs all of which localise to the Z- disk, linking it to have a key role within the pathology of MFM. Common hallmarks of MFMs include muscle fiber breakdown, protein aggregation, and centralised nuclei. Interestingly, characterization of the protein aggregates determined many critical sarcomeric proteins such as FLNC, Titin, and BAG3 to be part of these aggregates (Ruparelia, *et.al.*, 2012). The Z-disk

has been suggested to be the primary site affected in MFMs (Selcen and Engel, 2011). The data reviewed here suggests that the myopathy caused by *KY* mutation directly places it within the MFM category.

1.5 Immunoglobulin-like and fibronectin type III domain containing-1 (IGFN1)

1.5.1 Overview of IGFN1

Immunoglobulin-like and fibronectin type III domain containing 1 (IGFN1) was one of the first proteins to be identified to interact with *KY* alongside *FLNC*. These interactions were identified using a yeast two-hybrid assay (Baker, *et.al.*, 2010). Like *Ky*, *Igfn1* is highly conserved among species. *Igfn1* is specifically expressed within the skeletal muscle and is a complex gene, supporting many isoforms. The largest isoform (IGFN1) comprises an N-terminus containing three globular domains, a large unstructured region separating the C- and N-terminus, and a C-terminus which contains eight globular domains. The domain composition of IGFN1 resembles other sarcomeric proteins such as *FLNC*, which are associated with the actin cytoskeleton, and used to maintain protein interactions after contraction. IGFN1 has been shown to localise to the Z-disk and the nucleus. IGFN1 also lacks any enzymatic activity likely acting through the formation of protein complexes, such as the IGFN1-*FLNC*-*KY* complex (Baker, *et.al.*, 2010). IGFN1 is highly expressed in atrophic conditions, and correlates with myostatin signalling, which induces muscular atrophy (Rahimov, *et.al.*, 2011). Further to this, IGFN1 was suggested to down regulate protein expression via an inhibitory interaction with Eukaryotic translation elongation factor 1A (eEF1A) during times of muscle denervation (Mansilla, *et.al.*, 2008). DNA microarrays comparing gene expression from the vastus lateralis skeletal muscle in young and old healthy men, revealed a decrease in IGFN1 expression within the elderly (Welle, *et.al.*, 2003).

Previous work found a positive correlation between IGFN1 expression and myostatin which is known to promote muscle atrophy (Rahimov, *et.al.*, 2011). However, no direct evidence has shown that IGFN1 is capable of inducing muscle atrophy. IGFN1 is a known interacting partner of *ZAK*, a Kinase family with two isoforms *ZAK α* and

ZAK β . The ZAK β isoform is dominantly expressed within skeletal muscle, overexpression of ZAK β has been shown to increase muscle fiber size (Unpublished work, Li, 2016). Patients presenting with ZAK mutations, remarkably show a similar phenotype to that of *KY* deficient patients. These patients present with scoliosis, muscular weakness, fiber size variability, centralised nuclei, and a predominance of type I fibers. The typical phenotype displayed suggests that dysregulation of ZAK may contribute the phenotype observed within *ky/ky* patients. Further to this ZAK has been shown to be upregulated in the *ky/ky* mouse, providing further evidence that dysregulation of ZAK may contribute to the *ky/ky* phenotype. Recent data has also shown that IGFN1 can interact with the proteasomal subunits Psmd2,11 and 12 (Cracknell, *et.al.*, 2020). The ability of IGFN1 to localise to the nucleus and the Z-disk, provides evidence that IGFN1 could be able to respond to stimuli in a temporospatial way.

The large unstructured region IGFN1 holds is predicted to provide flexibility to the protein (Davey, 2019), therefore potentially allowing for a conformational change to occur. This conformational change could allow control over the localisation between the Z-disk and nucleus. IGFN1's correlation with myostatin's, and the ability to interact and modulate proteins involved within the UPS provides evidence that IGFN1 may be involved in protein turn over, by either recruiting or promoting expression of proteasomal proteins. Overall, the data provided suggests that IGFN1 may have a role in aiding muscular atrophy. There is also good reason to believe that IGFN1 is able to modulate gene expression through its inhibitory interaction with eEF1A.

1.6 Hypotheses

1.6.1 Does KY contain any enzymatic activity with the TGN/PROT domain?

Due to the high homology the TGN/PROT domain holds with active cysteine proteases and eukaryotic transglutaminases, it is plausible that KY may hold enzymatic ability. One inherent feature seen within the *ky/ky* pathology is Z-disk dysregulation and Z-disk thickening. If KY holds transglutaminase activity, this may aid Z-disk stabilisation, through an ability to cross-link proteins with proteinase resistant isopeptide bonds. The strong conservation of the key amino acids required for enzymatic activity suggest this domain may hold enzymatic ability.

1.6.2 Does KY's TGN/PROT domain control Z-disk localization?

KY is a known Z-disk associated protein, there is still a lack of evidence for the TGN/PROT domain to hold any enzymatic ability. The TGN/PROT domain is the most highly conserved region among species. Therefore, if KY does not hold any enzymatic ability within this region, it must be conserved to serve another function, such as aiding Z-disk localisation. As previously mentioned, some conserved transglutaminase domains are not enzymatically active but have in fact been re-assigned for a different role such as mediating protein-protein interactions (Anantharaman, *et.al.*, 2001). Therefore, it may be that KY conserves the TGN/PROT domain to aid Z-disk localisation.

1.6.3 Can the Human KY localize to the Z-disk?

The mouse KY protein has been shown to localise to the Z-disk. Little work has been published surrounding the human KY (hKY). Controversially, a human muscle biopsy showed the hKY to localise to the sarcolemma of the muscle fiber (Straussberg, *et.al.*, 2016). Due to the high level of conservation, and the similar pathologies seen within *Ky* deficient models it is unlikely there is a significant enough sequence difference to cause a drastic change to the KY protein. Therefore, it is highly likely the hKY will localise to the Z-disk.

1.7 Objectives and Aims.

KY still to this day remains highly uncharacterised and no specific role has yet been determined for KY within skeletal muscle. The primary aim of the project is to uncover the role of the highly conserved TGN/PROT domain. As the domain is highly conserved among species it must have a significant role within the function of KY. Uncovering the function of this region would give us a good insight into the function of KY, allowing us to further understand the method of action and the pathogenic mechanism contributing to the *ky/ky* phenotype. In order to study the TGN/PROT domain we will use a series of recombinant proteins, all of which contain mutations affecting the TGN/PROT domain. Two studies will be performed in-vivo, firstly assessing the ability of a series of these KY constructs to localise to the Z-disk, each construct having varying mutations within the TGN/PROT domain. Secondly assessing the ability of these constructs to rescue fibers within a new strain of *ky/ky* mouse.

In order to determine if the TGN/PROT domain is enzymatically active, York Universities Proteomics department will attempt to purify protein of the full-length KY, and a selection of a smaller region which will include the TGN/PROT domain. Thus, allowing us to test the enzymatic ability of the TGN/PROT domain within KY. If successful we will also aim to crystallise this protein in order to determine the structure.

Finally, as mentioned the hKY has been suggested to localise to the sarcolemma of a muscle fiber (Straussberg, *et.al.*, 2016). The mouse KY is known to localise to the Z-disk; thus, it is unlikely such a significant change in the protein localisation would be seen within the hKY. Loss of function *ky/ky* patients all present with Z-disk thickening, therefore suggesting a primary involvement of the Z-disk within the pathology. Furthermore, the phenotype in KY patients is highly similar to that of the *ky/ky* mouse. Thus, the evidence provided suggests the hKY is acting mechanistically similarly to the mouse KY, and if so should localise to the Z-disk like the mouse KY. Therefore, in order to better understand the human myopathies caused by loss of function KY mutations, we need to determine if the hKY does in fact localise to the Z-disk.

Chapter 2: Methods

2.1 Table of Reagents

2.1.1 Solutions and Reagents

Reagents	Formular
PBS	PBS tablets (Thermofisher, Cat #BR0014)
PBST	PBS with 0.1% Tween-20 (Sigma, Cat #P1379)
Full Serum Growth medium (GM)	DMEM (Thermofisher, Cat #31053044) with 10% FBS (Thermofisher, Cat #A4766) and 1% Penicillin Streptomycin (Gibco, Cat #15140122)
Freezing medium	10% DMSO, 40% Serum Free Growth Medium, 50% FBS
10X TBE	Diluted in 1L of ddH ₂ O: 108g of Tris base, 55g of Boric acid, and 7.5g of EDTA
TAE buffer	Diluted in 1L of ddH ₂ O: 242g of Tris base , 100ml of 0.5 EDTA, and 57.1mls of glacial acetic acid
LB Broth	Diluted in 1L of ddH ₂ O: 10g Tryptone, 10g NaCl, 5g Yeast Extract, pH 7
LB Agar	Diluted in ddH ₂ O: (Thermofisher, Cat #CM0003)
Polyacrylamide gels	15% polyacrylamide gels were made using 1.5mls of 10x TBE, 6mls of dH ₂ O, 7.5mls of 30% Acrylamide (Severn Biotech, Cat #20-2100-10), 150 μ l of 10% APS solution, and 15 μ l of TEMED
Mowiol	6mls of ddH ₂ O, 12mls 0.2M Tris, 2.4g Mowiol 4-88 (Fluka, Cat #81381), and 6g of glycerol

2.1.2 Antibodies

Primary Antibodies

Antibody	Concentration	Supplier
Anti- α -Actinin (EA53)	1:300	Abcam, Cat #ab9465
Anti-WGA (Wheat	1:250	Abcam, Cat #ab17844

Secondary Antibodies

Antibody	Concentration	Supplier
Anti-mouse-IgG (FITC)	1:100	Abcam, Cat #ab6717
Anti-mouse-IgG (TRITC)	1:100	Abcam, Cat #ab6718

2.1.3 Primers

Primer	Sequence	Purpose
Mouse <i>Ky</i> Forward	5' GGGGCCATTTGCAGCCTA	Genotyping the <i>Ky</i> colony
Mouse <i>Ky</i> Reverse	5' CGGAGAGGTTCCGATTAGCC	
Human <i>KY</i> Forward	5' CACCATGGAGCTGAAGAAG	Conformation of correct Gene
Human <i>KY</i> Reverse	5' GCCATTCACCTTTGTATTTTCAGGA	

2.2 Heteroduplex analysis

Identification of heterozygous and homozygous *Ky* mice was done using heteroduplex Analysis. Heteroduplex analysis relies on the formation of heteroduplexes during PCR amplification, the identification of these indicates a Heterozygous mouse. Notches from the mice were taken post weaning by the animal facility. Notches were then heated at 95°C in 0.05M NaOH for 25 minutes. Neutralised with 0.08M Tris-HCL and 0.2mM EDTA, samples were then quickly vortexed. Sonication was used to break down the ear notch.

2.2.1 PCR Amplification

PCR was used to amplify target sequences, using GoTaq green master mix (Promega, Cat #M7122) following the protocol provided. Primers were kindly provided by Gonzalo Blanco (Primer Table). We optimised the PCR protocol by changing the annealing temperature to 56°C. In order to verify a successful PCR amplification 1% agarose gels were run for 45 minutes at 80V. In order to identify the expected ~2.1Kb bands, we used a 1Kb DNA ladder (Biolabs, Cat #N3232L).

2.2.2 Polyacrylamide Gel Electrophoresis

After confirming the presence of a PCR product, PCR samples were loaded onto 15% polyacrylamide TBE gels, which were used to separate heteroduplexes from homoduplexes. 15% polyacrylamide gels were made using 1.5mls of 10x TBE, 6mls of dH₂O, 7.5mls of 30% Acrylamide (Severn Biotech, Cat #20-2100-10), 150µl of 10% APS solution, and 15µl of TEMED. Gels were run overnight for 16 hours at 45V. Gels were then stained with SYBR safe (Invitrogen, Cat #S33102) for 15 minutes. Gels were imaged with a BioRad EZ GelDoc system (BioRad. Cat #1708270EDU). The presence of multiple bands shows us the formation of a heteroduplex as they migrate slower through the gels, suggesting the presence of multiple alleles. Therefore, single bands were read as Homozygous wildtype or homozygous *ky/ky*. Two bands indicated Heterozygous WT/*ky* (Figure 4).

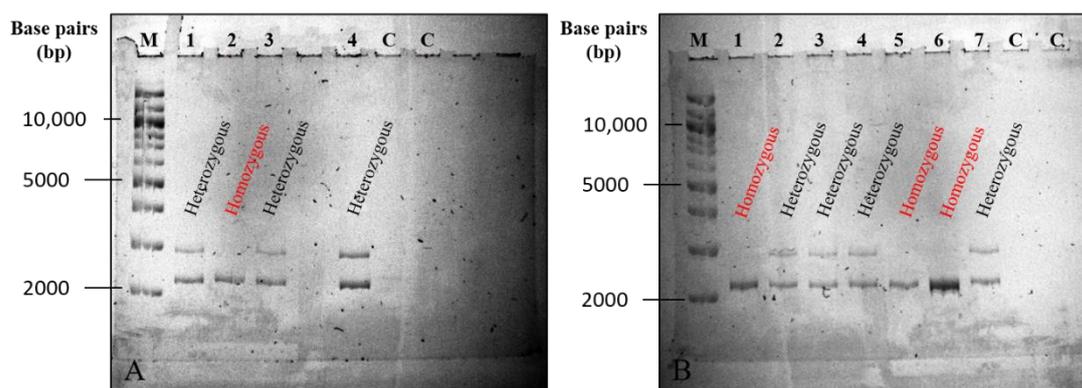


Figure 4. Polyacrylamide gels loaded with PCR amplified *Ky*. In each experiment 1Kb DNA ladders were used to ensure the correct product. Furthermore 2 PCR water controls were used in each experiment. Heterozygous (Black writing) and homozygous (Red writing) mice are marked on the gels. A) **Polyacrylamide gel 1.** Gel loaded with four test samples (3 Females; 1 Male). B) **Polyacrylamide gel 2.** Gel loaded with seven test samples (4 males; 3 Females).

This technique effectively allowed us to identify Homozygous and heterozygous mice. Visual phenotyping of the colony allowed us to successfully differentiate the homozygous WT and homozygous *ky/ky* mice.

2.3 DNA Cloning

The *ky* DNA constructs were amplified for cell transfections and *In-vivo* muscle electroporations. For successful electroporation high DNA concentrations (~1000ng/ml) were required, whilst cellular transfections required a lower DNA concentration (~300ng/ml). Therefore, we used a mixture of Maxi- and Mini-Preps, adjusting for the required DNA concentrations.

2.3.1 Transformation

Transformations were performed using Agilent competent cells and the provided reagents (Agilent, Cat # 200131), the protocol was followed precisely. Changes included LB Broth being used instead of SOC medium. Cells were thawed on ice, 50ul of cells were transferred to pre-chilled round bottomed falcon tubes and 0.850ul of B-Mercaptoetanol was added, cells were incubated on ice for 10 minutes. 2ul of DNA construct was then added to the tubes and left to incubate on ice for 30 minutes. Cells and constructs were heat shocked at 42°C for 45 seconds, then 450ul of preheated 42°C LB broth was added. Cells were then incubated at 37 degrees shaking at 225rpm for an hour. Cells were then spread on premade agar plates, which contained the relevant antibiotics and left overnight. All plates contained ampicillin, as all of our constructs had a selective ampicillin resistance.

2.3.2 DNA-preps

Single colonies were selected from overnight cultures and inoculated in a starter culture of 5 ml LB medium containing 1X the relevant antibiotic, this culture was incubated at 37°C on a shaker at 200rpm for 8 hours. Depending on the prep being undertaken, the starter culture diluted down in pre-warmed LB medium. The starting culture for mini preps was transferred into another 5mls of LB medium and left overnight. The starting culture for maxi-preps was diluted 6-fold into 30mls of prewarmed LB and left overnight. This culture was then transferred into ~800mls of prewarmed LB medium and left overnight. All cultures were left at 37°C on a shaker at 200rpm. For larger cultures 2L conical flasks were used.

Cultures were pelleted, then resuspended in resuspension solution from the provided Maxi- or Mini-prep kit. The Kit protocols were followed precisely. Whilst performing the Maxi-prep we also performed the optional DNA precipitation step to further concentrate our DNA to acquire high yields of DNA (~1000ng/ml). A nano drop was then used to measure DNA concentration.

2.4 Cell culture

2.4.1 Proliferation

C2C12 myoblasts and COS-7 fibroblasts were cultured in Dulbecco's modified Eagles medium (DMEM) (Thermofisher, Cat #41966-029), supplemented with 10 % foetal bovine serum (FBS) (Thermofisher, Cat #A4766) and 1% penicillin / streptomycin (Gibco, Cat #15140122) (full serum DMEM). Cells were left incubated at 37 °C and 5% CO₂.

Cells were split weekly before confluency was reached using 1% trypsin solution. Prior to splitting, cells were washed with pre-warmed PBS. Trypsin was added and left for 10 minutes in the incubator. Post incubation most cells would be detached, trypsin would be neutralised through full serum addition. Cells were then split to ~20% of their original confluency.

2.4.2 Freezing

Cells were routinely frozen for cell preservation. Cells would be trypsinised as described above, then removed, placed in a 15ml falcon tube and centrifuged at 1200rpm for 12 minutes. The supernatant would be removed, and the resulting pellet would be resuspended in either 1ml (T25 flask) or 3mls (T75 flask) of freezing media (Reagents Table), depending on the flask size. This would be immediately aliquoted into 1ml cryotubes and transferred into the -80°C freezer or liquid nitrogen.

2.4.3 Transfections

C2C12 myoblasts and COS-7 fibroblasts were cultured on glass slides within 24 well plates, myoblasts where seeded at 70-80% confluency before transfections. In order to transfect the *ky* constructs we used Genjet Invitro DNA transfection reagent

(SignaGen, Cat #SL100488). Following the protocol provided for 24 well plates and the appropriate cell line, cells were transfected. The transfection reagent was removed 12~18 hours post-transfection, being replaced with fresh full serum DMEM. Transfection efficiency was then checked 48 hours post transfection using a fluorescent microscope.

2.4.4 Fixation and Immunofluorescence.

Immunofluorescence analysis on C2C12 and COS-7 cells was carried out. Cells were fixed and mounted from 24 well plates onto glass slides before analysis.

Cells were washed twice with phosphate buffered saline (PBS) (Thermofisher, Cat #BR0014), then fixed using 4% paraformaldehyde (PFA) (Sigma, Cat #P6148) diluted in PBS for 15minutes. PFA was then removed and the cells were washed twice with PBS and once with dH₂O. Cells were mounted on slides, with mowiol (Fluka, Cat #81381) and 4',6-diamidino 2-phenylindole (DAPI) (Sigma, Cat #D9542) at 1:1000 dilution to stain the nuclei.

Cellular imaging took place on a Leica epifluorescent inverted laboratory LED microscope (Leica, Cat #DM IL LED)

2.5 In-Vivo Experiments.

2.5.1 In vivo electroporation

An hour before the electroporation procedure mice were injected with 50µl of hyaluronidase (Sigma, Cat #SLBX4632) (0.4U/µl diluted in 0.9% saline), alongside being given pre-operative analgesia in this case Bupavet (50µg/ml), 2 µl of Bupavet was administered per gram of mouse weight.

Mice were placed in an anaesthetising box with 4% isoflurane diluted in O₂, mice were left until deeply anaesthetised. Mice were transferred to a continually anaesthetised rodent face mask and a toe pinch reflex was used to insure deep anaesthetic. Throughout the procedure mice were kept on a heating pad maintained at 37°C. In

order to expose the tibialis anterior (TA) a subcutaneous incision was made. DNA constructs with a concentration above 1000ng/ μ l were injected directly into the TA (Figure 5.A), then three mm electrodes were placed within the muscles (Figure 5.B).

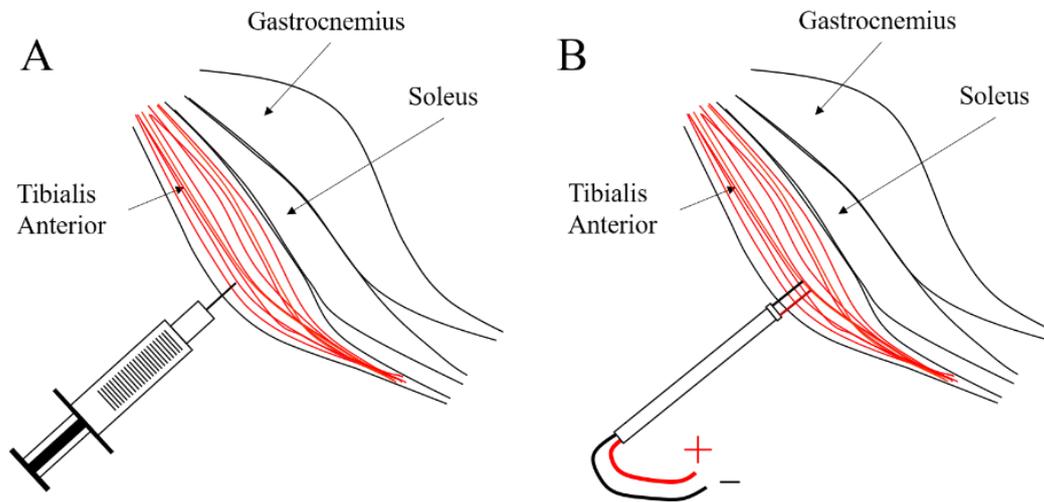


Figure 5. Mouse electroporation on a wildtype mouse. A) DNA microinjection. 20ul of a specified DNA construct would be injected directly into the exposed TA muscle. **B) Electroporation of the TA.** A current would be run through the TA using an electrode, allowing for an uptake of the DNA construct.

Pulses were delivered using a NEPA21 machine (Nepagene, Japan). Initially three 50-msec-long pulses at 80V followed by three more pulses of the opposite polarity were administered at each injection site at a rate of one pulse per sec. Mice were sacrificed eight days post electroporation.

Mice were sacrificed by a schedule one kill using cervical dislocation or rising CO₂. The TA/EDL of the mice was dissected. Muscles which had undergone electroporation would be fixed in 4% PFA for 15 minutes, then transferred to PBS. The whole muscle would be placed under a fluorescent microscope to check for a successful electroporation. If successful we would clearly see transfected fibers (Figure 6).

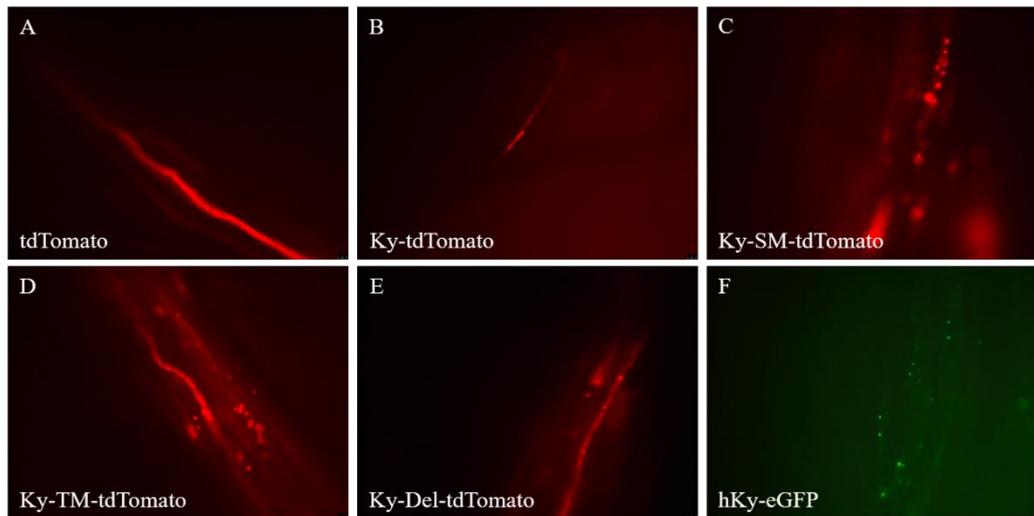


Figure 6. Whole Mount Images of successfully electroporated muscle fibres. Whole mount images where taken at a 4x magnification. The bottom left-hand corner of each panel shows the construct used. **A) Electroporated tdTomato. B-E) Electroporated KY constructs.** Each individual KY construct was tagged with tdTomato giving the red fluorescence. **F) Electroporated hKy-eGFP.**

Successfully electroporated muscles would then be snap frozen in liquid nitrogen cooled isopentane (Honeywell, Cat #M32631), then stored in the -80°C . Muscles which were to undergo histological analysis did not require fixation by PFA, instead they were instantly snap frozen in liquid nitrogen cooled isopentane.

2.5.2 Animal research ethics

All animal procedures were carried with approval from the University of York Ethics committee and followed the UK Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012. All procedures were performed by Gonzalo Blanco (Project licence number: P3FF2D943) within an approved establishment (Licence: X323ED8C8).

2.5.3 Muscle sectioning

Frozen muscles were removed from the -80 and kept on dry ice before sectioning with a cryostat (Leica, Cat # CM1520). Muscles were either sectioned cross sectionally or longitudinally depending on what we were looking for. Sections would typically be

around 12-14 μ m in width, being condensed on the superfrost adhesion slides (ThermoFisher, Cat #J1820AMNT). Sections were then placed in the -80 for storage, before mounting and staining.

2.5.4 Immunofluorescence of mouse muscle sections

In order to identify the Z-disk within the longitudinal sections, we needed to stain the α -actinin. Sections were thawed from the -80 freezer, then rehydrated with PBS. Sections were then permeabilised with 0.1% triton diluted in PBS for 5 minutes, sections were then washed twice with PBS. Post washing sections were blocked in 4% BSA (Diluted in PBS). Again, two washes were performed with PBS. Next the primary antibody EA53 (Abcam, Cat # ab9465) was added at a 1:300 dilution in PBS and left to incubate at 4°C overnight. Three 5-minute washes in PBS were undertaken before secondary antibody incubation. We used two secondary antibodies; FITC and TRITC (Reagents Table), both were left to incubate for two 25minutes at room temperature, as per the manufacturer's instructions. Finally, two five-minute PBS Washes and a ddH₂O washes were undertaken before mounting with DAPI diluted in mowiol.

In order to be able to identify individual fibers within the muscle cross-sections, we stained with Wheat Germ Agglutinin (WGA), which binds sialic acid and *N*-acetylglucosaminyl residues, commonly used for imaging of plasma membranes. The same method of thawing and rehydration was used before incubating sections for 20 minutes in WGA (1.400 diluted in PBS). Post WGA staining slides were washed 3 times with PBS and once with ddH₂O. Again, slides were mounted with DAPI diluted in mowiol.

All slides were imaged wither using a Leica epifluorescent inverted laboratory LED microscope (Leica, Cat #DM IL LED) or a Nikon Microphot-FXA.

2.6 Histology

Cross sectional sections of the TA were thawed from the -80 freezer, 15 minutes before Hematoxylin and Eosin H&E treatment. Slides were fixed in acetone for 10s before being transferred into Hematoxylin (Sigma, Cat# 290629) for 2 minutes. After Hematoxylin treatment the slides were washed for 1 minute with running water. The slides were then transferred to Scotts water for 1 minute, and immediately after they were washed with running water for 1 minute. After washing the slides were transferred into Eosin (Sigma, Cat# 861006) for 45 seconds. Immediately they were placed in 70% ethanol for one minute, then 100% ethanol for another minute. Finally, the slides were placed in HistoClear for a further minute, before mounting. Slides were then mounted with

Histological images were taken with a Lecia optical microscope (Lecia, Cat #DM2500), using the Insight 2Mp Monochrome FireWire Digital camera to capture images. This camera uses the SPOT basic software.

2.7 Enzymatic Tests

2.7.1 Transglutaminase Enzymatic test

The SIGMA-ALDRICH transglutaminase assay kit (Sigma, Cat #CS1070) was used to test for transglutaminase activity. The protocol was followed precisely when performing this test. We used a 1.2ug sample of the full-length protein due to limitations on the amount of protein. With the small domain we used a sample of 10ug, the maximum recommended amount of protein. Appropriate blanks were used, and the positive control derived from a guinea pig liver (provided by kit), was diluted in the appropriate buffer as per the manufacturer's instructions. A Dynex technologies MRX II Plate reader was used to measure the absorbance at 450nm.

2.7.2 Protease Enzymatic test

The ThermoFisher Scientific Pierce Protease Assay Kit (ThermoFisher, Cat#23263) was used to test for protease activity. Again, the protocol was followed precisely. The

full-length KY was tested at 50ug/ml, with an appropriate blank and positive control at the same concentration. The small domain KY and MBP control were tested at 500ug/ml, again with an appropriate blank and positive control at the same concentration. The positive control TPCK trypsin was provided by the kit. A Dynex technologies MRX II Plate reader was used to measure the absorbance at 450nm.

2.8 KY Protein Purification

All protein production was performed by the University of York Protein Production laboratory, with special thanks to Jared Cartwright and Rebecca Preece for their ongoing support and hard work.

2.8.1 Sample Prep

An o/n culture from transformed colonies was grown in 500ml of auto induction medium + Kanamycin (30ug/ml) at 37°C for 2hrs the 20°C in a shaker. The resulting solution was then pelleted down at 5,000rpm for 10mins. Protein pellets were resuspended in 100mls of extraction buffer which included a protease inhibitor cocktail. The suspension was sonicated five times for 30 seconds, being put on ice between pulses. Samples of 20ul were taken for use as a total gel sample. Cell debris was removed by centrifugation at 17,000g for 10mins at 4°C before 20ul being taken for a soluble gel sample.

2.8.2 Column Prep

Purification columns were prepped by adding 1.2mls (for full length KY) or 5ml (for MBP-KY(aa160-285)) of nickel beads to a tube. Beads were washed with 7 CV dH₂O whilst not disturbing the beads. 5CV His Buffer A was added to equilibrate the beads and allowed to settle, carefully the liquid was removed. Again, this procedure was repeated 7 times. The beads were then transferred to two 50ml falcon tubes, His buffer A was utilised to aid the transfer to the falcon tubes, any excess His buffer A was removed after the transfer. 100mls of lysate was then added to the two falcon tubes of beads, 50mls in each. The falcon tubes were then incubated for two hours at 4°C with mixing. Falcon tubes were then Centrifuged at 1000g for 1 minute at 4°C. Excess

liquid was then removed, leaving the beads and approximately 2mls excess liquid in each tube. At this point 20ul was taken to be used as the flowthrough gel sample.

A 1.2ml (for full length KY) or 5ml (for MBP-KY(aa160-285)) column was then packed with the remaining 2ml of flow through and the beads. By gravity flow with the flow dripped through, the remaining beads were washed with His Buffer A. Gradient elution was performed on a BioRad NGC purification system (BioRad). The gradient elution of 0-100% with His buffer B occurred over 10 column volumes where 1ml fractions were collected. UV read outs for each fraction were obtained and to confirm the presence of the protein, SDS-PAGE gels were run. Anti His blots were also performed for the full-length KY protein.

2.8.3 Analytical Size Exclusion

Analytical size exclusion was performed on the full-length KY protein. 100ul from the peak fraction of the His purification was loaded onto the AKTA 10 purification system using a 500ul superloop. A superdex 200 increases 10/300 GL column was used. An SDS-PAGE gel was run to confirm which fractions the KY protein was present in.

2.8.4 Size Exclusion

Size exclusion was used on the small MBP domain. The peak fractions were concentrated from the His Purification to approximately 4mls, using a 50ml falcon spin concentrator. This was then loaded onto the BioRad NGC purification system (BioRad, Cat #6286) using a 10ml superloop. A superdex 200 pg 16/60 size exclusion column was used. SDS-PAGE gels were run allowing identification of the fraction in which the protein is contained within.

Table 1 . Buffers used throughout protein purification and analysis. Left column shows the buffers, whilst the right shows their composition.

Buffer	Composition
Extraction Buffer	100mM Tris-HCl pH 8.0, 0.1% Tween20
His Buffer A (Full length protein)	50mM Tris-HCl pH 8.0, 300mM NaCl, 30mM Imidazole
His Buffer B (Full length protein)	50mM Tris-HCl pH 8.0, 300mM NaCl, 500mM Imidazole
His Buffer A (Small domain)	20mM Sodium Phosphate pH 7.4, 300mM NaCl, 30mM Imidazole
His Buffer B (Small domain)	20mM Sodium Phosphate pH 7.4, 300mM NaCl, 500mM Imidazole.
Size Exclusion Buffer (Full length protein)	50mM Tris-HCl pH 8.0, 300mM NaCl.
Size Exclusion Buffer (Domain)	20mM Sodium Phosphate, 300mM NaCl

2.9 Statistical analysis and Figure preparation.

2.9.1 Quantification of C2C12 speckled phenotype

A total of four slides were transfected providing us with 4 biological replicates for each construct. Each slide produced a large number of images, in order to count the cells quickly and accurately a specific method was used. Images were converted to 8 bits, then pushed to the threshold using the (Figure 7. A & B). The particle analyser was then used to count the number of cells. In order to count the speckled phenotype,

the contrast and brightness of the image was lowered drastically to a set point. At this point it clearly allowed us to identify the cells showing large aggregates (Figure 7.C).

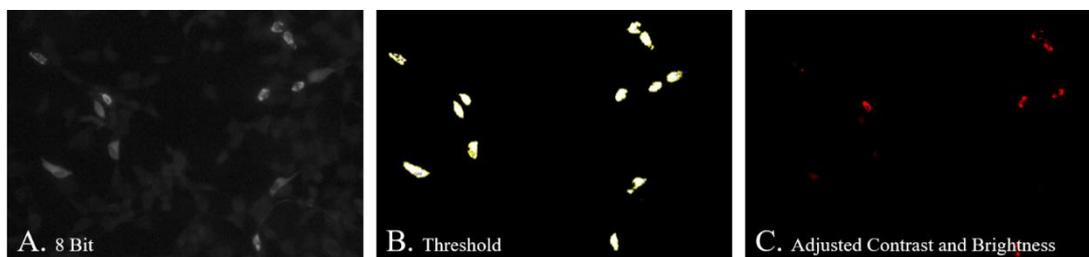


Figure 7. Counting of cells and cells showing a speckled phenotype. A) Conversion of images to 8 Bit. Images were converted to 8 bits. B) Images pushed to a threshold. Images then underwent the particle analyzer which had set parameters. C) Adjusted contrast and Brightness highlighting aggregates. After adjusting the contrast and brightness the aggregates were clearly visible, due to a brighter presence on the image.

Data was then compiled in an excel file, for each image the number of cells showing a phenotype was recorded, alongside the total number of cells for that image. Using this data, a percentage of cells showing the phenotype was calculated using the equation below.

$$\text{Percentage of cells with a phenotype} = \frac{\text{Nb of cells with a phenotype}}{\text{Total Nb of cells}} \times 100$$

The data gathered from each biological replicate was then compiled for each construct. A shapiro test was used to test normality, then a one-way ANOVA was used to compare the means.

2.9.3 Statistical tests performed on Enzymatic Assays

Both the Transglutaminase and Protease assays measured the end result as an absorbance at 450nm, the readout was proportional to the amount of enzymatic activity. Thus, in both cases three repeats for each test sample would be obtained. Overall, these assays were both run a minimum of three times, each revealing similar end results. Normal distribution was assumed due to the low replicate number. A one-way ANOVA allowed us to make valid comparisons between the means of each data

set. A post hoc Tukey's Test was also run for both enzymatic assays, backing up the findings of the ANOVA.

2.9.4 Quantification of colocalization

Again, four slides were co-transfected with the relevant constructs. This provided us with 4 biological replicates for each construct. Each slide produced ~5 images with co-transfected cells, due to the lower level of co-transfection seen.

Fiji was used to measure the level of co-localisation between the two proteins. Firstly, merged images were split into their individual channels, blue (DAPI), green, and red. After the channels were split into 8-bit images (Figure 8.A), the green channel was pushed to the threshold (Figure 8.B), and the cells expressing both constructs were selected, and stored in the ROI. After selecting the relevant cells, they were checked against the original image, ensuring all the co-transfected cells were selected. We performed the co-localisation test using the Coloc-2 plugin for Fiji, specifically using Costes threshold regression method (Costes, *et.al.*, 2004) was used. The region of interest for the test was specified through stored regions on the ROI manager, and the data was then collected as a spearman's rank correlation, the plugin also produced an image highlighting area of colocalization (Figure 8.C).

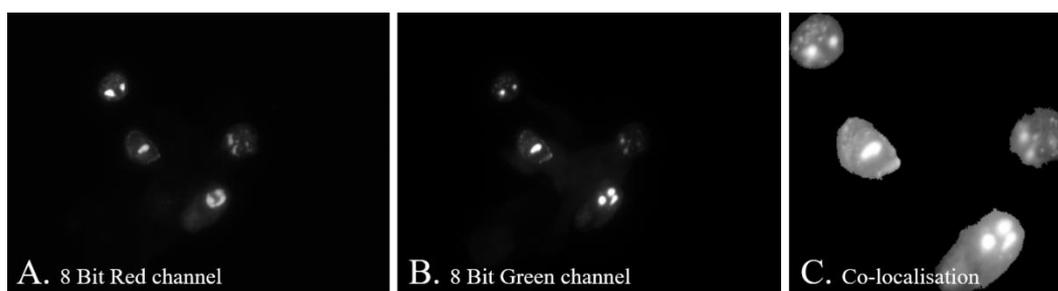


Figure 8. Co-localisation measurements between two proteins. A) Conversion of red channel images to 8 Bit. These images were subsequently pushed to the threshold to specify areas for the coloc2 test to run, these were stored within the ROI manager. B) Conversion of green channel images to 8 Bit. C) Resulting output after coloc-2 test. The areas highlighted in white show areas of high Co-localisation.

The collected data was then compiled into excel. In all cases the data was merged. A one-way ANOVA was run comparing the means of each construct. This data was then

represented as a box plot, allowing visualisation of outliers and how the data is dispersed.

2.9.5 Complementation quantification.

Successfully electroporated muscle fibers were quantified using Fiji (Schindelin, *et.al.*, 2012). Cross sectional muscle fiber sections were either stained with WGA or were overexposed in a FITC filter, allowing visualisation of muscle fibers membranes.

Images were then taken, and fiber size was measured using image J. Firstly the scale bar was set from the corresponding image. All transfected fibers were drawn around and stored in the region of interest (ROI) manger (Figure 9.A). Then using the ROI manager, each fiber's cross-sectional area was measured, and used to create an average fiber size for transfected fibers. Secondly the same procedure was used to measure the Untransfected fibers giving us an average size of Untransfected fibers (Figure 9.B).

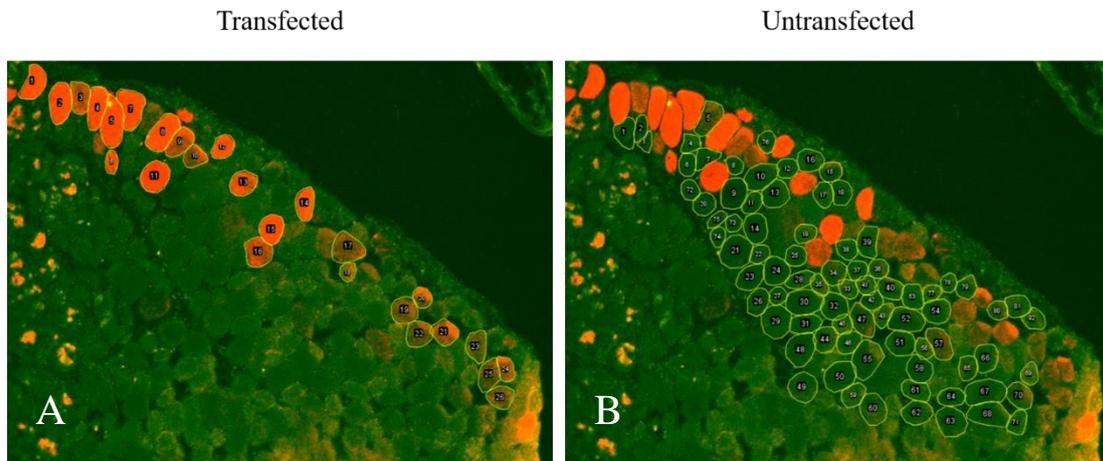


Figure 9. Quantification of cross-sectional area for transfected and untransfected fibers. **A) Measurements taken from transfected fibers.** Clearly transfected fibers were drawn around and stored in the ROI manger, each having its unique identification number. **B) Measurements taken from untransfected fibers.** Measurements would be taken from a large surrounding area of the muscle, in an area which looked to be of the same part of the muscle.

The data gathered was then used to create a ratio, which was used to compare transfected fiber size to untransfected fiber size. A ratio was used as it gave us a simple

way of seeing if a construct was increasing or decreasing the fiber size in the *ky/ky* mouse. If the Ratio > 1 the fiber size has increased. If the Ratio < 1 it has decreased.

$$\text{Ratio} = \frac{\text{Transfected Fibre Average}}{\text{Untransfected Fibre Average}}$$

The Ratio also allows for a valid comparison between mice as it normalises the untransfected fiber size for each mouse leg to 1. Therefore, we are able to perform statistics on multiple repeats. For each construct we had at least 3 replicates. In this case a one-way ANOVA, normal distribution was assumed due to the small replicate size.

2.10 In-Silico Analysis

2.10.1 Sequencing and Alignment

BLAST (Altschul, *et.al.*, 1990) was used to locate the hKY variants. The BLASTx tool takes a nucleotide sequence and compiles an output of highly similar sequences. Using the mouse *Ky* nucleotide sequence taken from ENSEMBL *Ky*-201 (ENSMUST00000039390.5)), a BLASTx was run, specified to look at sequences within Homo sapiens (Taxid: 9609). This allowed us to identify Homo sapiens *ky* variants.

BLAST was also used to compile a collection of *Ky* homologues from a vast repertoire of species. The *Ky* DNA sequences were acquired using BLASTx. Retrieved sequences were subsequently imported into the Molecular Evolutionary Genetic Analysis (MEGA-X) software (Kumar, *et.al.*, 2018). MEGA-X was used to convert these DNA sequences into protein sequences using the Standard genetic code. These protein sequences were aligned in MEGA-X using ClustalIW (Higgins, *et.al.*, 1994).

2.10.2 Phylogenetic Analysis

Phylogenetic analysis was performed using MEGA-X. Within all analysis the Jones, Taylor, and Thornton (JTT) model of evolution was selected (Jones, *et.al.*, 1992). Using the aligned *Ky* homologues, a neighbour joining tree was produced. This was produced using the PHYLOGENY tab, where the construct/test Neighbour joining tree option was selected, the parameters used are shown below (Figure 10.A). The produced tree was then edited and modified within MEGA-X.

MEGA-X was also used to estimate the rate of evolution at each protein site. Using the RATES tab within MEGA-X, the Estimate Position-by-Position Rates (Maximum likelihood) option was run using aligned KY protein homologues (Figure 10.B). The data output for the rate of evolution at each site was produced, this data was then collated in excel. The relative rate at each nucleotide site was then used to produce graphs and figures.

Phylogeny Reconstruction		Estimate Rate at Each Site (ML)	
Option	Setting	Option	Setting
A			
ANALYSIS			
Statistical Method →	Maximum Likelihood		
PHYLOGENY TEST			
Test of Phylogeny →	None		
No. of Bootstrap Replications →	Not Applicable		
SUBSTITUTION MODEL			
Substitutions Type →	Amino acid		
Genetic Code Table →	Standard		
Model/Method →	Jones-Taylor-Thornton (JTT) model		
RATES AND PATTERNS			
Rates among Sites →	Uniform Rates		
No of Discrete Gamma Categories →	Not Applicable		
DATA SUBSET TO USE			
Gaps/Missing Data Treatment →	Use all sites		
Site Coverage Cutoff (%) →	Not Applicable		
Select Codon Positions →	Not Applicable		
TREE INFERENCE OPTIONS			
ML Heuristic Method →	Nearest-Neighbor-Interchange (NNI)		
Initial Tree for ML →	Make initial tree automatically (Default - NJ/BioNJ)		
Initial Tree File →	Not Applicable		
Branch Swap Filter →	None		
SYSTEM RESOURCE USAGE			
Number of Threads →	7		
B			
ANALYSIS			
Tree to Use →	Automatic (Neighbor-joining tree)		
User Tree File →	Not Applicable		
Statistical Method →	Maximum Likelihood		
SUBSTITUTION MODEL			
Substitutions Type →	Amino acid		
Genetic Code Table →	Standard		
Model/Method →	Jones-Taylor-Thornton (JTT) model		
RATES AND PATTERNS			
Rates among Sites →	Gamma Distributed (G)		
No of Discrete Gamma Categories →	5		
DATA SUBSET TO USE			
Gaps/Missing Data Treatment →	Use all sites		
Site Coverage Cutoff (%) →	Not Applicable		
Select Codon Positions →	Not Applicable		
Branch Swap Filter →	None		

Figure 10. The parameters used within MEGA-X statistical tests. Both tests used the same substitution model. A) Parameters of Phylogeny Reconstruction. Selected from the PHYLOGENY tab B) Parameters of Estimate Rate at Each Site (Maximum Likelihood). Selected from the RATES tab.

Chapter 3: *In-vitro*
analysis of the
mouse KY

3.1 Overexpression of *Ky* variants within C2C12 mouse myoblasts

The current evidence suggests that the conserved TGN/PROT domain plays a significant role within KY's mechanistic action. To further investigate this, multiple mutagenised *Ky* constructs were utilised, previously produced by George Atkinson during his master's project. These constructs were originally produced from pDEST47-*Ky*-tdTomato (*Ky*-td), the full-length mouse KY tagged with tdTomato. These constructs each contain various mutations within TGN/PROT domain. These constructs are as followed: pDEST47-*Ky*-SM-tdtomato (*Ky*-SM) which holds a single Cystine to alanine (C[225]A) substitution within the first conserved residue of the catalytic triad. pDEST47-*Ky*-TM-tdtomato (*Ky*-TM) where all three conserved residues are substituted with alanine (C[225]A, H[267]A, and D[282]A). Finally, pDEST47-*Ky*-DEL-tdtomato (*Ky*-DEL) which had the whole TGN/PROT domain removed (DEL[217-285]) (Figure 11).

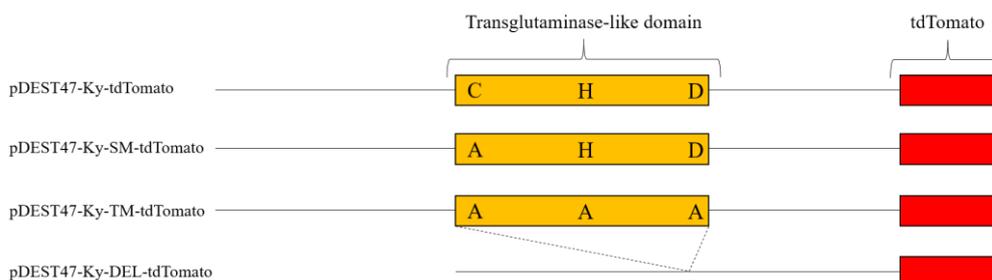


Figure 11. Mutagenised *KY* Constructs. pDEST47-*Ky*-tdTomato encodes the wild type full length mouse *KY* protein, the conserved residues of the catalytic triad within the TGN/PROT domain is shown (orange box), while tdTomato (red box) is attached via the C-terminus. pDEST47-*Ky*-SM-tdTomato holds a single (C[225]A) substitution in the catalytic triad of the conserved TGN/PROT domain. pDEST47-*Ky*-TM-tdTomato holds a triple substitution in all three conserved residues of the catalytic triad (C[225]A, H[267]A, and D[282]A). pDEST47-*Ky*-DEL-tdTomato has the whole conserved TGN/PROT domain removed.

The mutations and deletions were mapped using ENSEMBL *Ky*-201 (ENSMUST00000039390.5). An empty vector encoding tdTomato was used as control. The three catalytic residues held within the TGN/PROT domain, are known to be critical for enzymatic ability. The *Ky*-SM and *Ky*-TM which have these residues substituted with alanine will likely show no enzymatic ability. Previous work has

shown replacement of these residues with alanine greatly reduces enzymatic activity, shown by mutated catalytic residues within other TGN/PROT domain from similar proteins (Carter and Wells, 1988). Therefore, these constructs are predicted to have no enzymatic activity, if indeed KY is enzymatically active.

It has been predicted that a large number of eukaryotic proteins with transglutaminase domains have lost their enzymatic ability. These regions have likely evolved to mediate protein-protein interactions (Anantharaman, *et.al.*, 2001), which may explain how KY interacts with IGFN1 and Filamin C in the IGFN1-FLNC-Ky protein complex (Baker, *et.al.*, 2010). To investigate the functions of the TGN/PROT domain, C2C12 proliferating myoblasts were transfected with the aforementioned KY TGN/PROT domain mutants and assessed for protein localisation and cellular morphology when overexpressed. GenJet invitro transfection reagent was used to transfect the C2C12 myoblasts followed by incubation for 48 hours to allow each construct to express, before finally fixation and mounting with mowiol mounting agent and DAPI.

All KY constructs showed cytoplasmic expression, further observation revealed a high proportion of transfected myoblasts contained large cytoplasmic aggregates (Figure 12.B-E) when compared to the control, pDEST47-tdtomato (Figure 12.A). ImageJ was utilised to quantify these aggregates whereby a precise method was used to count the number of cells displaying these protein aggregates (Methods).

A one-way ANOVA showed each KY construct had a significantly increased amount of cytoplasmic aggregates compared to the tdTomato control. Further to this Ky-DEL showed a small significant increase in the percentage of aggregates seen when compared to the other KY constructs (Figure 12.F). No other morphological changes were observed within the transfected cells when compared to the control.

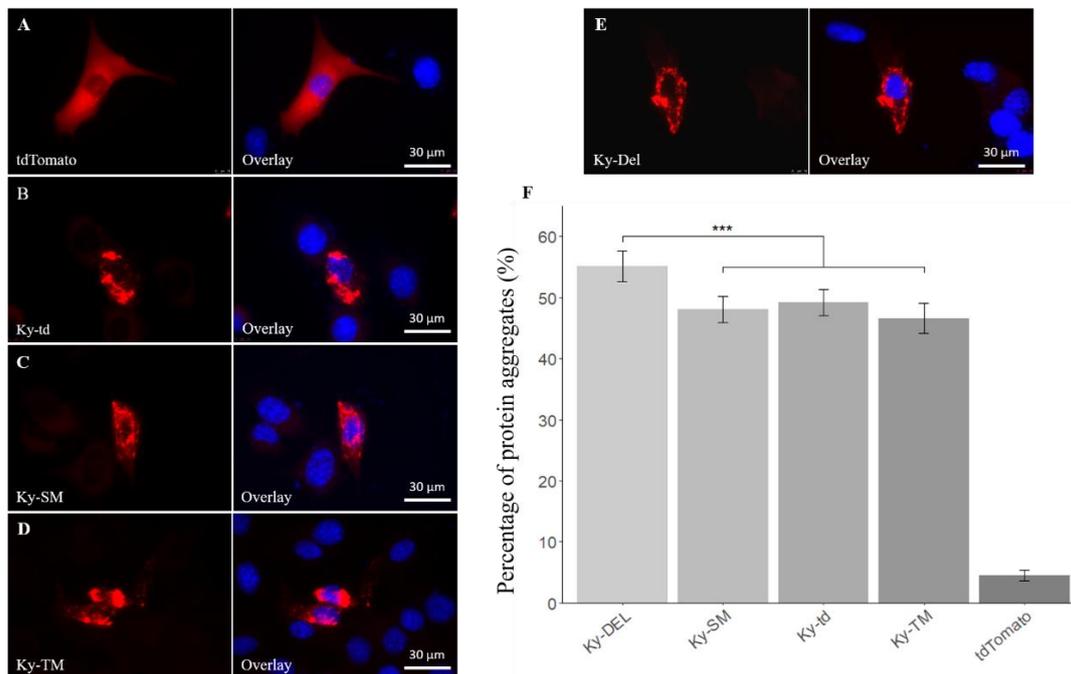


Figure 12. Overexpression of KY constructs within C2C12 myoblasts. A-E) Different Ky constructs expressed within C2C12 myoblasts. The Texas red channel reveals the Ky construct used. The Overlay with DAPI is shown on the right. A scale bar is shown in the bottom right-hand corner. F) **Quantification of phenotype.** A Bar graph is shown with different constructs. Error bars are shown, $P = \text{***} < 0.001$. ANOVA result: $F = 104.2$, $df = 4$, $p\text{-value} = 2 \times 10^{16}$.

3.2 Purification of Ky Protein constructs.

To understand the role of the conserved TGN/PROT domain, we needed to determine whether this domain is still enzymatically active. The conserved TGN/PROT domain still contains the catalytic residues Cystine, Histidine, and Aspartate which form the catalytic triad needed for enzymatic activity. This high level of conservation, seen specifically at the catalytic residue sites, suggest that KY may still hold enzymatic ability.

To test whether KY holds enzymatic abilities within the TGN/PROT domain, it was decided to purify different KY protein constructs. These constructs consisted of the full-length protein and a selection of smaller intermediates all of which contained the conserved TGN/PROT domain. Two constructs showed a good promise of being solubilised; a small MBP tagged version which ranged from amino acid (aa) 160-285

containing the TGN/PROT domain (MBP-KY(aa160-285)), and secondly a His tagged full-length Ky protein (KY) also showed a good level of solubilisation (Figure 13). In order to control for the MBP tag on the MBP-KY(aa160-285), purified MBP was also produced, for use as a control

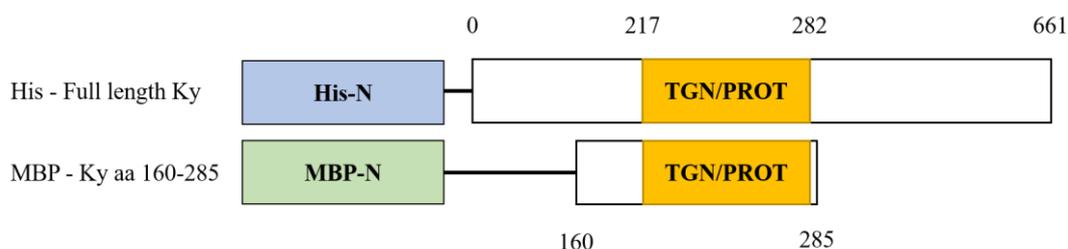


Figure 13. Schematic diagrams of purified KY constructs. Two constructs are shown the His tagged full-length KY and the MBP tagged small TGN/PROT domain. Both constructs were tagged via their N-terminus. Each protein tag has been highlighted, the TGN/PROT domain in both constructs is highlighted in orange. The numbers indicate the reflected protein residues, with the TGN/PROT domain held between residues 225-282 as indicated.

3.3 Transglutaminase Assay's on Purified KY protein.

Purified protein was tested with a transglutaminase assay kit. This kit tests the ability of proteins to catalyse the formation of a covalent bond between a free amine group of poly-L-lysine and the γ -carboxamide group of biotin-TVQQEL-OH. If the test protein holds any transglutaminase activity biotin is immobilized to the plate, then detected using Streptavidin-Peroxidase and 3,3',5,5'-Tetramethylbenzidine Liquid substrate. The amount of biotin adhered is detected using 450nm wavelength. The amount of biotin detected is then proportional to the enzymatic ability of the test protein.

The test specified 1-10 μ g of test protein was needed, each condition had 3 repeats. The yield of the purification of the MBP-KY(aa160-285) protein was very high, therefore 10 μ g of protein was used per repeat. The purification of the KY protein was not as efficient, therefore only 1 μ g of test protein could be used in each repeat. The MBP control was used at 10 μ g for each test. An active transglutaminase originating from a guinea pig liver was provided by the kit and used as a positive control. Each test was incubated at room temperature for 30 minutes as per manufacturer's instructions.

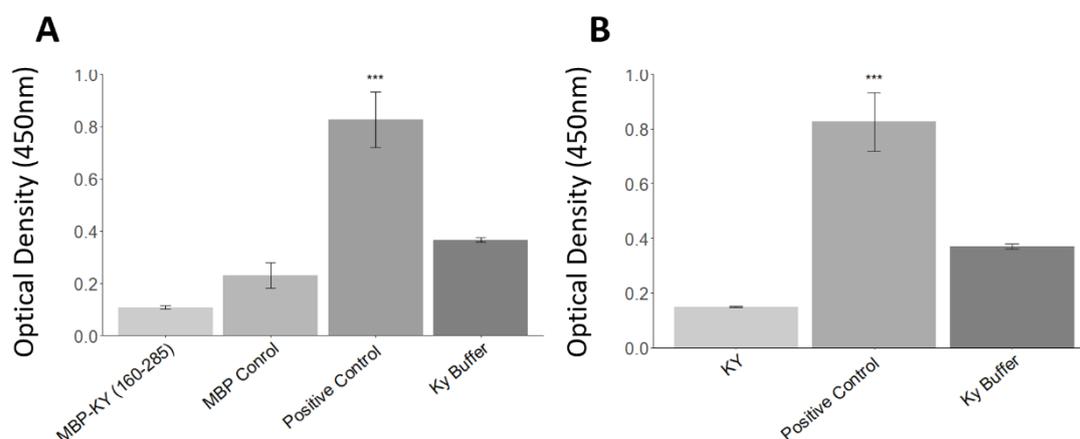


Figure 14. KY's TGN/PROT domain holds no transglutaminase enzymatic activity. In both cases enzymatic activity was measured throughout the detection of a substrate at 450nm. Error bars are shown alongside the significance difference, $P = \text{'***'} < 0.001$. **A) Transglutaminase assay on MBP-KYaa(160-285).** The different constructs tested are as shown. A MBP control was used due to the MBP tag include on the MBP-KYaa(160-285). ANOVA result: $F = 28.43$, $df = 3$, $p\text{-value} = 5.1 \times 10^{-3}$. **B) Transglutaminase assay on KY protein.** The different constructs tested are as shown. ANOVA result: $F = 31.02$, $df = 2$, $p\text{-value} = 6.86 \times 10^{-4}$.

Results from the transglutaminase activity test indicated that KY does not have any transglutaminase activity (Figure 14). The transglutaminase enzymatic activity of both KY and MBP-KY(aa160-285) was significantly lower than the positive control. Statistical analysis from a one-way ANOVA showed a significant difference between the positive control and each of the KY purified proteins. KY, MBP-KY, and the MBP control all showed significantly lower optical density of Biotin than their buffer controls.

3.4 Protease Assay's on Purified KY protein.

Again, purified protein was tested for protease enzymatic activity using the Pierce Protease Assay Kit. The test uses succinylated casein and trinitrobenzenesulfonic acid (TNBSA). Succinylated casein is native casein that has been treated with succinic anhydride to block primary amines on the surface of the protein. Proteases cleave the succinylated casein at peptide bonds, exposing the primary amines. TNBSA is then used to detect exposed primary amines, producing an orange colour.

For statistical significance, three repeats were used for each construct. Again, the maximum amount of protein was used if available. MBP-KY(aa160-285) was tested at 500 µg/ml, with the control and MBP buffer at the same concentration. Whilst KY was tested at 50 µg/ml with the positive control, a Trypsin standard provided by the kit, at the same concentration. In both cases the appropriate buffers were also tested. Each test was incubated at room temperature for 20 minutes as per manufacturer's instructions. Then Optical density, and therefore the level of orange colour produced in the reaction, was measured using the 450nm wavelength.

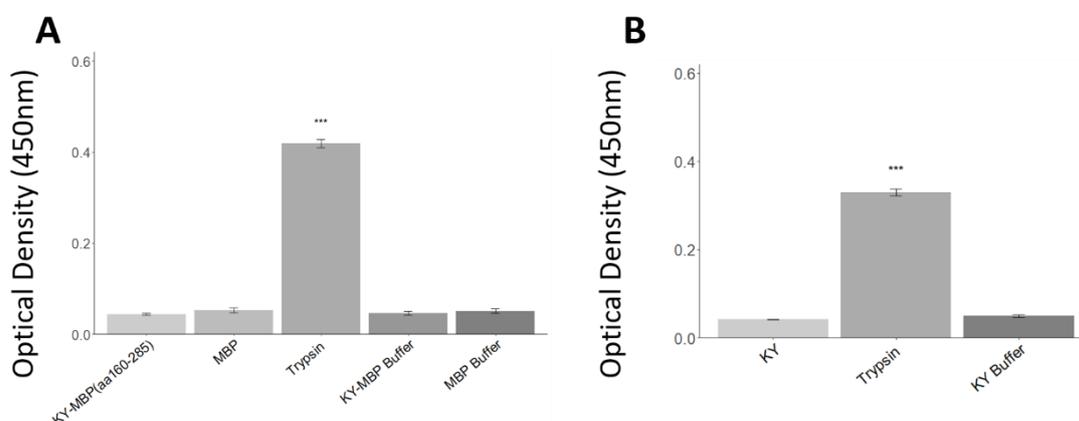


Figure 15. KY's TGN/PROT domain does not contain any protease activity. A) Protease assay for MBP-KY. The different tested constructs are shown. **B) Protease assay for purified KY protein.** In both cases enzymatic activity was measured throughout the detection of a substrate at 450nm. Error bars are shown alongside the significance difference, $P = \text{'***'} < 0.001$. **A) Protease assay for MBP-KY(aa160-285).** The different constructs tested are as shown. A MBP control was used due to the MBP tag include on the MBP-KY(aa160-285). ANOVA result: $F = 788.4$, $df = 4$, $p\text{-value} = 1.77 \times 10^{-13}$. **B) Protease assay for purified KY protein.** The different constructs tested are as shown. ANOVA result: $F = 1087$, $df = 2$, $p\text{-value} = 2.08 \times 10^{-8}$.

The results from the protease assay show that the Ky protein does not have any protease activity (Figure 15). The measured activity of MBP-KY(aa160-285) and the full KY protein were both significantly lower than the positive control. The little variance observed between both proteins provides further confidence in these results. Furthermore, this experiment was repeated twice with similar results.

Chapter 4: *In-vivo*
rescue tests with KY
TGN/PROT domain
variants

4.1 Histological analysis of the *Ky* deficiency seen within the C3 strain of mice.

The original homozygous *ky/ky* mouse arose due to a sporadic mutation within the BDL strain of mice first reported by Dickinson and Meikle (1973). This mutation led to the development of spinal curvature (kyphoscoliosis) leading to postural changes caused by postural muscle weakness. The *ky/ky* mice typically show pathology within slow contracting postural muscles such as the soleus. It has been shown that these muscles undergo a single postnatal phase of fiber degeneration and regeneration (Bridges, *et.al.*, 1992). Furthermore, *Ky* deficient mice show dramatic shifts in the type of MHC and MLC isoforms expressed within slow muscles, which leads to a complete shift to type I fibers within the soleus (Blanco, *et.al.*, 2001). Here it was suggested this shift in fibers type is an adaptive shift due to the low muscle to body weight ratio.

For this project, a new strain of *ky/ky* mice were derived from the C3H background of mice. These *ky/ky* mice may therefore be different to the original *ky/ky* mice. Early observations showed similarly overt features when compared to the original *ky/ky* BDL strain, these mice could be easily recognised post weaning. In most cases, mice were quickly identified due to their smaller size, and inability to perform a normal placing response test in which *ky/ky* mice were unable to reach towards a ledge, due to weakened paraspinal and postural muscles (Appendix). Interestingly within the C3H background the male *ky/ky* mice seemed to exhibit a more severe pathology. In the most severe cases mice would exhibit whole body tremors, a drastically smaller size, and exhibit an increased respiratory rate.

Histological analyses were performed to identify phenotypic differences between the BDL and C3H genetic backgrounds, and furthermore to investigate if in C3H background the males had a worst phenotype. The soleus was obtained from both male and female wildtype and *ky/ky* mice. The soleus was chosen as it is a postural slow contracting hind-limb muscle and is known to be severely pathological in the *ky/ky* mutant. Mice were taken at 8 weeks old; the TA and soleus were dissected which allowed easy access to obtain sections to observe the soleus. These sections were then stained with haematoxylin and eosin (H&E) (Methods).

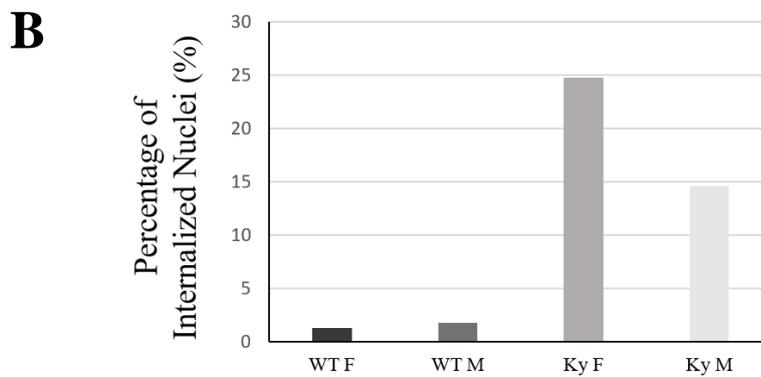
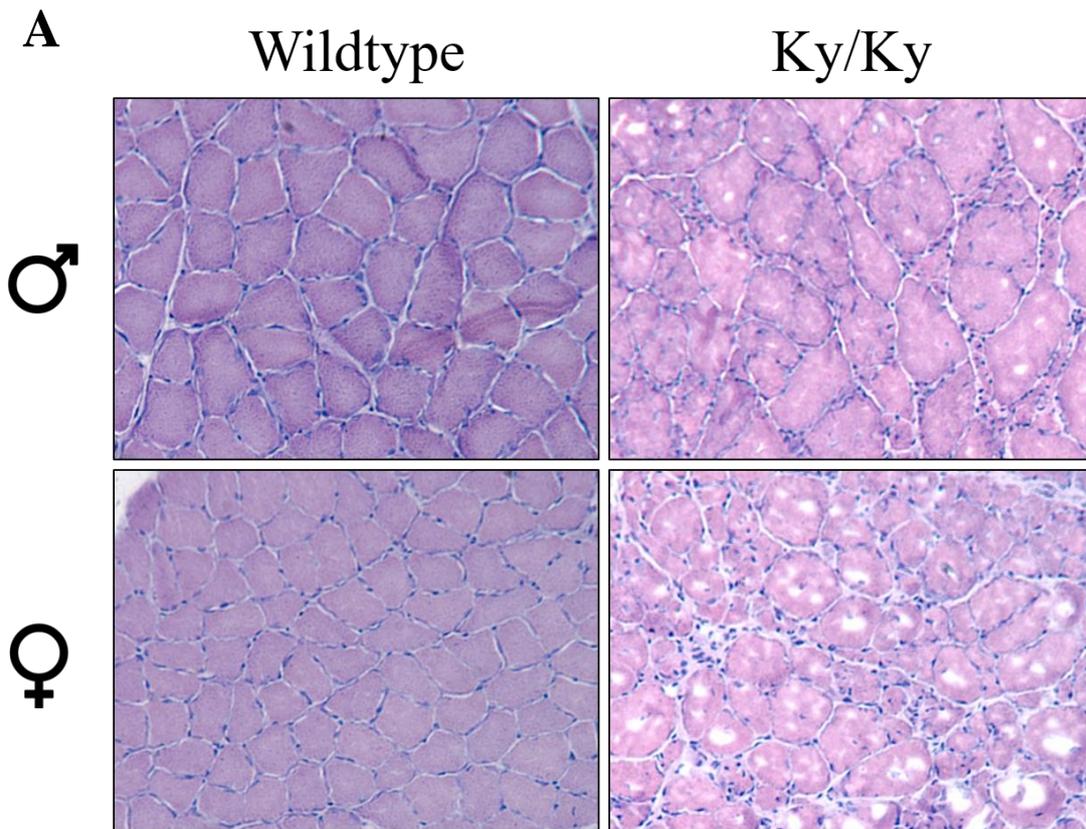


Figure 16. Histological analysis of the soleus of Wildtype and ky/ky mice. A) Haematoxylin and eosin-stained sections from wildtype and ky/ky mice soleus. The top row shows the soleus taken from wildtype and Ky mice. The bottom row shows the soleus taken from female wildtype and Ky mice. B) Quantification of the percentage of internalized nuclei from ky/ky mice and wildtype soleus. No statistical analysis was performed as these results were taken from 1 replicate in each sample.

Due to limited time only one replicate was taken for comparisons of WT and *ky/ky* male and female mice. Mice were taken at 8 weeks. As expected, H&E staining revealed a highly pathological soleus for both the female and male *ky/ky* mice (Figure 16.A). The first thing noticed was the extreme fiber size variability when compared to the ordered nature of the WT soleus. A WT soleus will typically have homogeneous fibers with consistent cross-sectional area. This was observed in both male and female WT soleus muscle sections. Conversely, the *ky/ky* mice show a vast heterogeneity of extremely small fibers alongside much large sized fibers. Unfortunately, due to limited repeats it was not possible to gather the data. But this phenotype is clearly seen within these fibers (Figure 16.A).

Next, we observed that a good proportion of the *ky/ky* fibers showed centralised nuclei, which is a typical hallmark of fiber regeneration. The presence of a high proportion of extremely small fibers further indicated that the muscle had been undergoing regeneration. To quantify these fibers undergoing regeneration, the number of fibers showing centralised nuclei were counted as a percentage of the total number of fibers in the muscle (Figure 16.B). Results showed both WT female and male mice had very few fibers showing centralised nuclei. When compared to WT mice, *ky/ky* mice had an increase in the number of centralised nuclei, which was observed to be higher in the female than the male.

Overall, there is a clear pathology within the soleus of this new strain of *ky/ky* mice when compared to wildtype mice. The pathology is highly reminiscent of the original pathology seen within the BDL strain of *ky/ky* mice.

4.2 In-vivo localisation of recombinant KY proteins.

KY is a known Z-disk associated protein, absence of this protein from the Z-disk is known to cause a multitude of abnormalities such as Z-disk thickening (Beatham, *et.al.*, 2004). The region within the KY protein which aids targeting to the Z-disk is yet to be elucidated. It has previously been theorised the TGN/PROT domain may aid targeting of KY to the Z-disk. Previously electroporated Ky-DEL in the tibialis anterior (TA) of a mouse showed a more diffuse pattern of expression when localising

to the Z-disk, when compared to the clear striations of Ky-td which localise clearly to the Z-disk (Unpublished work, Jokl, 2018). This data suggested that the conserved TGN/PROT domain may have a role in targeting KY to the Z-disk. Further being backed up by the evidence that many eukaryotic transglutaminases have likely evolved to serve another function (Anantharaman, *et.al.*, 2001).

In order to assess the ability of the TGN/PROT domain to target KY to the Z-disk, different KY constructs would be electroporated into mouse TA muscles, thus allowing visualisation of localisation. Using the mutagenised KY constructs (Figure 11) we tested each construct's ability to localise to the Z-disk. We theorised that were the TGN/PROT domain critical for Z-disk targeting, mutations within the region would result in less/absent localisation to the Z-disk, and the full deletion would likely potentially show no attachment to the Z-disk at all. The TA of wildtype mice was electroporated with these constructs, due to the fact it is easy to access for electroporation and dissection.

Mice had hyaluronidase injected into each leg which was to undergo electroporation, this aids the breakdown of the ECM, thus allowing for a higher electroporation efficiency (Cemazar, *et.al.*, 2012). Mice underwent the electroporation procedure through a surgical protocol allowing for an accurate and repeatable method of electroporation (Methods). Mice were then left for 8 days to express recombinant proteins. Mice were then sacrificed, and the TA was dissected.

Frozen muscles were sectioned longitudinally. Post sectioning, one slide was mounted with DAPI in order to identify regions of interest. It was clear that striations could be visualised for each construct except for tdTomato. Therefore, to confirm that these striations were visualised at the Z-disk, sections were stained with EA53 which is an anti- α -Actinin antibody. Sarcomeric α -actin is a vital component of the Z-disk and is therefore a good biomarker to identify the Z-disk. This would allow us to confirm that fiber striations seen with the KY constructs were indeed localised to the Z-disk, post staining the slides were mounted with DAPI and imaged.

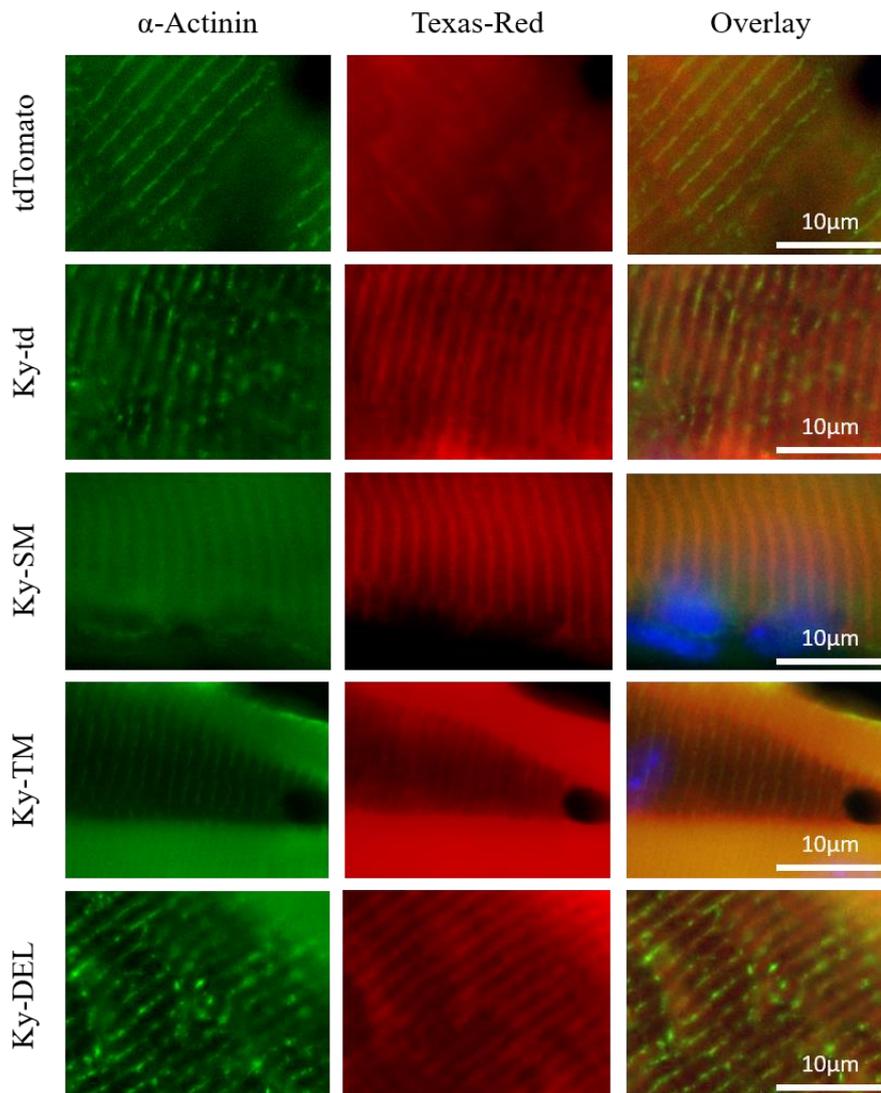


Figure 17. Z-disk localization of mutagenised KY constructs. Longitudinal sections of electroporated muscles stained. Columns from left to right: The first column is an α -Actinin stain highlighting the Z-disk. The Texas-red filter detects different KY construct used within each row. The Magnification column shows the area within the white box for each corresponding construct. The final column shows an Overlay of the α -actinin, Texas red, and DAPI. A scale bar is also shown in the overlay column. **Row A) Electroporated tdTomato.** tdTomato showed a diffuse expression profile along the fiber. **Row B) Electroporated Ky-td.** Ky-td shows extremely clear striations, which overlay with the α -actinin. **Row C) Electroporated Ky-SM.** Again, clear striations are seen with the fiber also lining up with the α -actinin. **Row D) Electroporated Ky-TM.** The striations within this construct are slightly harder to see but are still present. **Row E) Electroporated Ky-DEL.** Clear striations are seen throughout the fiber, lining up with the α -Actinin.

Results showed that each KY construct did in fact localise to the Z-disk (Figure 17). Compared to tdTomato each construct shows very clear striations, which co-localise with the α -actinin, showing Z-disk localisation. Interestingly the diffuse pattern observed previously with Ky-DEL was not observed here, whereby this construct produced a very clear striation pattern. This result suggests that the loss of the TGN/PROT domain does not seem to affect the localisation of KY to the Z-disk. All other KY constructs showed clear Z-disk localisation. Thus, it is unlikely the TGN/PROT domain is aiding Z-disk targeting.

4.3 In vivo complementation of a *ky/ky* deficiency

Unpublished work has suggested that the KY protein can induce a hypertrophic response to rescue atrophic muscle fibers within *ky/ky* muscles (Unpublished work, Jokl, 2018). Interestingly when Ky-td is overexpressed in wildtype muscles, no hypertrophic response is observed, suggesting KY protein amount does not correlate with the level of hypertrophy (Unpublished work, Jokl, 2018). Also showing that KY is not a regulator of hypertrophy but is necessary to maintain the muscle and initiate a hypertrophic response. Surgical overload within wildtype muscles causes an adaptive response whereby a change in the type of MHC expressed is observed, alongside muscle fiber hypertrophy. No hypertrophic response is observed in *ky/ky* mice when surgically overloaded (Blanco, *et.al.*, 2001). Hence, KY must have a significant role in inducing a hypertrophic response, indicating that it may be important for mechanosensing and the Z-disk.

The relevance of the conserved TGN/PROT domain still remains elusive, therefore we wanted to determine whether the TGN/PROT domain is important for fiber size rescue in *ky/ky* mice. To investigate this, the mutagenised TGN/PROT domain KY constructs previously used would be electroporated into the TA muscles of *ky/ky* mice, and evaluated for their ability to rescue normal fibre size.

Furthermore, this experiment can also be used to test by an independent method the importance of the conserved catalytic residues within the TGN/PROT domain. We have previously reported that mutations within the TGN/PROT domain are not

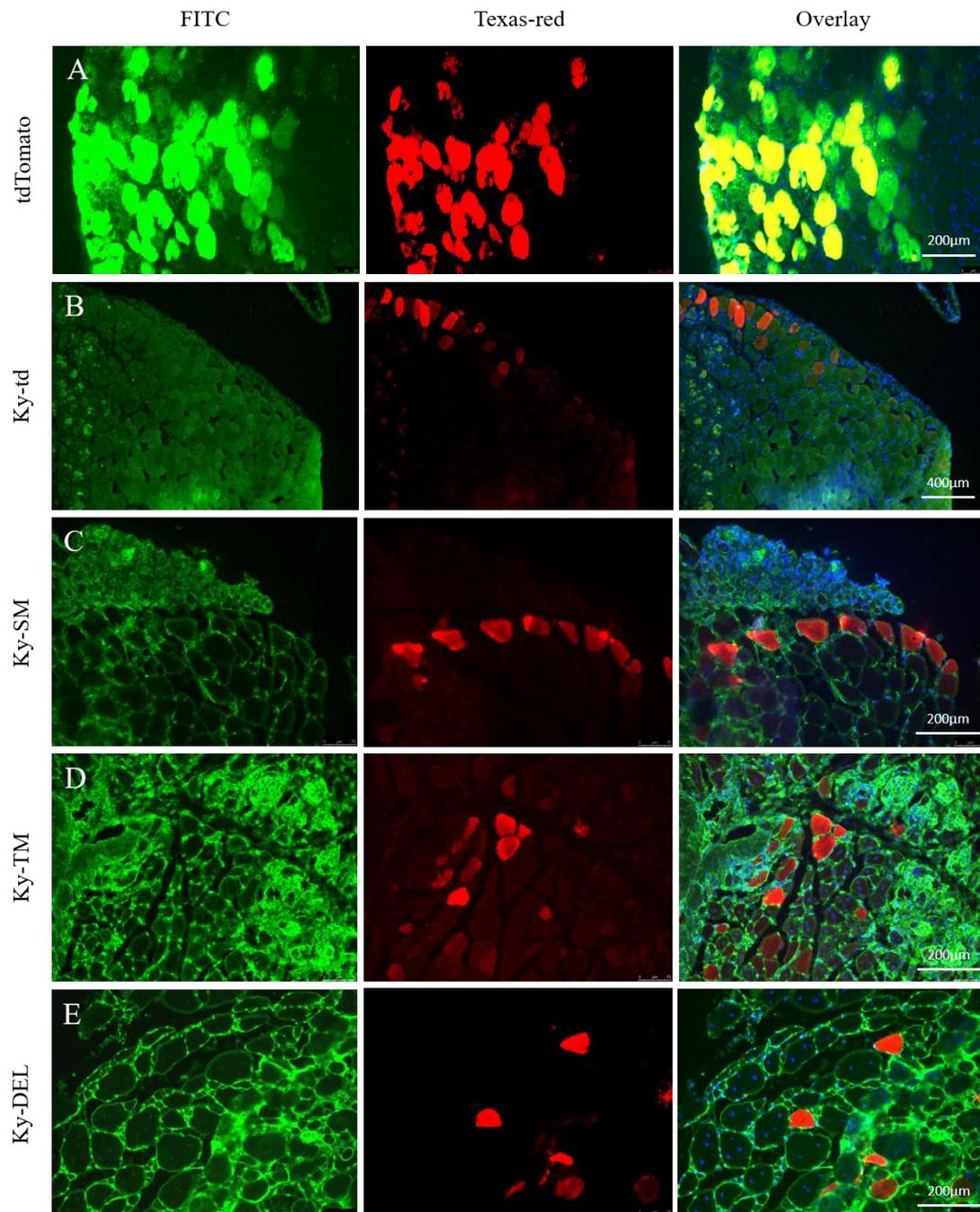
sufficient to alter Z-disk localization (Figure 17). Therefore, if we assume the mutations within these KY constructs are sufficient to remove or limit any enzymatic activity, we can utilise these to test if enzymatic ability is needed to induce a hypertrophic response. If a hypertrophic response is seen with enzymatically dead versions of Ky (Ky-SM & Ky-TM), we can conclude that the enzymatic ability and the catalytic triad is not needed to induce a hypertrophic response.

Again, we chose to electroporate the TA of *ky/ky* mice. The constructs, tdTomato, Ky-td, Ky-SM, Ky-TM, and Ky-DEL were electroporated in *ky/ky* mice using the same surgical procedure and pre-treatment with hyaluronidase. Mice were then left for 8 days allowing for construct expression and any subsequent phenotypic changes within the muscle to become apparent, before being euthanised. The TA muscles were dissected, frozen in liquid nitrogen-cooled isopentane. Cross sections of the TA muscle were obtained using a cryostat to allow for analysis of the fiber cross-sectional area. Finally, sections were imaged, and quantified (Methods).

The cross-sectional area of transfected and untransfected fibers was measured using ImageJ. The *ky/ky* phenotype shows vast fiber size variability, so comparisons between mice were impossible without normalisation. In order to make valid comparisons between different mice we opted to represent the difference in cross-sectional area between the transfected and untransfected fibers as a ratio (Methods). If the normalised value for the transfected fibers is above 1, then the average fiber size is greater than the untransfected. Conversely, if it is less than 1 the fiber size has reduced in comparison to the untransfected fibers.

In each case a minimum of three biological replicates were used. The tdTomato control showed a much higher electroporation efficiency than other constructs (Figure 18.A). All KY constructs showed a much lower electroporation efficiency when compared to eGFP, likely due to the mKY constructs' larger sizes (Figure 18.B-E). A one-way ANOVA was used on the ratio averages and showed a significantly higher cross-sectional areas for Ky-td, Ky-SM, and Ky-TM when compared to the control untransfected fibers, tdTomato and Ky-DEL. Neither Ky-DEL or tdTomato showed

any statistical difference when compared to the untransfected fibers indicating there was no change from the original size pre-electroporation. No Statistical difference was observed between Ky-td, Ky-SM, and Ky-TM. (Figure 18.F)



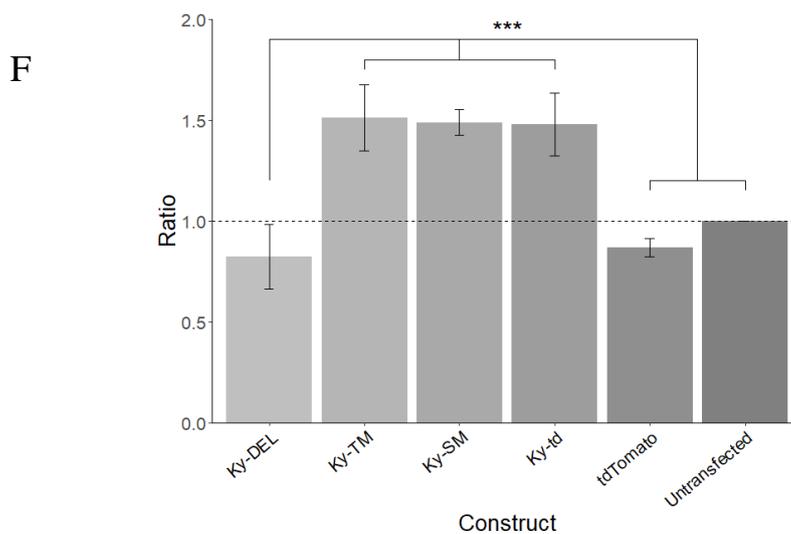


Figure 18. Enzymatically dead versions of Ky can induce a hypertrophic response within the muscles of Ky deficient mice. Each row shows a different Ky construct which has been electroporated into Ky/Ky mice. Columns from left to right: The FITC fluorescent filter was used to highlight cellular membranes. The Texas-red filter was then used to detect the different Ky constructs overexpressed in Ky/Ky mice. Finally, the last column is an overlay of the two filters including DAPI. A scale bar is shown at the end of each row and corresponds to each construct used A) *tdTomato* electroporated into Ky/Ky mice. Overexposure of the FITC filter allowed visualization of individual fibers. B) *Ky-td* electroporated into Ky/Ky mice. Again, the FITC filter was overexposed to outline of the fibers. C) *Ky-SM* electroporated into Ky/Ky mice. WGA was detected in the FITC filter to outline the cell membranes of individual fibers D) *Ky-TM* electroporated into Ky/Ky mice. WGA was detected in the FITC filter E) *Ky-DEL* electroporated into Ky/Ky mice. WGA was detected in the FITC filter. F) **Quantification of electroporated fiber size compared to untransfected Ky/Ky fibers.** A ratio was developed to represent a change in fiber size. If the ratio is higher than 1 the fiber has grown. A Box plot represents each Ky construct, compared to untransfected fibers. Data underwent a one-way ANOVA allowing valid comparisons of the ratios. Neither *Ky-DEL* or *tdTomato* are significantly different from the untransfected fibers. **ANOVA result:** $F = 6.977$, $df = 5$, $p\text{-value} = 1.03 \times 10^{-3}$.

These results indicate that the mutations of the conserved catalytic residues (Cystine, Histidine, and Aspartate) do not affect the ability of KY to induce hypertrophy in *ky/ky* muscle fibers. Interestingly, the removal of the whole conserved TGN/PROT domain does have a significant effect on the ability to induce hypertrophy, as the *Ky-DEL* construct is unable to induce the same hypertrophic response as the WT, *Ky-SM*, and *Ky-TM* counterparts. Therefore, we conclude that the whole domain has a role in inducing hypertrophy in *ky*-deficient muscle, but the individual catalytic residues do not.

Chapter 5: The human KY

5.1 In-silico human KY analysis

One report describing a patient with a *KY* myopathy suggested that the human *KY* (h*KY*) localizes to the sarcolemma (Straussberg, *et.al.*, 2016), contradicting previous research which shows *KY* is a Z-disk associated protein (Beatham, *et.al.*, 2004). Therefore, we wanted to unequivocally identify the subcellular localization of h*KY* in vivo. The h*KY* and mouse *ky* (m*Ky*) share a sequence identity of ~ 91 % and are 94% similar; thus, it is unlikely the small sequence differences observed within their alignment would produce a significant change in the protein's function or localization.

To find a suitable h*KY* candidate a pBLASTp of the m*Ky* was run against *homo sapiens*. Results showed all isoforms had greater than 80% homology with the m*Ky*. The multiple h*KY* isoforms identified all showed the exact same amino acid starting sequence as the m*Ky* (MELKKD). Intriguingly, one h*KY* Isoform (h*ky* X5) which also displayed a high homology with m*Ky* had the starting sequence of the m*Ky* aligned upstream of the start site for h*KY* X5. Further analysis showed that the h*KY* X5 contained an extra 59 amino acids preceding the MELKKD start sequence of the h*KY* (Figure 19.A) and m*Ky*. This newly identified region started with a Methionine which was in frame with the rest of the protein.

Interestingly, using nuclear localization signal predicting software (Nguyen, *et.al.*, 2009) we found that this region contained a nuclear localization signal (NLS). The extra region including the MELKKD motif, was run through multiple NLS predictors (Nguyen Ba, *et.al.*, 2009; Jih-Rong Lin, *et.al.*, 2012). Both predictors gave us readouts for an NLS within this extra region, with cut-off scores above 0.9. The NLS motif giving the highest cut-off score was RRPGRK (Figure 19.B) (Appendix). This newly identified extra sequence was predicted to enable the h*KY* X5 isoform to localize to the nucleus.

A	hKy	-----M	1
	hKy-X5	MACLAESAKSSALRLQRRGSALFRALALSRVGQLRARGHLQRKQPSVRRPRRGRKGGGIM*	60
	hKy	ELKKDINAVSIDMLLIVHSEKRRAAQGFQGVGNGVRRWQKLEGNDFHGTQLTVEVHPRDA	61
	hKy-X5	ELKKDINAVSIDMLLIVHSEKRRAAQGFQGVGNGVRRWQKLEGNDFHGTQLTVEVHPRDA *****	120
	hKy	MPQLLKKFSLAKRLQGDKNGNTRPRQGGKDAHAYPWRSSLSKMSLDLQQFEKLDIYAS	121
	hKy-X5	MPQLLKKFSLAKRLQGDKNGNTRPRQGGKDAHAYPWRSSLSKMSLDLQQFEKLDIYAS *****	180
	hKy	QVTAKSGLDELVSDLLQEHTDLERVRAIWIWICHHIEYDIAAAQEKDRQAFKPTDILRT	181
	hKy-X5	QVTAKSGLDELVSDLLQEHTDLERVRAIWIWICHHIEYDIAAAQEKDRQAFKPTDILRT *****	240
	hKy	QKTNC DG YAGL FERMCR LAGVQCMTVPGYSKGFQYQTGQSFSGEFDHAWNNAVYLEGRWHL	241
	hKy-X5	QKTNC DG YAGL FERMCR LAGVQCMTVPGYSKGFQYQTGQSFSGEFDHAWNNAVYLEGRWHL *****	300
	hKy	VDSTWGSGLVDTITSKFTFLYNEFYFLTHPALFIEDHFPDNKNWQLKPPQSLRQFENNM	301
	hKy-X5	VDSTWGSGLVDTITSKFTFLYNEFYFLTHPALFIEDHFPDNKNWQLKPPQSLRQFENNM *****	360
	hKy	YHKSEFYNGMLSAHPETSMIRTVNGKATVTIESCAPTLFMFMLNGKQEHGLLSLRKNGM	361
	hKy-X5	YHKSEFYNGMLSAHPETSMIRTVNGKATVTIESCAPTLFMFMLNGKQEHGLLSLRKNGM *****	420
	hKy	KLEVYPPMTGTHKLQIFAKGNSDIYSSVLEYTLKCNVDMGVQLPAELHQPVGPSWFSEQ	421
	hKy-X5	KLEVYPPMTGTHKLQIFAKGNSDIYSSVLEYTLKCNVDMGVQLPAELHQPVGPSWFSEQ *****	480
	hKy	MGIMKPSHPDPIIHTSDGRCSISFSVEEGINVLASLHGDDGPITEETQRRYIFQLHREKQ	481
	hKy-X5	MGIMKPSHPDPIIHTSDGRCSISFSVEEGINVLASLHGDDGPITEETQRRYIFQLHREKQ *****	540
	hKy	TELKVQLPHAGKFALKIFVKKRQEPGNVIFVFNVLVCCANTKVNWPMFPESFGNWDQNE	541
	hKy-X5	TELKVQLPHAGKFALKIFVKKRQEPGNVIFVFNVLVCCANTKVNWPMFPESFGNWDQNE *****	600
	hKy	LLEPLSGVLPANRNVPFKLLHGIKVLVKGQDTWPLTLNHEGYWEGSCSTAGCQEVVVM	601
	hKy-X5	LLEPLSGVLPANRNVPFKLLHGIKVLVKGQDTWPLTLNHEGYWEGSCSTAGCQEVVVM *****	660
	hKy	VLENANHNFYSYILKYKVNQAQ	622
	hKy-X5	VLENANHNFYSYILKYKVNQAQ *****	681

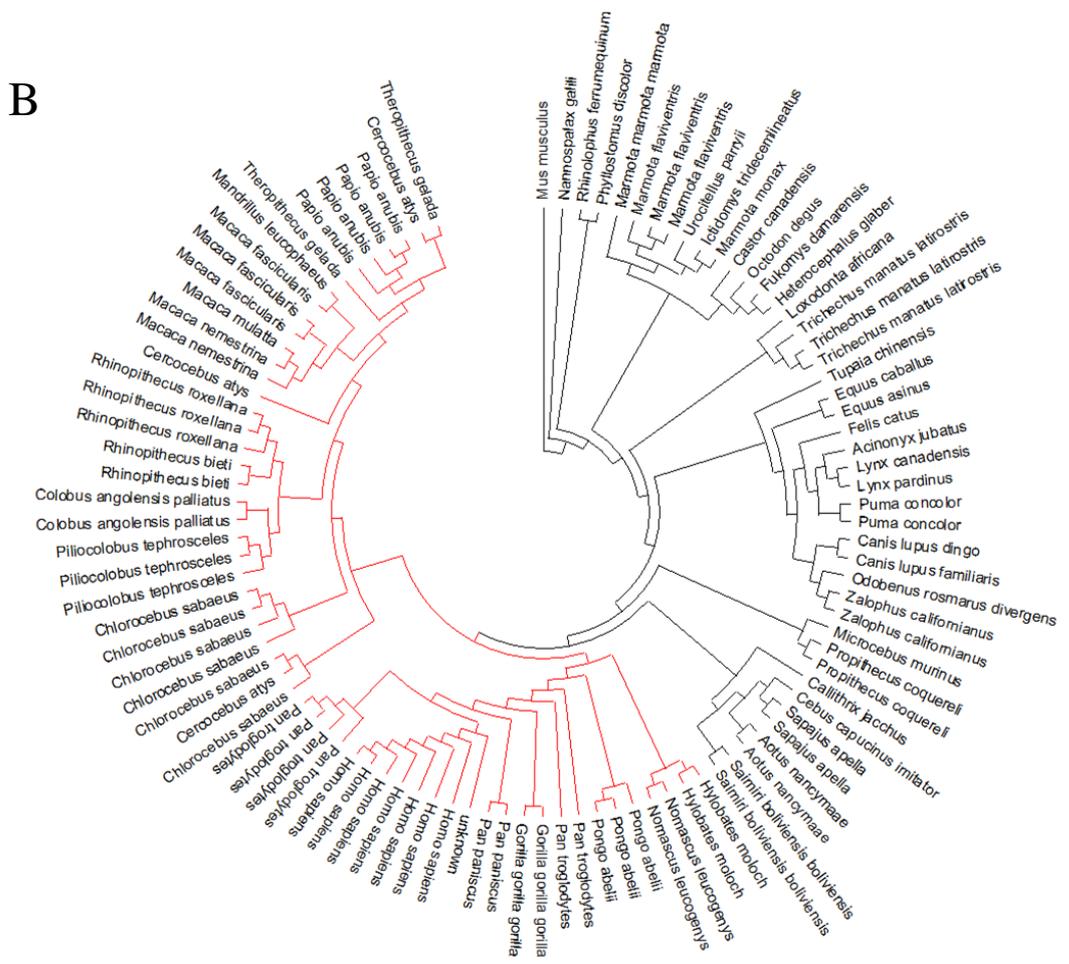
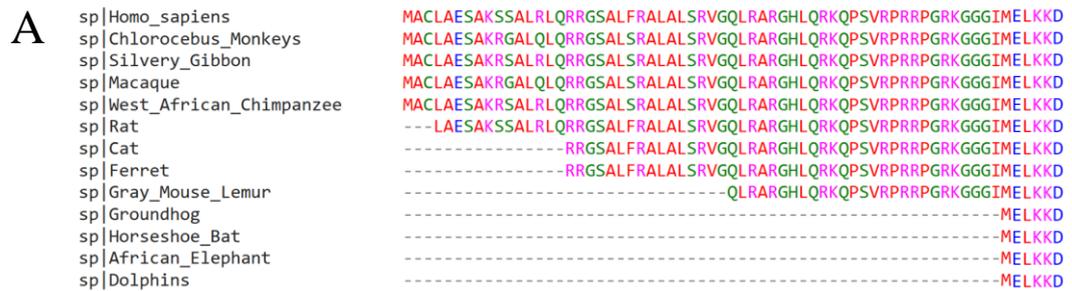
B

MACLAESAKSSALRLQRRGSALFRALALSRVGQL**RARGHLQRKQPSVRRPRRGRKGGGIMELK**

⏟
Nuclear localization signal

Figure 19 Protein sequences of Homo sapiens hKY-X5 against the hKY. A protein alignment shows the extra region identified within the N terminus of hky-X5. **B) KY isoform X5 nuclear localization signal.** The extra sequence identified was found to contain a nuclear localization signal (Orange text). The starting Methionine is highlighted in red along with the second Methionine. The (Blue text) has the highest NLS cut of score above 0.9.

Next, we wanted to look at what other species may contain this extra sequence. A tBLASTn search was run using the extra region found in hKY-X5. Results of the BLAST search showed that only primates conserved this sequence out of all other mammals (Figure 20.A). All identified Primates showed conservation of the first Methionine of the extra region, including the NLS in frame with the original start codon and motif MELKKD.



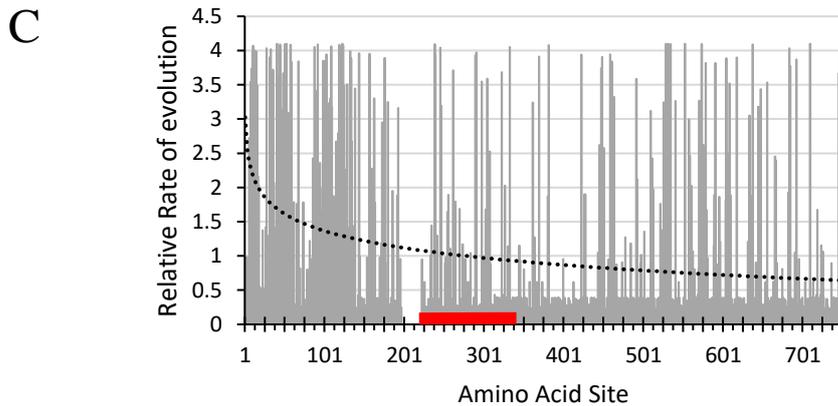


Figure 20. Conserved nuclear localization signal within primates shows high genetic divergence within other species. A) A group of Aligned ky Sequences. The beginning sequences from a collection of ky sequences taken from multiple species. B) Circular Newic phylogenetic tree from a collection of ky homologues from different species. The tree was generated using MEGA-X. Primates are highlighted in red showing their distinct group. C) Evolution rate at each protein site of aligned KY homologues. Using the Jones, Taylor, and Thornton model (JTT) model of evolution the rate of evolution at each protein site was predicted. Five gamma categories gave the relative rate. The TGN/PROT domain is highlighted by the red bar.

To assess conservation of this region, a collection of similar genes were identified through a pBLASTp using the full length hKY X5 isoform as the sequence of interest. The homologous sequences were a series of Ky variants from different species. Using MEGA-X these sequences were aligned and used to build a phylogenetic tree using the Newic format (Hehenberger, 2018). The polygenetic analysis revealed primates diverge from a common ancestor and constitute their own group (Figure 20.B).

To further understand the origins of this primate conserved region, extensive searches were run on ensemble and BLAST looking for conservation of this region within any other species. A tBLASTn search was run using the extra region containing the NLS and no hits were found within fish. Interestingly, some species showed partial conservation of this region (Figure 20. A). The only examples where the entire region with the NLS was conserved were, American Beavers, Dogs, and Guinea pigs although in each of these cases, the conserved region was out of frame with the rest of the KY

protein. Thus, concluding that this region containing the NLS sequence is solely conserved within primates.

Finally, we wanted to look at the rate of evolution at each amino acid site in an alignment of multiple KY homologues. The rate of evolution at a given site is calculated, which maximizes the probability of the given alignment under the selected model of evolution. In this case the Jones, Taylor, and Thornton (JTT) evolution model (Jones, *et.al.*, 1992) was used.

Results indicated a higher level of protein evolution at the beginning of the protein, shown by a logarithmic trendline (Figure 20.C). As you move along the protein, levels of evolution decrease. It is likely this reflects the fact that many species other than primates have either lost or only show partial conservation of the extra region which holds the NLS, while conserving the majority of the KY protein which starts at the second Methionine with the motif MELKKD. A gap at 200-220 amino acids represents a gap within the aligned data, where a single KY protein, derived from the *papio Anubis* (Olive Baboon), included an extra amino acid sequence, which no other species contain. This species being within the primates also included the extra region with the NLS.

5.2 In vitro overexpression of human KY

To assess the subcellular localisation of hKY and test whether the newly identified NLS could cause nuclear localisation, a new construct of the hKY X5 isoform was ordered from GenScript. The hKY X5 was cloned into the mammalian expression vector pcDNA3.1(+)-C-eGFP through a HindIII/BamHI cloning strategy. This subsequently produced the new hKY vector pcDNA3.1-hKy-eGFP (hKy-eGFP) (Figure 21). Two restriction digests were used to confirm the correct cloning. Finally, sequencing was used to confirm the flanking sequences of the cloning site, both of which were correct.

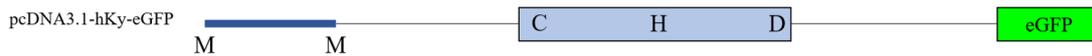


Figure 21. Cloned human KY isoform X5 (hKy-eGFP). The human ky isoform X5 was cloned in the vector pcDNA3.1-(+)-eGFP the cloned region shown above. The Highlighted blue region on the left indicates the extra region which contains the NLS, the two-start methionine's are shown above. The light blue box indicates the conserved TGN/PROT domain shown are the conserved Cysteine, Histidine and Aspartate. eGFP is attached to the C-terminus of the protein shown on the right.

To test the subcellular localisation of hKy-eGFP within myoblasts, lipofectamine transfections within C2C12 myoblasts were undertaken. We wanted to test whether the hKY-X5 would localise to the cytoplasm, similarly to the mKY within a proliferating C2C12 myoblast, or whether the predicted NLS sequence was sufficient to cause nuclear localisation, C2C12 myoblasts were grown to ~80% confluency, transfected with hKy-eGFP using GenJet *In-vitro* Transfection reagent according to the manufacturer's instructions. Cells were left to incubate for 48 hours before fixation and mounting. Cells were then imaged using fluorescent microscopes.

Results showed nuclear and cytoplasmic expression (Figure 22). All cases of cytoplasmic expression were observed as a normal diffuse pattern. Conversely the nuclear expression showed two variations; one variation showed a "speckled" pattern of expression (Figure 22.A&B), while the other showed a diffuse expression across the whole nucleus, similar to that observed in the cytoplasm (Figure 22.C&D).

Overexpression of hKy-eGFP within C2C12 myoblasts comes in two patterns. Nuclear and cytoplasmic expression is clear in both cases, but the differences lie within the nucleus. One pattern shows diffuse expression across the whole nucleus. Conversely the other shows a "Speckled" pattern of expression.

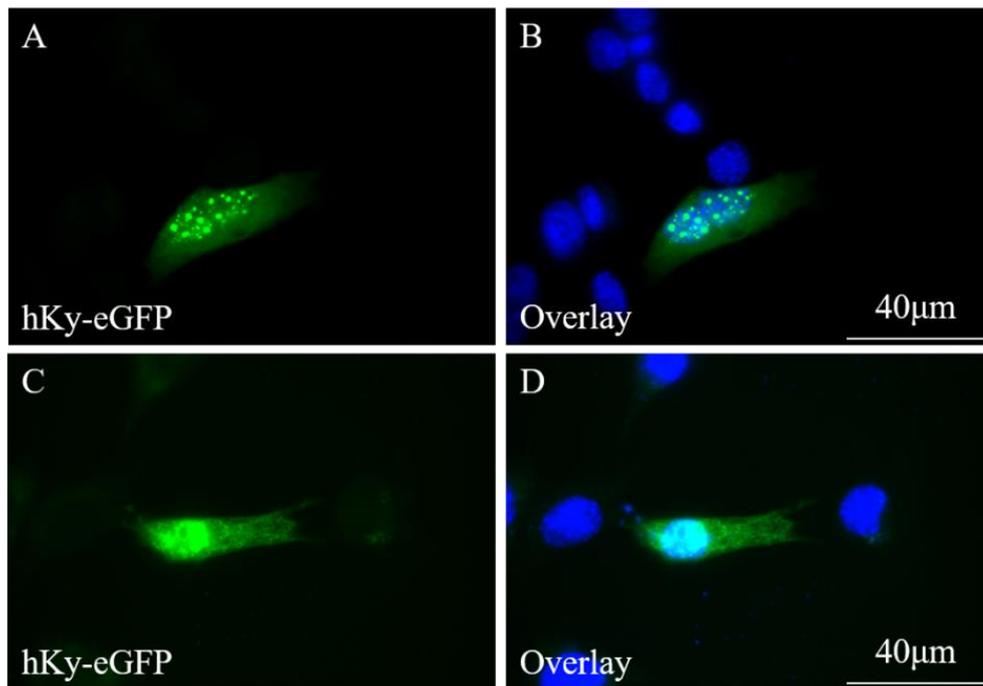


Figure 22. Cytoplasmic and nuclear localization of pcDNA3.1-hKy-eGFP. A-B) pcDNA3.1-hKy-eGFP transfected into C2C12 myoblasts. pcDNA-hKy-eGFP is detected using a FITC filter. Whilst DAPI is detected using a DAPI filter. A-B) Nuclear protein aggregates are seen showing a “speckled” pattern of expression. C-D) Cytoplasmic Expression is seen, whilst still showing nuclear expression.

5.3 Co-localisation of IGFN1 V1 and the human KY

The Immunoglobulin-like and fibronectin type III domain containing 1 (IGFN1) protein was one of the first proteins to be identified in a yeast-two-hybrid assay to interact with KY alongside other sarcomeric proteins such as FLNC (Beatham, *et al.*, 2004). IGFN1 is specifically expressed within skeletal muscle and is a complex gene, supporting many isoform variants. IGFN1 lacks any enzymatic activity, therefore it is likely that it initiates protein complex formations. One such formation is the IGFN1, FLNC and KY complex, which has been suggested to provide structural support at the Z-disk (Baker, *et al.*, 2010). IGFN1 is known to localise to the nucleus and is able to translocate to the cytoplasm (Baker, *et al.*, 2010).

As previously seen, expression of the hKY in C2C12 myoblasts led to a nuclear localisation, showing in some cases a “speckled” pattern of expression (Figure 22).

The pattern seen was similar to that of IGFN1_V1-tdTomato when overexpressed within COS7 myoblasts (Figure 23). To test if both proteins showed nuclear or cytoplasmic colocalisation or affected each other's localisation, a co-transfection with both the hKy-eGFP and IGFN1_V1-tdTomato was performed.

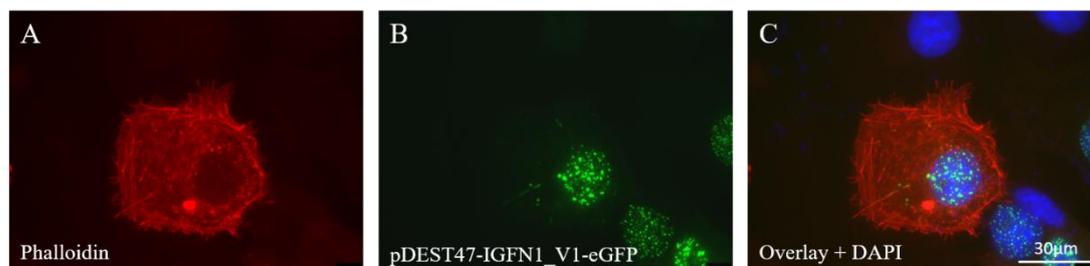


Figure 23. Nuclear localization pattern of IGFN1_V1. IGFN1_V1 transfected into COS-7 fibroblasts. **A) Phalloidin stain highlighting the actin.** **B) pDEST47-IGFN1_V1-eGFP expression profile.** **C) Overlay of the phalloidin stain, IGFN1_V1, and DAPI.** A scale bar is shown in the overlay.

In this transfection, C2C12 myoblasts were transfected with pDEST47-IGFN1_V1-tdtomato (IGFN1_V1-tdTomato), a smaller isoform of IGFN1 without the disordered middle region, which has a higher transfection efficiency. GenJet transfection reagent was used to co-transfect IGFN1_V1-tdTomato alongside hKy-eGFP into C2C12 myoblasts. Appropriate controls were set up: IGFN1_V1-tdTomato was transfected alongside pDEST-eGFP, whilst hKy-eGFP was co-transfected with pDEST47-tdTomato. Post-transfection, myoblasts were left to express the constructs for 48 hours, then fixed and mounted with DAPI.

Observations of co-transfected myoblasts indicated that hKy-eGFP and IGFN1_V1-tdTomato were colocalising (Figure 24. A-I). Large colocalised complexes formed within the nucleus, whilst cytoplasmic expression seemed more diffuse. ImageJ was then used to quantify protein colocalization. Images were subjected to a specific method to test colocalization. The Coloc2 plugin was utilised within ImageJ, Coloc2 uses Pearson's correlation coefficient, with costes method (Costes, *et.al.*, 2004).

Results showed a high level of co-localisation of hKy-eGFP and IGFN1_V1-tdTomato, indicated by a high Pearson's correlation coefficient score, alongside

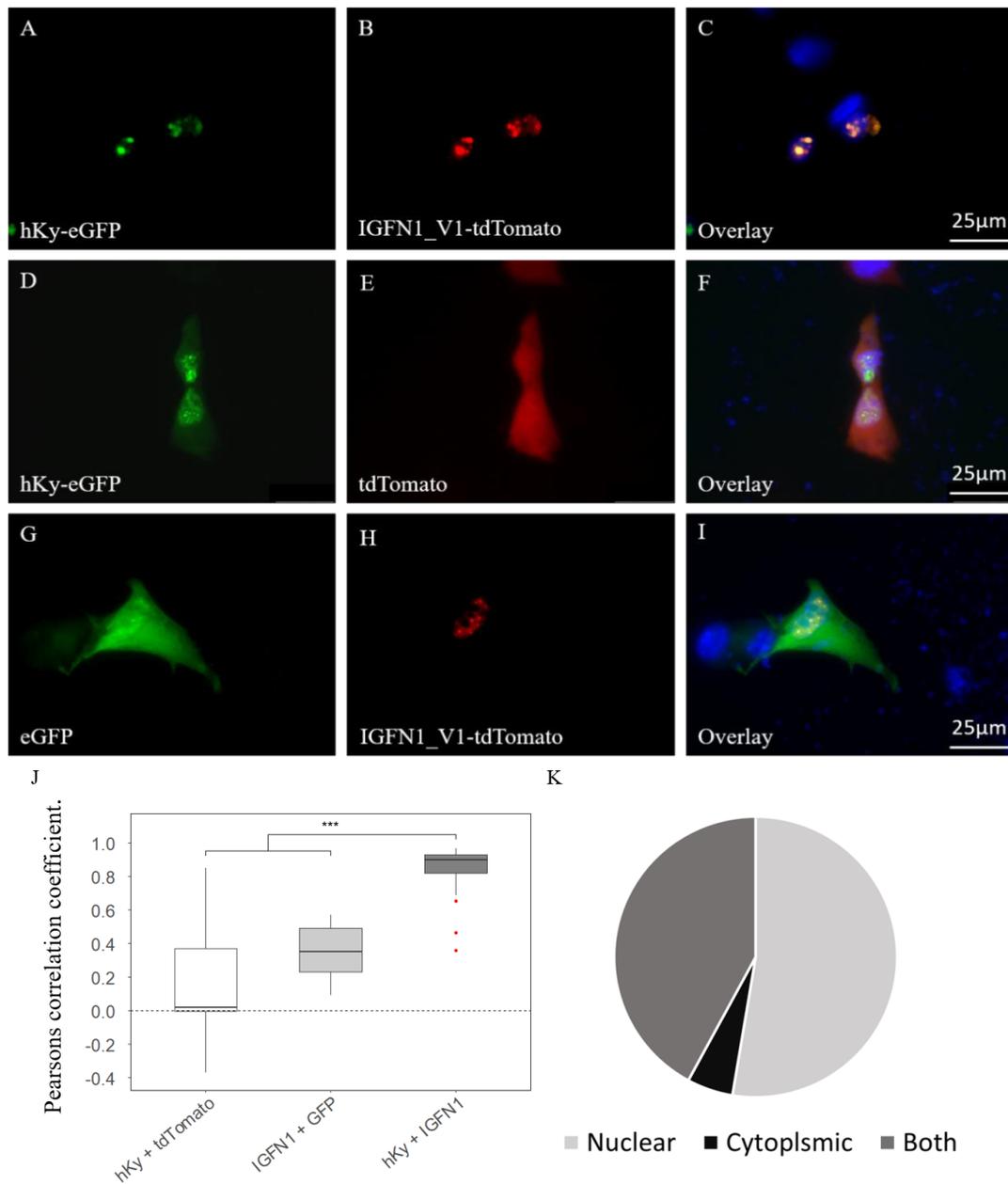


Figure 24. Colocalization of *pcDNA3.1-hKy-eGFP* and *pDEST47-IGFN1_V1-tdTomato*. A-I) Fluorescent images of co-transfected *IGFN1_V1* and *hKy-eGFP*. Rows consist of different co-transfected constructs; each construct is shown in bottom left of each panel. The right column consists of an overlay of both constructs and DAPI. A scale bar is shown in each overlay and corresponds to that row J) Box plot of Co-localization between co-expressed proteins. Pearson correlation coefficient (no threshold) was used to quantify co-localisation. The median is shown as the middle line inside the box, the lower and upper quartile are shown by the upper and lower boundaries of the box. The lines indicate the range of the data, outliers are shown, $P = \text{*} < 0.001$. ANOVA result: F value = 55.66, $df = 54$, p -value = 0.0063. K) Pie chart of the distribution of co-localised clusters. A count represented as a percentage; no statistical analysis could be performed.**

having a small variance (Figure 24. J). Both controls also showed significantly lower Pearson's correlation coefficients and indicated near to no co-localisation. Further to this both showed a much higher variance compared to hKy-eGFP and IGFN1_V1-tdTomato.

A count was performed to give an indication as to where the aggregates of hKy-eGFP and IGFN1_V1-tdTomato were forming. Three possibilities were available: nuclear localisation, cytoplasmic localisation or both nuclear and cytoplasmic. Results showed a large amount of these aggregates localised to the nucleus (Figure 24. K). Second to that, aggregates were observed in both the nucleus and cytoplasm. Lastly a very small number of aggregates were observed only within the cytoplasm.

The results here show clear colocalised aggregates between hKy-eGFP and IGFN1_V1-tdTomato, most of these colocalised aggregates form within the nucleus. Statistical analysis further supports a high level of colocalization between these two constructs. In both co-transfection controls, the expression profile of the hKy-eGFP and IGFN1_V1-tdTomato are highly comparable.

5.4 Co-Transfection of IGFN1 V1 and the mouse KY

The mKY does not contain an NLS and has never been observed within the nucleus. Therefore, we wanted to test whether IGFN1_V1-tdTomato alone could cause nuclear translocation of the mKY.

The same method for transfection was used, as previously stated, but pDEST47-IGFN1_V1-tdtomato was swapped for pDEST47-IGFN1_V1-eGFP, as the Ky-td (pDEST47-Ky-tdTomato) contained tdTomato. pDEST47-IGFN1_V1-eGFP and Ky-td were co-transfected into C2C12 myoblasts. Additionally, pDEST47-IGFN1_V1-eGFP was co-transfected alongside tdTomato and Ky-td was co-transfected with pDEST47-eGFP as controls. Post transfection, myoblasts were left to express the constructs for 48 hours, fixed, then mounted with DAPI.

No clear colocalization could be seen between Ky-td and pDEST47-IGFN1_V1-eGFP (Figure 25.A-C). The typical cytoplasmic expression profile of Ky-td was seen, and importantly, in no cases was Ky-td seen within the nucleus. IGFN1_V1-eGFP

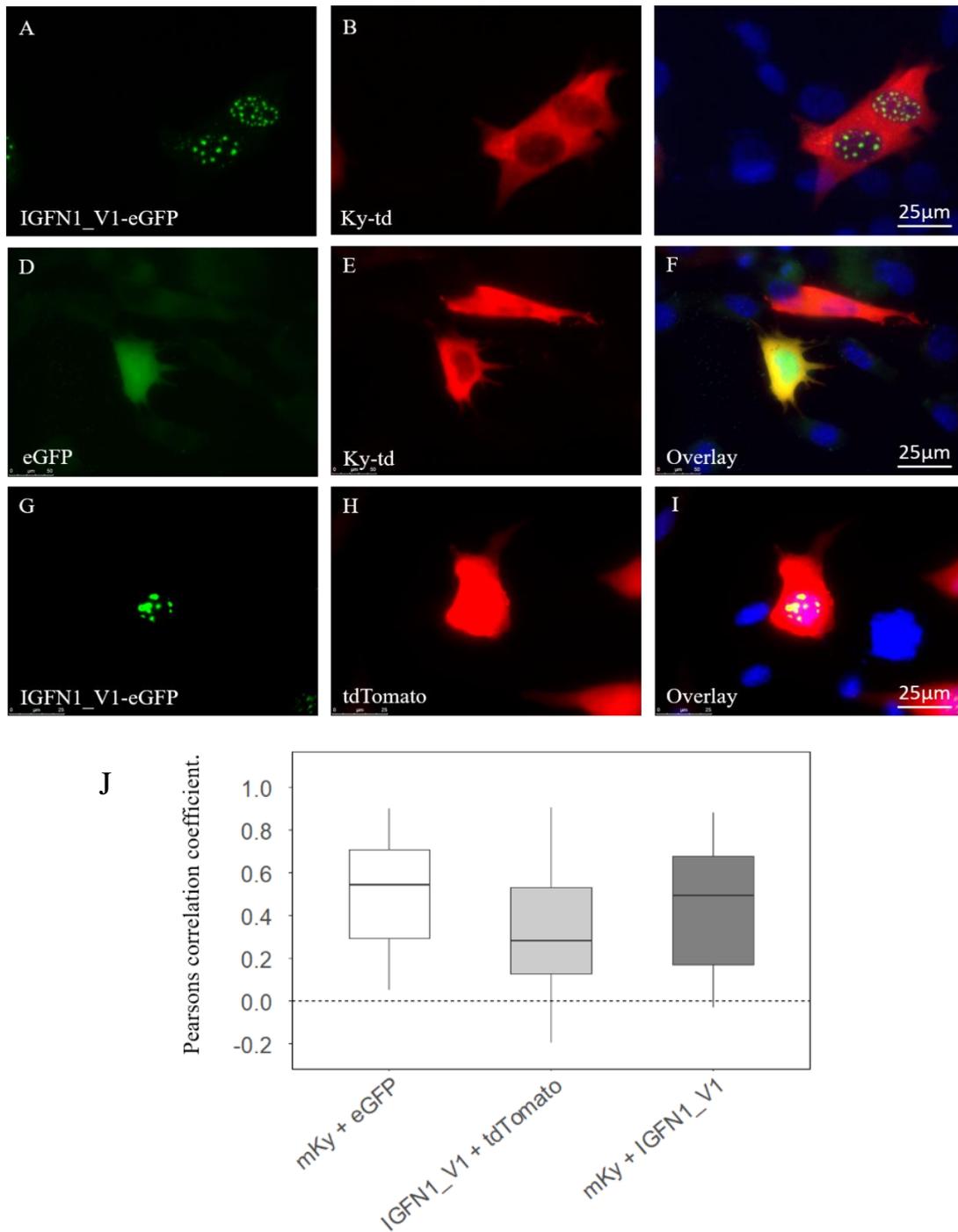


Figure 25. Co-expression of pDEST47-mKy-tdTomato and pDEST47-IGFN1_V1-eGFP. A-I) Fluorescent images of co-transfected mKy-tdTomato and IGFN1-eGFP. Rows consist of different co-transfected constructs; each construct is shown in bottom left of each panel. The right column consists of an overlay of both constructs and DAPI. Scale bars are shown in the bottom right-hand corner of each row. J) Box plot of Co-localization between co-expressed proteins. Pearson correlation coefficient (no threshold) was used to quantify co-localisation. The median is shown as the middle value inside the box, the lower and upper quartile are shown by the box. The lines show the maximum and minimum values, outliers are shown. ANOVA result: F value= 1.778 =, df =61, p =0.178.

showed its typical “Speckled” pattern of expression (Figure 23). Again, the Coloc2 plugin for image J was used to analyse colocalization utilising the Costes method (Costes, *et.al.*, 2004) (Methods). Results show a low level of colocalization between all colocalised constructs, with no significant difference between each group (Figure 25.J).

Results here suggest that Ky-td and pDEST47-IGFN1_V1-eGFP do not colocalise, as no visible colocalised aggregates were observed with the data analysis revealing a low Pearson’s correlation coefficient. Furthermore pDEST47-IGFN1_V1-eGFP is unable to induce nuclear translocation of Ky-td. These results confirm previous observations indicating that the mKY is a cytoplasmic protein (Baker, *et.al.*, 2010).

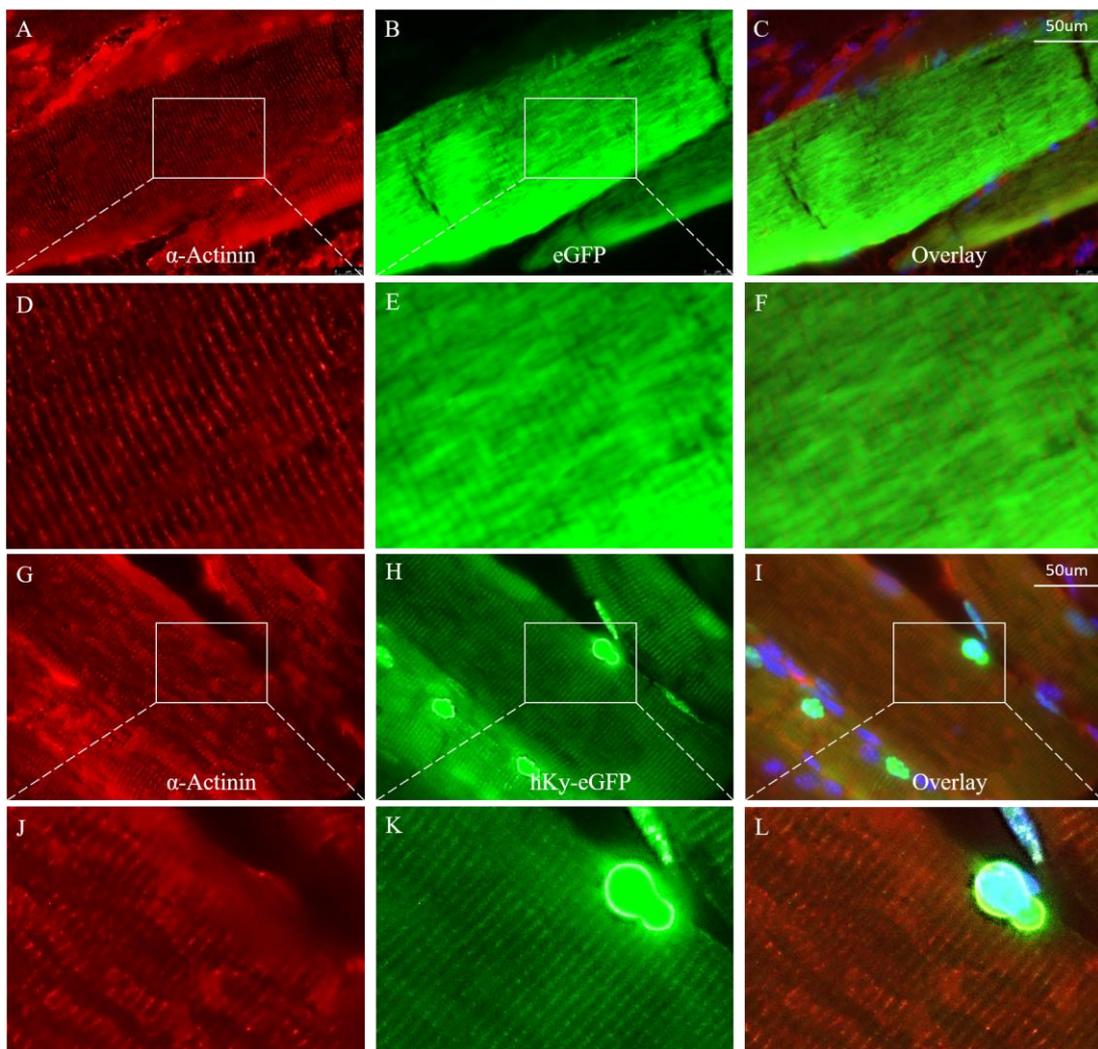
5.5 In Vivo localisation of human KY

In-silico analysis and *In-vitro* experiments have revealed the hKY X5 isoform contains an extra region, in which a functional nuclear localisation signal is present (Figure 19 & 22). It is yet to be confirmed whether this protein can localise to the Z-disk. Due to the high homology with the mouse *Ky*, it is highly likely Z-disk localisation will be observed along with nuclear localisation, when overexpressed *in-vivo* in fully differentiated skeletal muscle.

To test this theory, we electroporated hKy-eGFP into mouse TA muscles, using the same surgical method as before, eGFP was used as a control for hKy-eGFP. Post-electroporation, mice were left for 8 days to recover and express the recombinant proteins. After 8 days, mice were sacrificed, and the TA was dissected and frozen in liquid nitrogen-cooled isopentane. Longitudinal muscle sections of 12 μ m were taken, immunostained, and mounted with DAPI. The Z-disks were again marked with an EA53 antibody against α -Actinin. High magnification images were then taken using a fluorescent microscope (Figure 26.A-L).

Images of hKy-eGFP electroporated fibers showed clear Z-disk striations, which overlaid perfectly with the α -actinin. These results suggest that like the mKY, hKY X5 too is able to localise to the Z-disk. When compared with eGFP, which also shows some faint striations (Figure 226D-F), it is clear there is a significant difference

between the patterns of hKy-eGFP and eGFP. The striations seen within electroporated eGFP fibers are likely an artefact, due to the Z-disk being such a prominent structure within differentiated fibers. Furthermore, nuclear localisation was also observed with hKy-eGFP. Higher magnification images also revealed a “speckled” pattern of expression seen in most nuclei (Figure 26.M-O) similar to that which was observed *in vitro* (Figure 22).



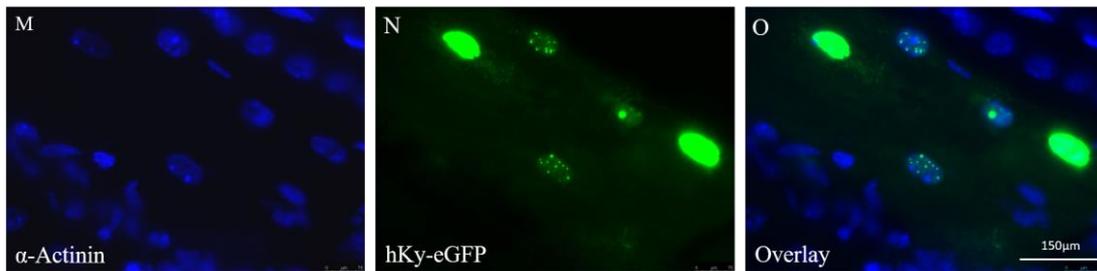


Figure 26. *pcDNA3.1-hKy-eGFP* localizes to the Z-disk and the nucleus *in vivo*. Longitudinal sections of electroporated wildtype muscle fibers. The relevant a scale bar is shown, corresponding to the panels within that row **A-C) Electroporated eGFP**. The left panel shows α -actinin highlighting the Z-disk, the middle panel shows electroporated eGFP. The final panel shows an overlay image of the two images including DAPI. **D-F) Magnified images of A-C**. Magnified images are highlighted in the original part of the image taken the white box. **G-I) Electroporated hKy-eGFP**. The left panel shows α -actinin highlighting the Z-disk, the middle panel shows electroporated hKy-eGFP. The final panel shows an overlay with DAPI. **J-L) Magnified images of G-I**. White box represents magnified area. **M-O) Low exposure images of electroporated hKy-eGFP**. Low exposure allowed visualization of the nuclear localization. The left panel shows DAPI, the middle panel shows electroporated eGFP. The final panel shows an overlay with a scale bar.

It is interesting to note that the expression profile observed with the hKy-eGFP highly resembles the *in vivo* expression profile of electroporated IGFN1_V1-tdTomato, which also shows nuclear and Z-disk localisation (Figure 24).

Here we have seen that hKy-eGFP is able to localise both to the Z-disk and myonuclei in fully differentiated skeletal muscle. The eGFP tag does not affect localisation of the protein exhibited by the different expression profile of the eGFP control. Furthermore, the profile observed with hKy-eGFP resembles the expression pattern of electroporated IGFN1_V1-tdTomato (Figure 27). This data, alongside what we previously knew regarding the interaction of KY and IGFN1 *in vivo*, suggests that there is most definitely a role for the interaction of these two proteins, which is not exclusive to the myoplasm, it is just waiting to be discovered.

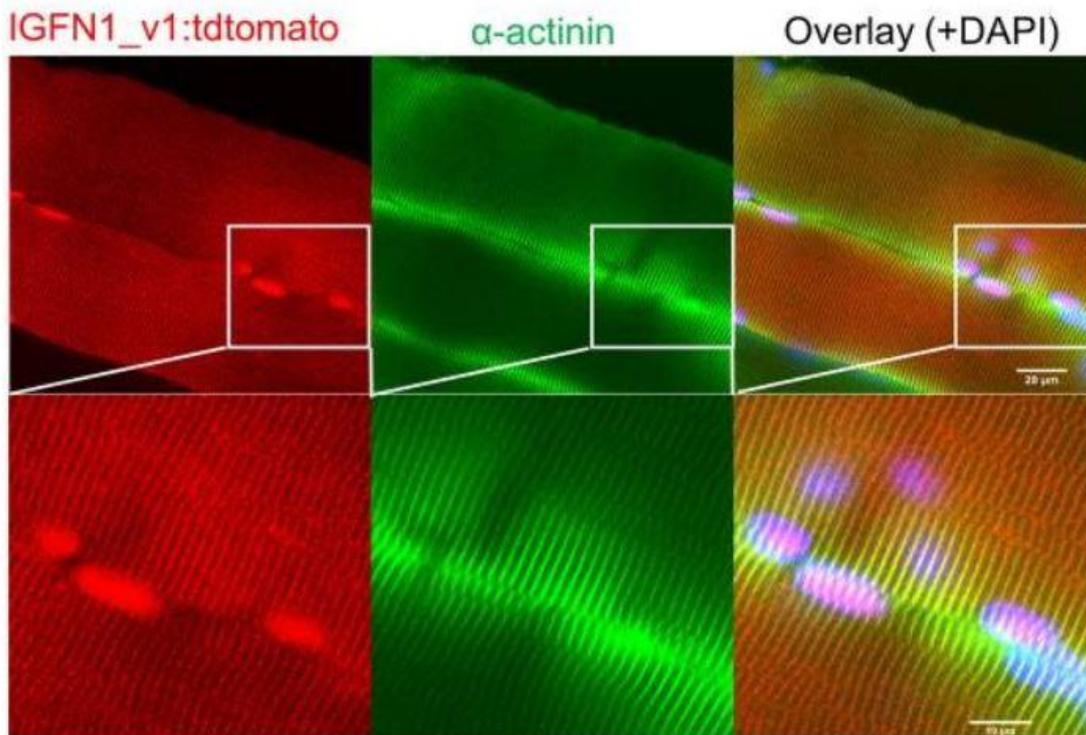


Figure 27. IGFN1_V1 localises to the Z-disk. Longitudinal sections of IGFN1_V1 electroporated wildtype muscle fibers. Each column shows the indicated construct, the α -Actinin was stained with an EA53 antibody. The bottom row shows an enlarged section of the fiber indicated within the white box. A relevant a scale bar is shown for each corresponding row. This image was taken from Li, et.al., 2017 and is available at <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0180217>.

Chapter 6:
Discussion and
Future work

The primary aim of this thesis was to gain a better understanding of the mechanism by which KY acts, through defining a definitive role for the highly conserved TGN/PROT domain. The secondary aim of this thesis was to explicitly define the subcellular localisation of the hKY, as a previous report has suggested the hKY was localising to the sarcolemma of a muscle fiber (Straussberg, *et.al.*, 2016).

6.1.1 KY does not contain any enzymatic activity within the conserved TGN/PROT domain

The specific residues (Cysteine [225], Histidine [267], and Aspartate [282]) forming the catalytic triad within KY's TGN/PROT are conserved among species. These residues are required for enzymatic ability (Carter and Wells, 1988), and as previously discussed, suggest that KY may hold transglutaminase or protease enzymatic activity. Here full length purified KY protein and MBP-KY(aa160-285) have both been tested for enzymatic activity. Protease and transglutaminase assays performed on both KY and MBP-KY(aa160-285) suggest that KY has lost any potential enzymatic abilities (Figures 14 & 15). Furthermore, the complementation studies performed on KY's TGN/PROT domain mutants (Figure 18) would also suggest that enzymatic ability is not needed for its mechanistic action. Thus, Ky-TM which had all three catalytic residues substituted to alanine, would have likely lost any potential enzymatic ability (Carter and Wells, 1988), and yet was able to rescue fiber size in *ky/ky* mice. Thus, suggesting that these individual residues and in turn enzymatic activity isn't required for KY protein function. Collectively this data provides strong evidence that KY is not enzymatically active, and that replacement of the conserved catalytic residues has no effect on the overall function of the resulting protein.

The formation of catalytic triads is known to be dependent on the distance between the catalytic residues. Other enzymatically active eukaryotic transglutaminases which contain the transglutaminase catalytic triad consisting of Cystine, Histidine, and Aspartate all conserve 23 amino acids between there Histidine (H) and Aspartate (D) catalytic residues. The *homo-sapien* and *mus musculus* (Mouse), TGM family of proteins (UniProt, #P22735) and the coagulation factor XIII protein (UniProt,

#P00488), all have 23 amino acids between Histidine and Aspartate, whilst conversely KY only contains 15 amino acids between Histidine and Aspartate (UniProt, #Q8NBH2). This could cause a significant disruption within the shape of the active site and thus prevent enzymatic activity from occurring. One interesting observation is the fact the catalytic residues have been conserved throughout evolution. It has been shown that these specific residues are not required for the rescue of *ky/ky* muscle fibres, and do not serve any enzymatic function. Therefore, the question remains as to why these specific residues have been conserved, as they seem to serve no apparent function within the KY protein.

6.1.2 KY's TGN/PROT domain is not required for Z-disk localisation

KY has been shown to be a Z-disk protein (Beatham, *et.al.*, 2004), highly conserved among species. The *Ky* homologues all hold a TGN/PROT domain, which we now know to not be enzymatically active. As KY's only distinguishable feature, it was hypothesised that the TGN/PROT domain was causing Z-disk localisation. Here we have shown this not to be the case, and therefore indicating that the region causing Z-disk localisation is conserved elsewhere within the protein. Looking at the amount of evolution along the KY protein (Figure 20.C), the C-terminus appears to have evolved the least. Therefore, the domain controlling Z-disk localisation is likely to be contained within the C-terminus, due to the higher level of conservation compared to the rest of the protein.

6.1.3 The human KY can localise to the Z-disk

Previous reports have suggested that the hKY localises to the sarcolemma of the muscle fiber (Straussberg, *et.al.*, 2016). Being a highly conserved protein, we predicted that the hKY, would localise to the Z-disk. This thesis has shown that primates contain an atypical version of the *Ky* gene, holding an enclosed nuclear localisation signal within an extra 49 amino acids upstream of the original start site. Here it has been unveiled that the hKY localises to the Z-disk like its counterparts, but also localises to myonuclei within a fully differentiated muscle fiber and myoblasts.

The data here has clearly shown that the human KY is capable of localising to the Z-disk.

6.2 Overview

This thesis has once again highlighted the complex nature of KY's role within muscles. This elusive protein has evolved down two separate pathways; primates conserving a NLS signal, while other mammals seem to have removed this NLS. *Ky* is known to be imperative for normal muscular function shown by the severe pathology seen within the *ky/ky* mouse and the human *Ky* myopathies, the similarities between these phenotypes also highlight the conserved function of KY in primates and other mammals. Here we have shown the TGN/PROT domain is integral to the function of KY. Complementation experiments show that when the whole TGN/PROT domain is removed, KY is unable to rescue fiber size through hypertrophy. Conversely, when wildtype KY is electroporated into *ky/ky* muscle, rescue occurs through hypertrophy. Thus, the TGN/PROT domain must have an integral role within KY's mechanism of action to maintain normal muscle pathology.

Muscular hypertrophy is essential to muscular health and is a normal response to functional overload of a muscle. To date most of the signalling cascade which allows for muscular hypertrophy has been identified, although the initiating signals remain elusive and highly debated (Wackerhage, *et.al.*, 2019). The *ky/ky* mouse has been shown to lack a hypertrophic response (Blanco, *et.al.*, 2001), thus KY was identified as a regulator of the hypertrophic response. However, a direct link between KY and hypertrophy had not been established. The mKY has never been seen within the nucleus and is not known to interact with any proteins involved within induction of hypertrophic pathways. The finding of a functional NLS within primates' *KY* locus now provides evidence that KY itself may be able to modulate hypertrophic pathways, due to a functional nuclear link. KY is known to interact with a multitude of sarcomeric proteins (Beatham, *et.al.*, 2004), such as IGFN1 and FLNC. Like many other sarcomeric proteins, KY localises to the Z-disk, where it has been implicated to provide structural support to the Z-disk, through forming large protein complexes

(Baker, *et.al.*, 2010). The clear myofiber pathologies seen within *ky/ky* mice and the human myopathies highlight the critical role of KY within maintaining sarcomeric stability. These pathologies also implicate the Z-disk as the primary site affected. The Z-disk is a known site of intracellular signalling and has been suggested to sense and respond to stretch and mechanical overload (Luther, 2009; Knöll, *et.al.*, 2011). Within the *ky/ky* pathology, thickened and abnormal Z-disks are commonly seen (Beatham, *et.al.*, 2004). This thickening of the Z-disks might be an adaptive response to try and further aid sarcomeric stability or conversely may reflect the adaptive switch to type I muscle fibers, a common hallmark of the *ky/ky* pathology. Type I muscle fibers are known to have wider Z-disks (Luther, *et.al.*, 2003).

In the broader context, KY's association with the Z-disk has led to the hypothesis that KY is being used for stretch sensing (Blanco, *et.al.*, 2001), which allows for a response to mechanical overload. If KY is not present within muscles, the myofibers are unable to undergo a hypertrophic response. In wildtype muscles, mechanical overload and stretch induces vast changes to cellular morphology, typically occurring through hypertrophy. In order for these changes to occur, the damage that stretch induces, or the stretch present at the Z-disk during the time of mechanical overload must be detected and responded too. This response is absent in *ky/ky* mice. Therefore, the evidence reviewed suggests that KY is being utilised to detect stretch, or the damage produced from mechanical overload, and thus allows for muscular adaptation via hypertrophy.

6.3 KY's TGN/PROT domain

A specific function for KY's TGN/PROT domain still remains unclear, yet it is vital for the function of KY. Therefore, this region must serve a critical function underpinning KY's mechanism. As we know, KY's TGN/PROT domain is not enzymatically active, thus we can assume that similarly to other eukaryotic transglutaminases, KY has repurposed the function of this domain (Anantharaman, *et.al.*, 2001). In this case potentially mediating KY's interactions with other sarcomeric proteins, and allowing for the formation of complexes, such as the IGFN1-FLNC-KY

complex (Baker, *et.al.*, 2010). Therefore, if the TGN/PROT domain is mediating the formation of these complexes, this would explain the inability of Ky-Del to rescue fiber size within the complementation studies. Further to this it would also explain why Ky-TM and Ky-SM were able to rescue fiber size within the complementation studies; with these mutations being unlikely to change protein folding drastically, and are unlikely to affect potential protein-protein interactions. If this is to be believed, it would suggest that the TGN/PROT domain and the catalytic residues have been conserved to maintain the precise 3D structure. This would in turn imply that small sequence variances seen within other species, should not impact function, as protein interactions will depend on the secondary and tertiary structure of the TGN/PROT domain. Therefore, large complexes can still form, like the IGFN1-FLNC-KY complex, facilitating Z-disk stabilisation, despite the mutations and small sequence differences seen between species.

One hallmark commonly seen within *Ky* deficiencies in both mice and humans is a large FLNC aggregates. It was suggested by Jokl, *et.al.*, (2018) that these FLNC aggregates occurred due to dysregulation of the CASA pathway, which has been suggested to clear damaged FLNC. It could also be plausible that KY could aid the localisation of FLNC to the Z-disk, occurring through complex formation with IGFN1 and FLNC (Baker, *et.al.*, 2010). Therefore, if KY is unable to bind FLNC this could lead to the FLNC aggregation observed within *ky/ky* mice and humans. Furthermore, within FLNC myopathies, which result in myofibrillar disintegration (a hallmark observed within *ky/ky* mice and *KY* myopathies; it was suggested that the sequestering of FLNC into large aggregates, and therefore away from Z-disks, preferentially caused the phenotype (Ruparelia *et.al.*, 2016). The authors showed that overexpression of the mutant FLNC (FLNC^{W2710X}) which typically forms these large aggregates could correctly localise to the Z-disk, and rescue the fiber disintegration phenotype within a zebrafish model of a FLNC knockdown. Interestingly these researchers also showed that BAG3 was contained within FLNC^{W2710X} aggregates, which is the same finding identified within the *ky/ky* soleus fibres (Jokl, *et.al.*, 2018). Hence, if FLNC is not able

to properly localise within *Ky* myopathies, this may contribute to the severe phenotype seen within these cases.

6.4 KY's Z-disk Localisation

KY's only distinguishable feature is the highly conserved TGN/PROT domain. Hence it was hypothesised that this domain was causing Z-disk localisation. Here we have shown this not to be the case, which has indicated that the region causing Z-disk localisation is held elsewhere within the protein. Looking at the amount of evolution along the KY protein (Figure 20.C), the C-terminus appears to have evolved the least. Thus, the domain controlling Z-disk localisation may be contained within the C-terminus, due to the fact there is a higher level of conservation compared to the rest of the protein.

6.5 The human *KY* gene

It was always likely the hKY would localise to the Z-disk, as the main body of the protein (not including the extra region in which the NLS is held), is highly similar to the mouse KY. It is probable that this region is mechanistically acting in a similar fashion to the mKY, with the TGN/PROT domain performing a similar, if not the same role. Again, correlating with the fact that similar pathways would be disrupted within the human *KY* myopathies and the *ky/ky* mouse, and thus producing the similar phenotype seen. The striking similarities seen between the muscular pathology of *ky/ky* mice and humans, reiterates a relationship between gene similarity and function. Specifically, the development of extremely small myofibers with internalised nuclei, alongside large FLNC and Xin aggregates. Although there are some subtle differences between these phenotypes. The mouse phenotype in most cases is more severe, *ky/ky* mice seem to develop scoliosis extremely quickly post weaning, and within the worst cases showing whole body tremors. Unlike the human *KY* myopathies, where mild muscular weakness and muscular atrophy is present, seeming to primarily affect lower limbs, and cause a milder scoliosis which in general developed much later within the disease (Yogev, *et.al.*, 2017; Hedberg-Oldfors, *et.al.*, 2016).

These subtle differences in the myopathies may be attributed to the NLS found within the primate KY. In each of the human KY myopathies, the nonsense mutation occurred within the main body of the KY protein, not affecting the NLS which is located at the N-terminus. These nonsense mutations are unlikely to affect the expression of KY. Both the 7-year-old girl with a nonsense mutation within KY (Hedberg-Oldfors, *et.al.*, 2016) and the 12 Bedouin individuals who had nonsense mutations within exon 1 (Yogev, *et.al.*, 2017), showed high levels of KY mRNA transcripts within muscular organs. Straussberg, *et.al.*, (2016) reported that non-sense mediated decay may be occurring but could not confirm this due to unsuccessful antibodies used within western blots. Finally, no western blots or rt-PCR were performed on the last reported KY myopathy within the Iranian patient who held a nonsense mutation within exon 4 (Ebrahimzadeh, *et.al.*, 2018). This data suggests these KY variants are not being degraded through nonsense mediated decay. If these transcripts are not being degraded it would inevitably result in the production of truncated proteins and would therefore leave the region encoding the NLS. Therefore, this may allow the truncated proteins to localise to the nucleus and perform part of KY's role within primates, and this may explain the minor pathological differences observed between *ky/ky* mice and humans KY myopathies.

6.6 The Nuclear localisation signal conserved within primates

In order for muscular hypertrophy to occur there must be an upregulation of synthetic pathways allowing for protein synthesis (Chen, *et.al.* 2016). Therefore, the hKY could be acting in a similar fashion to other well-defined nucleocytoplasmic proteins. The previously mentioned proteins Ankrd2, and Csrp3 are examples of nucleocytoplasmic shuttles, both of which have been implicated as stretch receptors (Cenni, *et.al.*, 2019; Boateng, *et.al.*, 2009). Both of these proteins localise to the nucleus through the use of a NLS in response to different stressors. Ankrd2 responds to mechanical stimulation and reactive oxygen species (ROS) (Cenni, *et.al.*, 2019). Furthermore Ankrd2, Csrp3, and FLNC were all shown to be upregulated in hypertrophy and regrowth conditions (Chaillou, *et.al.*, 2015).

Another well-defined muscular nucleocytoplasmic protein is the muscle glycogen synthase (MGS), which is known to accumulate and aggregate within the nucleus in times of glucose deprivation. MGS like the hKY also contains NLS which causes a “speckled” nuclear localisation pattern, highly reminiscent of hKY and IGFN1_V1. MGS accumulates and aggregates within nuclear compartments giving the “speckled” appearance. Furthermore, they showed its nuclear localisation was not determined by the enzymatic sites, but through an independent region on the protein. In conclusion the authors suggested that MGS may act as a sensor of cellular energy reserves (Cid, *et.al.*, 2005). These cases highlight that a NLS can be utilised in order to respond to stressors independently of its primary function, providing good evidence that the hKY could be acting as a stress response protein, in this case the stress being stretch. It's plausible to believe that h could be detecting mechanical stress at the Z-disk, a known anchor for nucleocytoplasmic proteins and a site for intracellular signalling (Luther, 2009). Then, in response to mechanical stress, translocate to the nucleus and initiate a transcriptional response.

It is highly intriguing that only primates have evolved to maintain a NLS within their KY protein. There are many significant differences between primates and other placental mammals. The number and complexity of muscles that primates hold differ to other species. Also, primates are one of few species that are able to move regularly within three dimensions. All non-human primates rely on the ability to climb, living, sleeping, and evading predators within the trees. It has been theorised that early in the primate evolution their morphological features and locomotor patterns evolved due to climbing (Hanna and Schmitt, 2011), making primate locomotion very unique. The most distinct locomotion difference being that primates hold a different walking gait pattern (Larson, 1998), where a diagonal-sequence diagonal couplet gait is used. These specific gait characteristics are thought to have evolved due to specific mechanical challenges linked to climbing and the movement on vertical supports (Hanna, *et.al.*, 2017). Furthermore, a study comparing the total energy expenditure from a range of primates and other placental mammals, has shown that on average, primates expend half the predicted amount (Pontzer *et.al.*, 2014). The authors suggested this

represented a “systemic metabolic adaptation for low energy expenditures” which was not explained by differences in physical activity.

Overall, the evidence presented shows primates have evolved to be able to perform a wide range of physical activities, highly different to other mammals. These significant differences may represent the reason as to why primates have evolved to conserve a NLS within KY, potentially answering the reason behind the significant energy expenditure differences between primates and other placental mammals.

6.7 IGFN1 and KY’s Interaction

KY’s interaction with IGFN1 still remains elusive. The fact that *Ky* deficient muscles show muscle atrophy and an inability to grow through muscle hypertrophy, could suggest that KY is able to modulate IGFN1’s ability to promote muscular atrophy. More recent data has shown IGFN1 interacts with the proteasomal subunits (Psm2, 11 and 12) (Cracknell, *et.al.*, 2020), it has been suggested that IGFN1 may be acting as a scaffold for proteasomal subunits at the Z-disk. Non primate KY does not localise to the nucleus, suggesting non primate KY’s interaction with IGFN1 would solely occur within the myoplasm. Potentially IGFN1 could be directed to misfolded proteins at the Z-disk via KY, then utilise interactions with proteasomal subunits to aid their clearance. Thus, if this is true it may explain the wider Z-disk observed within the *ky/ky* pathology, as misfolded proteins may be contained within the Z-disk. Therefore, this could suggest that the IGFN1-FLNC-KY complex may not only be required as a protein scaffold but may also allow for the detection and clearance of misfolded proteins within the Z-disk. Conversely, KY could be protecting proteins from premature degradation, through the formation of complexes.

IGFN1 expression positively correlates with atrophic signalling (Rahimov, *et.al.*, 2011), and has been suggested to down regulate protein synthesis via an inhibitory interaction with eEF1A occurring within the myoplasm (Mansilla, *et.al.*, 2008). Cumulatively, this data provides evidence that IGFN1 may be involved within the inhibition of protein synthesis in atrophic conditions. Therefore, if IGFN1 is inhibiting the induction of proteins synthesis via eEF1A and this pathway is overstimulated this

could prevent muscle fibers from being able to grow via hypertrophy, as new proteins would need to be produced for this to occur. Therefore, if KY is able to bind IGFN1 and prevent the interaction with eEF1A, this could explain the atrophy seen within *ky/ky* muscles. As IGFN1 is not inhibited by KY, this would therefore lead to a constant inhibition placed upon eEF1A.

Intriguingly the mKY has never been seen within the nucleus, always localising to the cytoplasm. Here it was shown that the hKY can localise to the nucleus where it is able to form clusters with IGFN1_V1 and hKY (Figure 24). Therefore, it was asked if overexpression of IGFN1_V1 could cause the mKY to localise to the nucleus. Results indicated no co-localisation of these two proteins (Figure 25) and reiterating the mKY to be a cytoplasmic protein. This data shows that IGFN1 does not induce the nuclear localisation of KY. This distinct lack of colocalization between IGFN1_V1 and mKY was extremely interesting, as they are known interacting partners. Curiously, previous attempts to immunoprecipitated individual complexes of IGFN1_V1 and KY from mouse skeletal muscle extracts were not successful. Although affinity purified KY was able to form complexes with IGFN1_V1 and FLNC from C2C12 myoblast extracts (Baker, *et.al.*, 2010). Furthermore, in this case the authors then suggested that IGFN1, FLNC, and KY form a complex which provides support to the Z-disk. This may explain why an interaction between the mKY and IGFN1_V1 was not seen when co-transfected into C2C12 myoblasts, as their interaction may be exclusive to the Z-disk within a fully differentiated myotube.

6.8 Ky as a regulator of protein turnover

The *ky/ky* mouse and human *KY* myopathies are known to have a prominent pathology. The build-up of large sarcomeric protein aggregates (such as FLNC) is one of these hallmarks, commonly seen within myofibrillar myopathies. Previous work has shown an upregulation of the co-chaperone BAG3 in *Ky* deficient models (Jokl, *et.al.*, 2018). The cochaperone BAG3 is part of the CASA system and allows for the formation of the CASA complex. BAG3 has been reported to be essential for muscle maintenance (Arndt, *et.al.*, 2010) and allows for mechanotransduction (Ulbricht, *et.al.*, 2013).

Furthermore, BAG3 has been suggested to increase the transcription of FLNC through interactions with components of the Hippo system (Ulbricht, *et.al.*, 2013). It has been reported in *ky/ky* mice that a transcriptional upregulation of BAG3 was observed, and within the pathological soleus, there was evidence of impaired BAG3 turnover (Jokl, *et.al.*, 2018).

6.9 Future Work

6.9.1 In-Silico Analysis

Evolutionary analysis on the NLS conserved within primates

Extensive phylogenetic analysis should be undertaken to further understand the evolutionary history of *Ky*. The current research indicates that many species have removed the NLS conserved within primates. Identification of the time point at which this occurred could better aid our understanding as to why this happened, and why primates have specifically conserved this region. Furthermore, identification of the key motifs conserved among species may help to understand the function of each conserved region, identifying a potential region which allows for Z-disk localisation. Using the MEME program (Bailey and Elkan, 1994), predictions of potential patterns within the full amino acid sequence can be acquired. Thus, these can be used to identify key regions which may have a critical role within *KY*'s function. Overall extensive *in-silico* analysis could better aid our understanding of this complex protein, furthering our knowledge of conserved regions. To date only one highly conserved region has been identified this being the TGN/PROT domain.

6.9.2 In-vitro Experiments

Conformation that *KY*'s TGN/PROT domain can aid protein interactions

Here it has been assumed that the TGN/PROT domain mediates protein-protein interactions. In order to confirm this, we could map the interacting domains in the *KY* protein for IGFN1 using a yeast two hybrid.

Validation that the h*KY* is acting as a stress response protein

Therefore, in order to validate that the hKY is responding to mechanical tension, C2C12 myoblasts which have been transfected with hKy-eGFP, should be exposed to osmotic shock through the use of Sorbitol, which has been previously used to induce osmotic shock (Wang, *et.al.*, 1999). Thus, if the hKY is translocating to the nucleus to under mechanical tension, we should observe a higher proportion of hKY localised within the nucleus within conditions of osmotic shock. This would allow us to validate that this NLS is being used to respond to mechanical stress.

6.9.3 *In-vivo* Experiments

Identification of the region within KY allowing for Z-disk localisation

As suggested above the minimal domain required for Z-disk localisation could be present within the C-terminus. Therefore, in order to identify this minimal domain, we can electroporate a series of KY constructs containing a series of deletions across the C-terminus. Then assess the ability fo these constructs to correctly localise to the Z-disk.

Conformation of nuclear interaction between IGFN1_V1 and hKY *in-vivo*

Here it has been shown that the hKY is able to form nuclear clusters with IGFN1_V1 when co-transfected within C2C12 myoblasts, likely representing a nuclear protein interaction between these two proteins. In order to confirm this is not exclusive to myoblasts and can occur in fully differentiated myotubes, both constructs need to be electroporated into the TA of mice. Although, it is likely only a few fibers will contain both constructs, these should have a significant enough expression for valid conclusions to be made.

7. Reference list

- Altschul, S. F. et al. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215 (3), pp.403–410.
- Anantharaman, V., Koonin, E. V. and Aravind, L. (2001). Peptide-N-glycanases and DNA repair proteins, Xp-C/Rad4, are, respectively, active and inactivated enzymes sharing a common transglutaminase fold. *Human molecular genetics*, 10 (16), pp.1627–1630.
- Arndt, V. et al. (2010). Chaperone-assisted selective autophagy is essential for muscle maintenance. *Current biology: CB*, 20 (2), pp.143–148.
- Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings /... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 2, pp.28–36.
- Baker, J. et al. (2010). Identification of a Z-band associated protein complex involving KY, FLNC and IGFN1. *Experimental cell research*, 316 (11), pp.1856–1870.
- Beatham, J. et al. (2004). Filamin C interacts with the muscular dystrophy KY protein and is abnormally distributed in mouse KY deficient muscle fibres. *Human molecular genetics*, 13 (22), pp.2863–2874.
- Biga, L. M. et al. (2020). *Anatomy & physiology*. [Online]. Available at: <https://open.oregonstate.edu/aandp/chapter/10-2-skeletal-muscle/>.
- Bilodeau, P. A., Coyne, E. S. and Wing, S. S. (2016). The ubiquitin proteasome system in atrophying skeletal muscle: roles and regulation. *American journal of physiology. Cell physiology*, 311 (3), pp.C392–C403.
- Blanco, G. et al. (2001). The kyphoscoliosis (ky) mouse is deficient in hypertrophic responses and is caused by a mutation in a novel muscle-specific protein. *Human molecular genetics*, 10 (1), pp.9–16.
- Blanco, G. et al. (2004). Molecular phenotyping of the mouse ky mutant reveals UCP1 upregulation at the neuromuscular junctions of dystrophic soleus muscle.

Neuromuscular disorders: NMD, 14 (3), pp.217–228.

Boateng, S. Y. et al. (2009). Myocyte remodeling in response to hypertrophic stimuli requires nucleocytoplasmic shuttling of muscle LIM protein. *Journal of molecular and cellular cardiology*, 47 (4), pp.426–435.

Bodine, S. C. et al. (2001). Akt/mTOR pathway is a crucial regulator of skeletal muscle hypertrophy and can prevent muscle atrophy in vivo. *Nature cell biology*, 3 (11), pp.1014–1019.

Bonaldo, P. and Sandri, M. (2013). Cellular and molecular mechanisms of muscle atrophy. *Disease models & mechanisms*, 6 (1), pp.25–39.

Bridges, L. R. et al. (1992). The neuromuscular basis of hereditary kyphoscoliosis in the mouse. *Muscle & nerve*, 15 (2), pp.172–179.

Cai, D. et al. (2004). IKKbeta/NF-kappaB activation causes severe muscle wasting in mice. *Cell*, 119 (2), pp.285–298.

Carter, P. and Wells, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature*, 332 (6164), pp.564–568.

Cemazar, M. et al. (2012). Hyaluronidase and collagenase increase the transfection efficiency of gene electrotransfer in various murine tumors. *Human gene therapy*, 23 (1), pp.128–137.

Cenni, V. et al. (2019). Ankrd2 in Mechanotransduction and Oxidative Stress Response in Skeletal Muscle: New Cues for the Pathogenesis of Muscular Laminopathies. *Oxidative medicine and cellular longevity*, 2019, p.7318796.

Chaillou, T. et al. (2015). Identification of a conserved set of upregulated genes in mouse skeletal muscle hypertrophy and regrowth. *Journal of applied physiology*, 118 (1), pp.86–97.

Chen, J. et al. (2019). A mutation in the filamin c gene causes myofibrillar myopathy with lower motor neuron syndrome: a case report. *BMC neurology*, 19 (1), p.198.

Cid, E. et al. (2005). Determinants of the nucleocytoplasmic shuttling of muscle glycogen synthase. *The FEBS journal*, 272 (12), pp.3197–3213.

Clark, K. A. et al. (2002). Striated muscle cytoarchitecture: an intricate web of form and function. *Annual review of cell and developmental biology*, 18, pp.637–706.

Costes, S. V. et al. (2004). Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophysical journal*, 86 (6), pp.3993–4003.

Cracknell, T. et al. (2020). Proteomic resolution of IGFN1 complexes reveals a functional interaction with the actin nucleating protein COBL. *Experimental cell research*, 395 (2), p.112179.

Davey, N. E. (2019). The functional importance of structure in unstructured protein regions. *Current opinion in structural biology*, 56, pp.155–163.

De Larichaudy, J. et al. (2012). TNF- α - and tumor-induced skeletal muscle atrophy involves sphingolipid metabolism. *Skeletal muscle*, 2 (1), p.2.

Dickinson, A. G., and V. M. Meikle. (1973). Genetic kyphoscoliosis in mice. *The Lancet*.

Ebrahimzadeh-Vesal, R. et al. (2018). Identification of a novel nonsense mutation in kyphoscoliosis peptidase gene in an Iranian patient with myofibrillar myopathy. *Genes & diseases*, 5 (4), pp.331–334.

Foss, C. A. et al. (2005). Radiolabeled small-molecule ligands for prostate-specific membrane antigen: in vivo imaging in experimental models of prostate cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 11 (11), pp.4022–4028.

Frank, D. and Frey, N. (2011). Cardiac Z-disc signaling network. *The Journal of biological chemistry*, 286 (12), pp.9897–9904.

Goodman, C. A. and Hornberger, T. A. (2014). New roles for Smad signaling and phosphatidic acid in the regulation of skeletal muscle mass. *F1000prime reports*, 6, p.20.

Griffin, M., Casadio, R. and Bergamini, C. M. (2002). Transglutaminases: nature's biological glues. *Biochemical Journal*, 368 (Pt 2), pp.377–396.

Guilbert, S. M. et al. (2018). HSPB8 and BAG3 cooperate to promote spatial sequestration of ubiquitinated proteins and coordinate the cellular adaptive response to proteasome insufficiency. *FASEB journal: official publication of the Federation of*

American Societies for Experimental Biology, 32 (7), pp.3518–3535.

Hall, M. N., Corbett, A. H. and Pavlath, G. K. (2011). Regulation of nucleocytoplasmic transport in skeletal muscle. *Current topics in developmental biology*, 96, pp.273–302.

Hall, Z. W. and Ralston, E. (1989). Nuclear domains in muscle cells. *Cell*, 59 (5), pp.771–772.

Haller, R. G. and DiMauro, S. (2012). Chapter 75 - Metabolic and Mitochondrial Myopathies. In: Hill, J. A. and Olson, E. N. (Eds). *Muscle*. Boston/Waltham: Academic Press. pp.1031–1041.

Hanna, J. B. et al. (2017). The evolution of vertical climbing in primates: evidence from reaction forces. *The Journal of experimental biology*, 220 (Pt 17), pp.3039–3052.

Hanna, J. B. and Schmitt, D. (2011). Locomotor energetics in primates: gait mechanics and their relationship to the energetics of vertical and horizontal locomotion. *American journal of physical anthropology*, 145 (1), pp.43–54.

Hedberg-Oldfors, C. et al. (2016). A new early-onset neuromuscular disorder associated with kyphoscoliosis peptidase (KY) deficiency. *European journal of human genetics: EJHG*, 24 (12), pp.1771–1777.

Hehenberger, E. (2018). *Phylogenetic trees in newick format*. [Online]. Available at: doi:10.5061/dryad.26bv4/11 [Accessed 13 December 2020].

Ho, R. C. et al. (2005). Regulation of I κ B kinase and NF- κ B in contracting adult rat skeletal muscle. *American Journal of Physiology-Cell Physiology*, 289 (4), pp.C794–C801.

Holbrook, J. J. et al. (1975). 4 Lactate Dehydrogenase. In: Boyer, P. D. (Ed). *The Enzymes*. 11. Academic Press. pp.191–292.

Jokl, E. J. et al. (2018). Transcriptional upregulation of Bag3, a chaperone-assisted selective autophagy factor, in animal models of KY-deficient hereditary myopathy. *Disease models & mechanisms*, 11 (7). [Online]. Available at: doi:10.1242/dmm.033225.

- Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences: CABIOS*, 8 (3), pp.275–282.
- Kadmas, J. L. and Beckerle, M. C. (2004). The LIM domain: from the cytoskeleton to the nucleus. *Nature reviews. Molecular cell biology*, 5 (11), pp.920–931.
- Klont, R. E., Brocks, L. and Eikelenboom, G. (1998). Muscle fibre type and meat quality. *Meat science*, 49S1, pp.S219–S229.
- Knöll, R., Buyandelger, B. and Lab, M. (2011). The sarcomeric Z-disc and Z-discopathies. *Journal of biomedicine & biotechnology*, 2011, p.569628.
- Kumar, S. et al. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution*, 35 (6), pp.1547–1549.
- Larson, S. G. (1998). Unique Aspects of Quadrupedal Locomotion in Nonhuman Primates. In: Strasser, E. et al. (Eds). *Primate Locomotion: Recent Advances*. Boston, MA: Springer US. pp.157–173.
- Lasich, C. (2014). *Elevated CPK and Other Diagnostic Clues to Muscle Pain*. [Online]. Available at: <https://www.healthcentral.com/article/elevated-cpk-and-other-diagnostic-clues-to-muscle-pain> [Accessed 3 December 2020].
- Leber, Y. et al. (2016). Filamin C is a highly dynamic protein associated with fast repair of myofibrillar microdamage. *Human molecular genetics*, 25 (13), pp.2776–2788.
- Luther, P. K. et al. (2003). Heterogeneity of Z-band structure within a single muscle sarcomere: implications for sarcomere assembly. *Journal of molecular biology*, 332 (1), pp.161–169.
- Luther, P. K. (2009). The vertebrate muscle Z-disc: sarcomere anchor for structure and signalling. *Journal of muscle research and cell motility*, 30 (5-6), pp.171–185.
- Makarova, K. S., Aravind, L. and Koonin, E. V. (1999). A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. *Protein science: a publication of the Protein Society*, 8 (8), pp.1714–1719.
- Mansilla, F. et al. (2008). Translation elongation factor eEF1A binds to a novel

- myosin binding protein-C-like protein. *Journal of cellular biochemistry*, 105 (3), pp.847–858.
- Maréchal, G. et al. (1996). Isoforms of myosin in growing muscles of kyphoscoliotic mice. *European journal of biochemistry / FEBS*, 241 (3), pp.916–922.
- Masiero, E. and Sandri, M. (2010). Autophagy inhibition induces atrophy and myopathy in adult skeletal muscles. *Autophagy*, 6 (2), pp.307–309.
- McPherron, A. C., Lawler, A. M. and Lee, S. J. (1997). Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature*, 387 (6628), pp.83–90.
- Morissette, M. R. et al. (2009). Effects of myostatin deletion in aging mice. *Aging cell*, 8 (5), pp.573–583.
- Morton, J. P. et al. (2009). The exercise-induced stress response of skeletal muscle, with specific emphasis on humans. *Sports medicine*, 39 (8), pp.643–662.
- Musarò, A. et al. (1999). IGF-1 induces skeletal myocyte hypertrophy through calcineurin in association with GATA-2 and NF-ATc1. *Nature*, 400 (6744), pp.581–585.
- Nguyen Ba AN, Pogoutse A, Provar N, Moses AM. (2009). *NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction*. [Online]. Available at: <http://www.moseslab.csb.utoronto.ca/NLStradamus/> [Accessed November 2019].
- Nilsson, M. I. et al. (2013). Xin is a marker of skeletal muscle damage severity in myopathies. *The American journal of pathology*, 183 (6), pp.1703–1709.
- Palmio, J. and Udd, B. (2016). Myofibrillar and distal myopathies. *Revue neurologique*, 172 (10), pp.587–593.
- Panzhinskiy, E. et al. (2013). Chapter 22 - Role of Mammalian Target of Rapamycin (mTOR) in Muscle Growth. In: Bagchi, D., Nair, S. and Sen, C. K. (Eds). *Nutrition and Enhanced Sports Performance*. San Diego: Academic Press. pp.217–227.
- Pavlat, G. K. et al. (1989). Localization of muscle gene products in nuclear domains. *Nature*, 337 (6207), pp.570–573.

Perillo, M. and Folker, E. S. (2018). Specialized Positioning of Myonuclei Near Cell-Cell Junctions. *Frontiers in physiology*, 9, p.1531.

Pocock, G. et al. (2013). *Human Physiology*. OUP Oxford.

Pontzer, H. et al. (2014). Primate energy expenditure and life history. *Proceedings of the National Academy of Sciences of the United States of America*, 111 (4), pp.1433–1437.

Rahimov, F. et al. (2011). Gene expression profiling of skeletal muscles treated with a soluble activin type IIB receptor. *Physiological genomics*, 43 (8), pp.398–407.

Rashid, M. M. et al. (2015). Muscle Lim Protein (MLP)/CSRP3 at the crossroad between mechanotransduction and autophagy. *Cell death & disease*, 6, p.e1940.

Rebbapragada, A. et al. (2003). Myostatin signals through a transforming growth factor beta-like signaling pathway to block adipogenesis. *Molecular and cellular biology*, 23 (20), pp.7230–7242.

Rommel, C. et al. (2001). Mediation of IGF-1-induced skeletal myotube hypertrophy by PI(3)K/Akt/mTOR and PI(3)K/Akt/GSK3 pathways. *Nature cell biology*, 3 (11), pp.1009–1013.

Ruparelia, A. A. et al. (2016). FLNC myofibrillar myopathy results from impaired autophagy and protein insufficiency. *Human molecular genetics*, 25 (11), pp.2131–2142.

Ruparelia, A., Vaz, R. R. and Bryson-Richardson, R. J. (2012). Myofibrillar myopathies and the Z-disk associated proteins. In: *Skeletal Muscle-From Myogenesis to Clinical Relations*. In-Tech. pp.317–358. [Accessed 16 November 2020].

Sacco, A. et al. (2010). Short telomeres and stem cell exhaustion model Duchenne muscular dystrophy in mdx/mTR mice. *Cell*, 143 (7), pp.1059–1071.

Sandri, M. et al. (2004). Foxo transcription factors induce the atrophy-related ubiquitin ligase atrogin-1 and cause skeletal muscle atrophy. *Cell*, 117 (3), pp.399–412.

Sanger, J. W. et al. (2017). Assembly and Maintenance of Myofibrils in Striated

Muscle. In: Jockusch, B. M. (Ed). *The Actin Cytoskeleton*. Cham: Springer International Publishing. pp.39–75.

Schoenfeld, B. J. (2010). The mechanisms of muscle hypertrophy and their application to resistance training. *Journal of strength and conditioning research / National Strength & Conditioning Association*, 24 (10), pp.2857–2872.

Schoenfeld, B. J. (2013). Potential mechanisms for a role of metabolic stress in hypertrophic adaptations to resistance training. *Sports medicine*, 43 (3), pp.179–194.

Scott, W., Stevens, J. and Binder-Macleod, S. A. (2001). Human skeletal muscle fiber type classifications. *Physical therapy*, 81 (11), pp.1810–1816.

Selcen, D. and Engel, A. G. (2011). Chapter 11 - Myofibrillar myopathies. In: Griggs, R. C. and Amato, A. A. (Eds). *Handbook of Clinical Neurology*. 101. Elsevier. pp.143–154.

Simoneau, J. A. and Bouchard, C. (1989). Human variation in skeletal muscle fiber-type proportion and enzyme activities. *The American journal of physiology*, 257 (4 Pt 1), pp.E567–E572.

Stitt, T. N. et al. (2004). The IGF-1/PI3K/Akt pathway prevents expression of muscle atrophy-induced ubiquitin ligases by inhibiting FOXO transcription factors. *Molecular cell*, 14 (3), pp.395–403.

Straussberg, R. et al. (2016). Kyphoscoliosis peptidase (KY) mutation causes a novel congenital myopathy with core targetoid defects. *Acta neuropathologica*, 132 (3), pp.475–478.

Sun, X. et al. (2002). Expression and binding activity of the glucocorticoid receptor are upregulated in septic muscle. *American journal of physiology. Regulatory, integrative and comparative physiology*, 282 (2), pp.R509–R518.

Tasca, G. et al. (2012). Novel FLNC mutation in a patient with myofibrillar myopathy in combination with late-onset cerebellar ataxia. *Muscle & nerve*, 46 (2), pp.275–282.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids*

research, 22 (22), pp.4673–4680.

Ulbricht, A., Arndt, V. and Höhfeld, J. (2013). Chaperone-assisted proteostasis is essential for mechanotransduction in mammalian cells. *Communicative & integrative biology*, 6 (4), p.e24925.

Ullrick, W. C. (1967). A theory of contraction for striated muscle. *Journal of theoretical biology*, 15 (1), pp.53–69.

Vafiadaki, E., Arvanitis, D. A. and Sanoudou, D. (2015). Muscle LIM Protein: Master regulator of cardiac and skeletal muscle functions. *Gene*, 566 (1), pp.1–7.

Vierck, J. et al. (2000). Satellite cell regulation following myotrauma caused by resistance exercise. *Cell biology international*, 24 (5), pp.263–272.

Wackerhage, H. et al. (2019). Stimuli and sensors that initiate skeletal muscle hypertrophy following resistance exercise. *Journal of applied physiology*, 126 (1), pp.30–43.

Wang, D.-T. et al. (2015). Myostatin Activates the Ubiquitin-Proteasome and Autophagy-Lysosome Systems Contributing to Muscle Wasting in Chronic Kidney Disease. *Oxidative medicine and cellular longevity*, 2015, p.684965.

Wang, H. L. et al. (1999). Effect of sorbitol induced osmotic stress on the changes of carbohydrate and free amino acid pools in sweet potato cell suspension cultures. *Botanical bulletin of Academia Sinica. New series. Zhong yang yan jiu yuan. Zhi wu yan jiu suo*. [Online]. Available at:

https://www.researchgate.net/profile/Heng_Long_Wang/publication/281222708_Effect_of_sorbitol_induced_osmotic_stress_on_the_changes_of_carbohydrate_and_free_amino_acid_pools_in_sweet_potato_cell_suspension_cultures/links/571aeb9208ae6eb94d0c803c.pdf.

Wang, J. et al. (2005). Dynamics of Z-band based proteins in developing skeletal muscle cells. *Cell motility and the cytoskeleton*, 61 (1), pp.34–48.

Welle S, Brooks AI, Delehanty JM, Needler N et al. (2003). Muscle function and aging - male. *Muscle function and aging - male*. [Online]. Available at: <https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS288>.

White, J. P. et al. (2009). Overload-induced skeletal muscle extracellular matrix

remodelling and myofibre growth in mice lacking IL-6. *Acta physiologica* , 197 (4), pp.321–332.

Wilkins, J. T. et al. (2001). Contractile properties of adjacent segments of single human muscle fibers. *Muscle & nerve*, 24 (10), pp.1319–1326.

Yaron, A. et al. (1998). Identification of the receptor component of the IkappaBalpha-ubiquitin ligase. *Nature*, 396 (6711), pp.590–594.

Yogev, Y. et al. (2017). Progressive hereditary spastic paraplegia caused by a homozygous KY mutation. *European journal of human genetics: EJHG*, 25 (8), pp.966–972.

Zhang, G. and Li, Y.-P. (2012). p38 β MAPK upregulates atrogin1/MAFbx by specific phosphorylation of C/EBP β . *Skeletal muscle*, 2 (1), p.20.

8. Appendix

Placing reflex.

The placing reflex was commonly used to allow for the phenotyping of Ky/Ky mice. As Ky/Ky mice are known to have weak postural muscles, they are unable to reach out and grab a ledge in front of them.

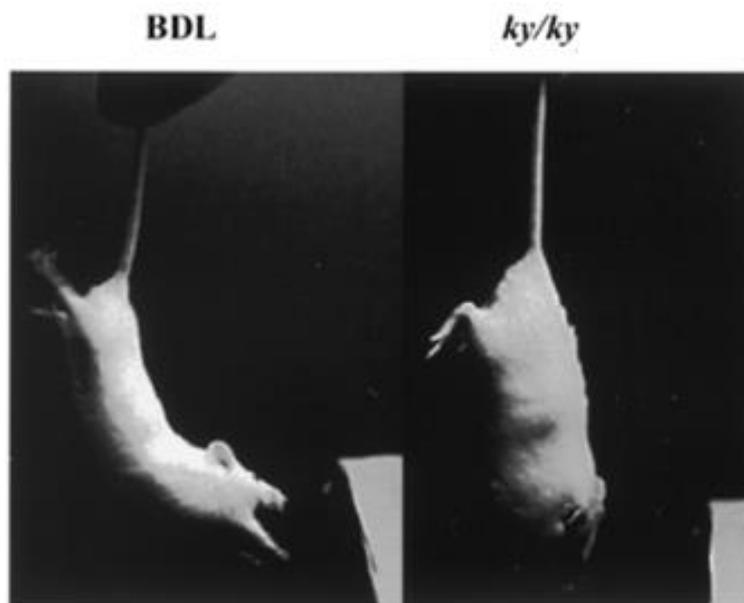


Figure 1. Placing reflex. Image taken from (Blanco, et.al., 2001). The BDL (Wildtype) strain of mice show a normal placing reflex, where mice will reach out towards a ledge. The Ky/Ky mouse shows an inability to reach towards the ledge due to weak paraspinal muscles (postural muscles).

In-silico Analysis of the NLS within hKy-X5.

Two NLS signal predicting software's were used to identify the NLS with hKy X5. Firstly, NLStradamus was used and gave us the readout below (Figure 2). The software showed us the motif most likely to cause nuclear localisation. This software is available at <http://www.moseslab.csb.utoronto.ca/NLStradamus/>

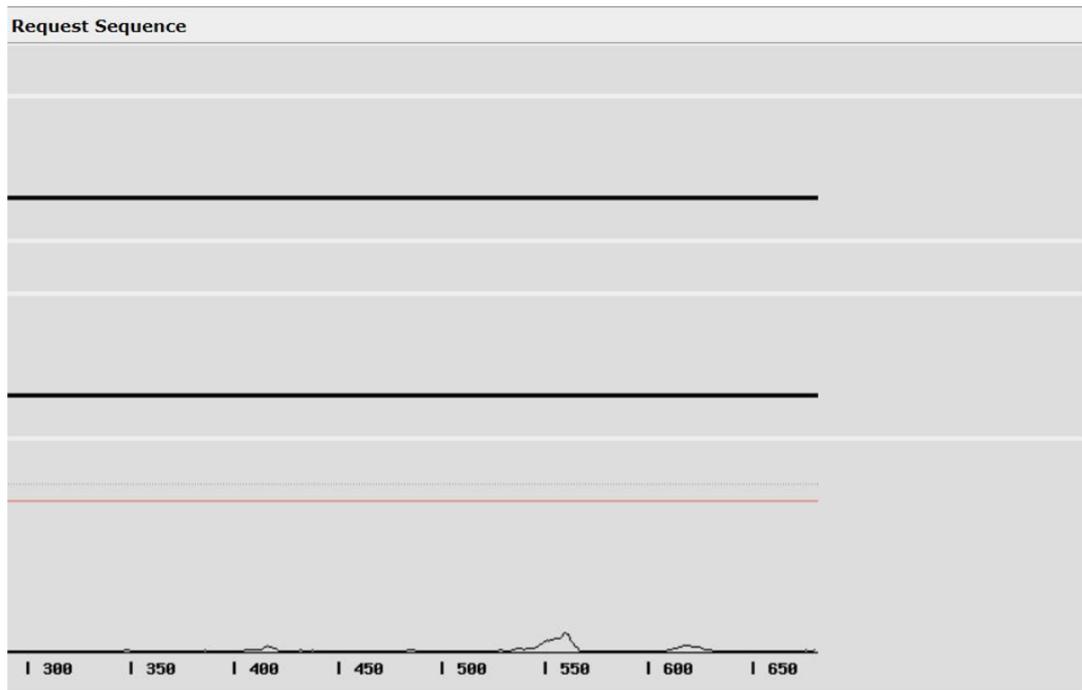
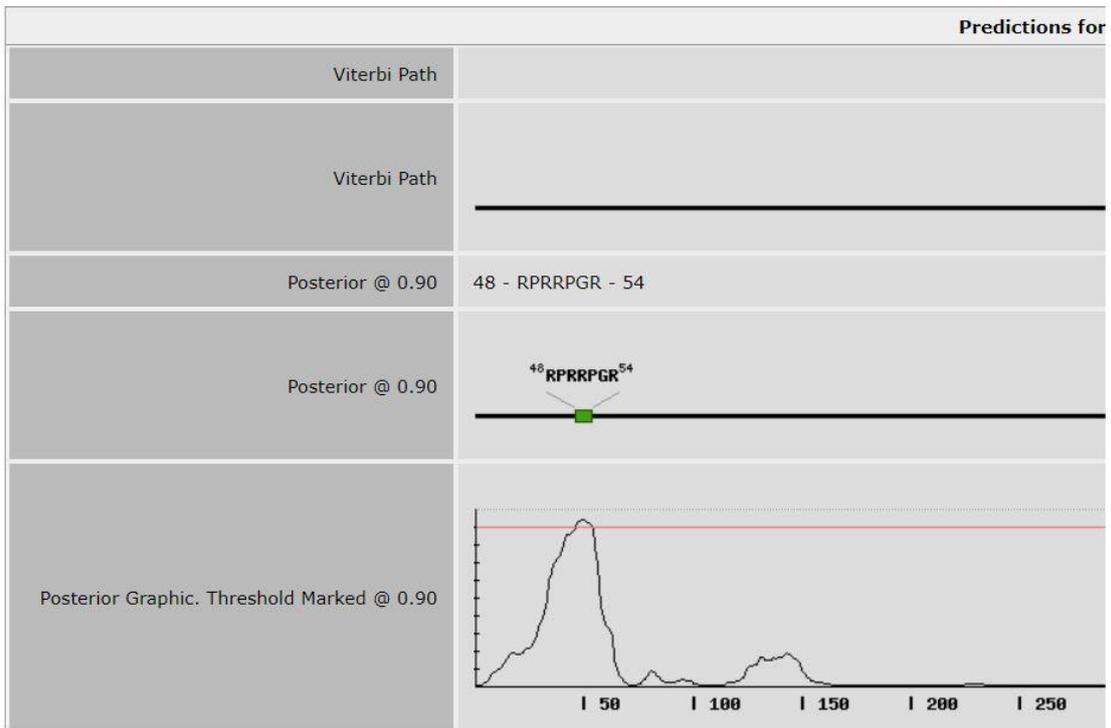


Figure 2. NLS software results. A cut of score of 0.9 was applied, as seen on the graphic. This gave us the motif that was most likely to cause nuclear localisation.

Muscle sections from Ky/Ky patients

Patients with loss of function Ky mutations have an extremely prevalent phenotype. The muscular pathology presenting with heterogenous fiber sizes, including extremely small myofibers. Also seen is internalised nuclei.

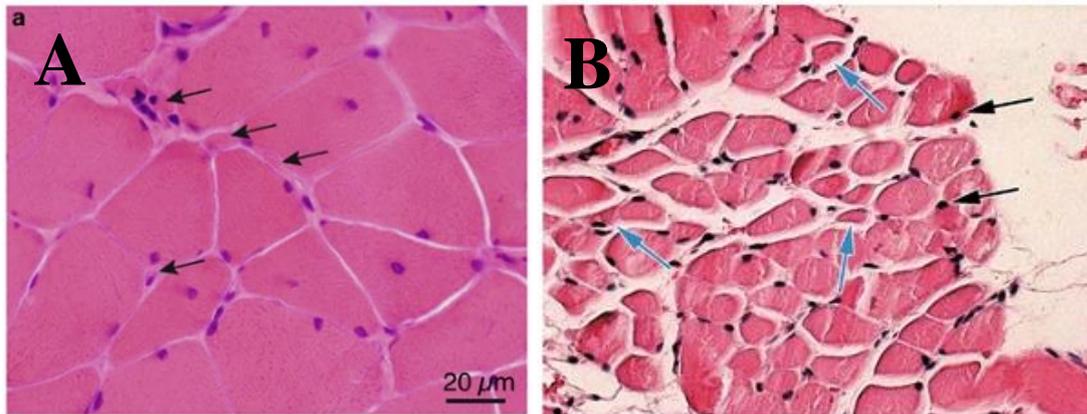


Figure 3. Heterogenous fibers and internalised myonuclei within Ky loss of function patients. Both patients present with extremely small myofibers alongside internalised nuclei. **A)** Image taken from Herdberg-Oldfors, et.al., 2016. Available at: <https://www.nature.com/articles/ejhg201698>. **B)** Image taken from Yogev, et.al., 2017. Available at: <https://www.nature.com/articles/ejhg201785>