Curtin Medical School

Investigating Genetic Causes of Mendelian Congenital Myopathies

Lein Dofash

0000-0002-4381-1909

This thesis is presented for the Degree of
Master of Research
of
Curtin University

January 2022

Declaration

To the best of my knowledge, this thesis contains no material previously published by any other person except where due acknowledgment has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Human Ethics

This study received human ethics approval by the University of Western Australia Human Research Ethics Committee (RA/4/20/1008) and the Curtin University Human Research Ethics Office (HRE2019-0566).

Signature

Date **19/12/21**

Abstract

Congenital myopathies are a clinically and genetically heterogeneous group of rare neuromuscular disorders. Patients typically suffer from early-onset muscle dysfunction and weakness which can result in severe morbidity and early mortality. Identification of the genetic basis of disease enables an early and accurate diagnosis, which facilitates appropriate genetic counselling, family planning, and where available, treatments to improve patient outcomes and quality of life. Over 40 genes have been associated with congenital myopathies. However, the genetic basis for ~50% of affected patients remains unresolved. It is expected that massively parallel sequencing and functional genomics can facilitate identification of novel variants and genes associated with congenital myopathy. In this study, whole exome sequencing data was analysed from a cohort of patients with genetically unsolved congenital myopathies. For two families, two genetic candidates were identified and further characterised using a range of functional assays.

Firstly, in a Spanish patient with genetically unresolved nemaline myopathy, we identified biallelic compound heterozygous variants in the kelch-like family member 40 gene (*KLHL40*). These include a single nucleotide variant (c.*152G>T) in the 3' untranslated region (3' UTR) *in trans* with a 10.9 kb deletion spanning exons 2-6 of *KLHL40*. RNA sequencing and western blotting of patient muscle biopsy respectively revealed significantly lower *KLHL40* transcript and protein levels in the patient compared to controls. *In silico* tools SpliceAI and Introme predicted the c.*152G>T variant creates a donor splice site, which was confirmed by *de novo* alignment of patient RNA sequencing data. Our data suggests that the c.*152G>T variant likely results in 3' UTR splice-mediated nonsense mediated decay, a mechanism that may be underrecognised in Mendelian disease.

Secondly, in two Spanish siblings with severe rigid spine syndrome, we identified a homozygous missense variant (p.Ser447Pro) in the 3-hydroxy-3-methylglutaryl-coenzyme A synthase gene (*HMGCS1*), encoding a key enzyme in the mevalonate pathway. Through international collaborations and matchmaking platforms, we subsequently identified two additional families with unsolved rigid spine syndrome and biallelic variants in *HMGCS1*. Whilst *HMGCS1* has not been associated with

disease in human to date, myopathic manifestations have been associated with defects of the mevalonate pathway. We showed that *HMGCS1* is expressed, and protein is abundant in skeletal muscle and muscle-derived cells compared to some tissues and cells. *Hmgcs1* null mice are embryonic lethal, thus we hypothesise that *HMGCS1* variants identified in these patients likely result in partial loss-of-function. Using size exclusion chromatography, circular dichroism, and enzymatic assays respectively, I investigated whether the p.(Ser447Pro) variant disrupts the dimerisation, secondary structure, thermal stability, and enzyme activity of recombinantly produced wildtype and mutant HMGCS1. I show that HMGCS1^{S447P} maintains its 3D structure, dimerized state and retains ~92% of the wildtype's thermal stability. However, HMGCS1^{S447P} activity appeared to be moderately dysregulated and retained ~56% of the wildtype's activity when saturated with acetoacetyl-CoA. Whether this difference in *HMGCS1* function is sufficient to account for the phenotype observed in our patients remains unknown. It is possible that defects in other, currently unknown function/s of HMGCS1 may also contribute to disease seen in our patients. Ongoing analyses include recombinant investigations of the other HMGCS1 substitutions identified in additional families as well as phenotyping *hmgcs1* null zebrafish models.

Overall, this study highlights the utility of whole exome sequencing in combination with RNA sequencing, clinical data sharing and functional genomics when investigating genetic causes of congenital myopathies. This thesis encourages the investigation of 3' UTR variants in known congenital myopathy genes and highlights the critical role played by the mevalonate pathway in skeletal muscle function and disease.

Acknowledgements

I would like to thank a number of people who have supported me throughout my MRes.

Foremost, I thank my primary supervisor Dr. Gina Ravenscroft for this rewarding research experience. I am grateful for all her guidance, insight, time, and constructive criticism throughout the years. I thank her for our regular meetings, for pushing me to think critically and enhance my writing through *one pagers*, and for all the feedback she has provided at phenomenal speed. I also thank her for entrusting me to implement new assays, encouraging me to participate in conferences and for providing opportunities to collaborate with researchers and clinicians worldwide.

I am also incredibly grateful for the support of my co-supervisors at the Perkins. Special thanks to Dr. Josh Clayton for everything he taught me in the lab, for all his troubleshooting suggestions, for his thorough responses to any little or big questions I had, and for all his feedback throughout the years. My courage to explore and implement new techniques in the lab has grown immensely thanks to his moral support. I extend my appreciation to Dr. Rhonda Taylor for all her input over the years, including analysing qPCR data, writing, presenting, and defending my research. I am also grateful for Prof. Nigel Laing for all his involvement and input over the years, as well as for his inspiring anecdotes and valuable chats during the occasional Saturday afternoons when I came in for lab work.

I also extend my deep gratitude to my supervisory team at Curtin. I thank my primary supervisor, A/Prof. David Groth for all his support and encouragement throughout my degrees. For our coffee meet ups and our interesting discussions about statins, mitochondrial disorders, spider naming and human history. I also thank my co-supervisor Dr. Danielle Dye and my Chair Dr. Carl Mousley for their involvement and moral support. Thank you for reminding me of my potential and for believing in me.

I would like to acknowledge and thank all co-authors and collaborators for their invaluable input, contributions, and support throughout my master's.

A special thanks goes to Dr. Joel Haywood, Prof. Joshua Mylne and all members of the Mylne laboratory for the great times I have had while producing recombinant

protein at Bayliss. I am especially grateful for Joel for everything he taught me about recombinant protein production, for his mentorship and for our interesting conversations about HMGCS1 interactions.

To the RAVING students and RA's, thank you for all the laughs and good times we have had. For our regular Lion's eye coffee and Presotea breaks. For the weekly quizzes, the tennis nights, the Friday dress ups and more. Special mentions to Elina who kept me company while we pursued lab work during odd times of the night. Special mentions also go to my friends Shahama, Hanane, Maria, and Somayra. Thank you for being by my side during this thrilling rollercoaster ride.

To my dear family, I could not have come this far without your unconditional support. My deepest gratitude goes to my beloved grandparents (Nahla Nayfeh, Ahmad Daour, Sobheye Marei, and Husni Dofash). I thank them for their continuous prayers, for routinely checking in with my research and inspiring me to pursue my dreams. I am profoundly grateful to my dear parents (Abeer Daour and Nedal Dofash), and my siblings (Sara and Maen). I thank them for their endless encouragement, for cheering me on during the lows and for celebrating the highs. I also wish to acknowledge my great uncle, Prof. Ali Nayfeh who was an inspiring figure for me, who emphasised the value of knowledge despite hardships. I thank him and my grandpa Ahmad for empowering me to pursue molecular genetic research.

Acknowledgement of Country

We acknowledge that Curtin University works across hundreds of traditional lands and custodial groups in Australia, and with First Nations people around the globe. We wish to pay our deepest respects to their ancestors and members of their communities, past, present, and to their emerging leaders.

Table of Contents

List of figures

Supplementary figures

List of tables

Supplementary tables

Abbreviations

ACMG-AMP	American College of Medical Genetics and Genomics and Association of Molecular Pathology
AUC	analytical ultracentrifugation
CADD	combined annotation dependent depletion
CCD	central core disease
CFTD	Congenital myopathy with fibre type disproportion
CNM	centronuclear myopathy
FANTOM5	Functional Annotation of the Mammalian genome consortium
gnomAD	Genome Aggregation Database
GTEx	Genotype-Tissue Expression portal
HEK	human embryonic kidney
IVCT	*in vitro* contracture testing
LoF	loss of function
LOVD	Leiden Open Variation Database
MmD	multiminicore disease
MRI	magnetic resonance imaging
NEM	nemaline myopathy
NGS	next generation sequencing
OMIM	Online Mendelian Inheritance in Man
PCR	polymerase chain reaction
PDB	Protein Data Bank
pLI	probability of loss of function intolerance
POLYPHEN	Polymorphism Phenotyping
qPCR	quantitative PCR
RT-PCR	reverse transcriptase PCR
SIFT	Sorting Intolerant from Tolerant
UTR	untranslated region
VUS	variant of unknown significance
WES	whole exome sequencing
WGS	whole genome sequencing

Chapter 1

Introduction and literature review of Mendelian congenital myopathies

1.1 Introduction

Congenital myopathies are a clinically, histologically, and genetically diverse group of early-onset neuromuscular disorders primarily affecting skeletal muscle[1]. Classical manifestations include early-onset muscle dysfunction, hypotonia and weakness with static or slowly progressive disease progression[2]. Affected individuals may require chronic care, including feeding and breathing support[2]. A majority of these disorders are incurable and can result in death during the first years of life[2,3].

Most congenital myopathies are monogenic in origin, arising from pathogenic monoallelic or biallelic variants in genes critical for normal muscle structure and function[4,5]. For the purposes of this thesis, we focus on the Mendelian congenital myopathies associated with autosomal dominant, autosomal recessive, and X-linked inheritance[2]. Over 40 genes have been implicated in Mendelian congenital myopathies[6], many of which can cause multiple congenital myopathy subtypes and/or other neuromuscular disorders[7]. Knowledge of the underlying genetic cause enables an accurate diagnosis which can benefit families by facilitating appropriate genetic counselling, family planning, and preventing disease recurrence[3,8,9]. In addition, a genetic diagnosis can significantly improve patient outcomes by improved disease management and available therapies[10,11].

Applications of next generation sequencing (NGS), such as whole exome sequencing (WES) represent a non-invasive method for molecular disease diagnosis and gene discovery in clinical and research settings[3,12]. However, ~50% of affected patients who are screened by NGS do not receive a genetic diagnosis for their congenital myopathy[13]. We postulate their disorders may be caused by variants currently classified as of unknown clinical significance, or by novel congenital myopathy genes that are yet to be identified and validated[5].

In this study, we used a combined WES and functional genomics approach to investigate genetic causes of congenital myopathies in patients who remained without a genetic diagnosis following diagnostic screening of known neuromuscular disease genes. In this chapter, I provide a literature review of congenital myopathies, focussing on the subtypes observed in this study. In addition, I discuss the utility of NGS and functional genomics in congenital myopathy diagnosis and disease gene discovery.

1.2 Congenital myopathy subtypes

Congenital myopathies are clinically diagnosed on the basis of structural features identified in skeletal muscle biopsy[1]. There are four classical congenital myopathy subtypes highlighted in the literature which account for a large proportion of congenital myopathies[2]. These include: 1) nemaline myopathy, 2) core myopathy, 3) centronuclear myopathy, and 4) congenital myopathy with fibre type disproportion[2].

1.2.1 Nemaline myopathy

Nemaline myopathies (NEM) are diseases generally caused by dysfunction of the skeletal muscle thin filament[2]. Histologically, NEM are characterised by rod-like inclusions or bodies within the sarcoplasm of myofibres[14]. In rarer instances, they may be present within myonuclei[2]. The latter manifestation is termed intranuclear rod myopathy[15]. Nemaline bodies can be readily detected by electron microscopy of patient muscle biopsy and in some instances, they may be visible with light microscopy and Gomori trichome staining[16]. The localisation of the rods in some cases appears to correlate with the causative gene[16].

To date, 13 genes have been associated with NEM (Table 1.1)[7]. The majority of these genes encode proteins comprising the sarcomeric thin filament, as well as BTB-kelch proteins. Severity of NEM generally depends on the gene affected and ranges from mild and non-progressive manifestations to severe and fatal myopathy[2]. *KLHL40* and *LMOD3* in particular appear associated with severe nemaline myopathy with arthrogryposis, hypokinesia, bulbar and respiratory insufficiency and in most cases lead to neonatal death[17-19]. There are several subtypes of NEM which follow a numbered classification (NEM1-NEM11) corresponding to the implicated gene (Table 1.1), though this classification has not yet been updated to include MYO18B and RYR3-related nemaline myopathies.

Table 1.1 Genes associated with nemaline myopathies[7]

Gene	Protein	Disease	Inheritance	OMIM#
TPM3	Slow α-tropomyosin	NEM1	AD, AR	609284
NEB	Nebulin	NEM2	AR	256030
ACTA1	Skeletal muscle α-actin	NEM3	AD, AR	161800

TPM2	β-tropomyosin	NEM4	AD	609285
TNNT1	Slow troponin T	NEM5	AR	605355
KBTBD13	Kelch-repeat and BTB-[POZ]-domain containing 13	NEM6	AD	609273
CFL2	Cofilin 2	NEM7	AR	610687
KLHL40	Kelch-like family member 40	NEM8	AR	615340
KLHL41	Kelch-like family member 41	NEM9	AR	607701
LMOD3	Leiomodin 3	NEM10	AR	616112
MYPN	Myopallidin	NEM11	AR	617336
MYO18B[20]	Myosin XVIIIB	-	AR	-
RYR3[21]	Ryanodine receptor 3	-	AR	-

Abbreviations include: NEM, nemaline myopathy; AD, autosomal dominant; AR, autosomal recessive.

1.2.2 Core myopathy

Core myopathies are characterised by the presence of single or multiple, structured, or unstructured cores that can be accompanied by myofibrillar disorganisation[1,22]. Cores appear negative with oxidative enzyme staining and reflect areas that are devoid of mitochondria and glycogen[22]. The two classical and most common groups include central core disease (CCD) and multiminicore disease (MmD)[1].

Central core disease is classically characterised by large, well-defined cores within the centre of myofibres[2]. The predominant genetic cause of CCD are biallelic variants in *RYR1* (Table 1.2), though this gene has also been implicated in MmD as well as in malignant hyperthermia and other non-core related myopathies[23].

Multiminicore disease is characterised by multiple, small, and often unstructured cores which can give muscle a "moth-eaten" appearance when analysed by oxidative enzyme staining[1]. The majority of classical MmD cases identified thus far are associated with biallelic variants in *SELENON* (also known as *SEPN1*), followed by biallelic variants in *RYR1*[4]. As with *RYR1*, there is also considerable overlap between the SELENON-related myopathies, which include rigid spine muscular dystrophy (OMIM# 602771) and congenital fibre-type disproportion (OMIM# 255310)[24-26].

Altogether, over 10 genes have been associated with core myopathy (Table 1.2)[2,7]. However, it is estimated that the genetic basis for ~50% of core myopathies remains to be identified[1].

Table 1.2 Genes associated with core myopathies

Gene	Protein	Disease	Inheritance	OMIM#
ACADS	Acyl-CoA Dehydrogenase Short Chain	MmD	AR	201470
ACTA1	Skeletal muscle α-actin	Core-rod	AD, AR	161800
CACNA1S[27]	Calcium channel, voltage dependent, L type, alpha 1S subunit	Core-like	AD	114208
CFL2	Cofilin 2	MmD	AR	610687
FXR1	FMR1 Autosomal Homolog 1	MmD	AR	618823
MEGF10	Multiple EGF-like domains 10	MmD	AR?	614399
MYH2	Myosin, heavy chain 2	MmD	AD, AR	605637
MYH7	Myosin, heavy chain 7	CCD	AD	160760
RYR1	Ryanodine receptor 1	CCD, MmD	AD, AR	117000
SELENON	Selenoprotein N	MmD	AR	606210
TTN	Titin	CCD	AR	611705

Abbreviations include: CCD, central core disease; MmD, multiminicore disease; AD, autosomal dominant; AR, autosomal recessive.

1.2.3 Centronuclear myopathy

Centronuclear myopathies are characterised by increased numbers of nuclei located internally or centrally within myofibres. At least seven genes are currently implicated in CNM (Table 1.3). However ~30% of patients with CNM remain without a genetic diagnosis suggesting that additional CNM genes are yet to be identified[1].

As with NEM, the classification for most CNM subtypes follows a numbered nosology (CNM1-CNM6) corresponding to the causal gene (Table 1.3). However, the classification does not include all CNM-related genes for example *RYR1* and *TTN*, which are associated with other histopathological abnormalities such as cores and fibre

type disproportion[28].

Table 1.3 Genes associated with centronuclear myopathies

Gene	Protein	Disease	Inheritance	OMIM#
MTM1	myotubularin	CNMX	X-linked	310400 300415
DNM2	Dynamin 2	CNM1	AD	160150
MTMR14	Myotubularin related protein 14	CNM1, modifier	AD	160150
BIN1	Bridging integrator 1	CNM2	AD, AR	601248
CCDC78	Coiled coil domain containing protein 78	CNM4	AD	614666
SPEG	Striated muscle preferentially expressed protein kinase	CNM5	AR	615959
ZAK	Sterile alpha motif and leucine zipper containing kinase AZK	CNM6	AR	617760
RYR1	Ryanodine receptor 1	-	AD, AR	255320
TTN	Titin	-	AR	611705

Abbreviations include: CNM, centronuclear myopathy; AD, autosomal dominant; AR, autosomal recessive.

1.2.4 Congenital myopathy with fibre type disproportion

Congenital myopathies with fibre type disproportion (CFTD) are defined by the predominance of disproportionally smaller sized type 1 (slow) myofibres compared to type 2 (fast) myofibres[29]. At least 10 genes have been implicated in CFTD (Table 1.4)[2]. The expression of some of these genes for example, *TPM3* which encodes a slow tropomyosin isoform, are selectively expressed in type I myofibres and are thus consistent with the fibre type pathology. However, some CFTD-associated genes appear expressed in both myofiber types, for example *ACTA1*[30]. It remains largely unclear why or how such genes selectively affect type I myofibres despite being expressed in all myofiber types.

Table 1.4 Genes associated with congenital myopathy with fibre type disproportion

Gene	Protein	Inheritance	OMIM#
ACTA1	Skeletal muscle alpha actin	AD, AR	255310
CACNAS1[2]	Dihydropyridine receptor	AR	-
MYH7	Myosin, heavy chain 7	AD	160760
RYR1	Ryanodine receptor I	AR	117000
SELENON	Selenoprotein N	AR	255310
TPM2	Tropomyosin 2	AD	190990
TPM3	Tropomyosin 3	AD, AR	255310
ZAK	Sterile alpha motif and leucine zipper containing kinase AZK	AR	617760

Abbreviations include: AD, autosomal dominant; AR, autosomal recessive.

1.3 Clinical and genetic heterogeneity of congenital myopathies

As apparent above, attaining an accurate congenital myopathy diagnosis according to traditional nosology is a major challenge given the heterogeneity of these disorders[3]. In some cases, histopathological assessments from muscle biopsies are also variable and inconsistent, and thus may not offer a representative or definitive diagnosis[1]. Accordingly, congenital myopathies have become increasingly classified and diagnosed according to the affected gene[5]. However, the heterogeneity of congenital myopathies can also exist at a gene level, whereby a single gene is associated with various congenital myopathy subtypes[31] (Figure 1.1). In addition, several genes may be associated with the same phenotype[32]. To further complicate matters, this heterogeneity can extend across other neuromuscular disorders[33]. The heterogeneities of congenital myopathies are further discussed below.

1.3.1 One gene, variable phenotypes

Variants in a single gene for example *RYR1* can be associated with diverse disease manifestations, onset, severity, progression, and prognosis[2,34].

At a monogenic level, the phenotypic variability has been suggested to be influenced by the variant type and location within a gene. For example, most *RYR1* variants associated with CCD are localised towards the C-terminus of *RYR1,* whereas variants

implicated in MmD and CNM are dispersed throughout the gene[34]. It also appears that biallelic missense variants in *RYR1* are more commonly associated with MmD, and that loss of function variants such as nonsense or frameshift variants in *RYR1* are more frequently observed in CCD, as well as in cases of lethal fetal akinesia[2,35].

It has also been suggested that different substitutions throughout a protein may affect specific protein-protein interactions and consequently give rise to different phenotypes[36,37]. This is particularly implicated in genes encoding large structural proteins such as *RYR1*[38] and *TTN*[36].

Interestingly, a single variant can also be associated with variable manifestations even within a family (i.e., same variant, different presentation)[39,40]. This intrafamilial variability suggests there are other factors that influence disease manifestations[40]. In particular, genetic modifiers are becoming increasingly recognised as contributors to Mendelian disease variation and may partly explain the heterogeneity of congenital myopathies[41-43]. Variants in enhancer or suppressor genes have been discovered by NGS and shown to exacerbate or alleviate disease severity[41,44,45]. For example, variants in *MTMR14* have been associated with neonatal onset of centronuclear myopathy 1, (CNM1; Table 1.3), a DNM2-related disorder that typically manifests later during childhood or adulthood[46].

1.3.2 Multiple genes, specific phenotype

Alternatively, multiple genes can be associated with the same congenital myopathy phenotype (Figure 1.1)[4]. This partly results from changes in genes that encode proteins with a similar function or are involved in similar pathways[2,47]. For example, deficiency of KLHL40 and proteins of the sarcomere such as LMOD3 have been associated with similar nemaline myopathy presentations including distinctly-shaped nemaline bodies[32]. Although initially there was no known link between BTB-kelch proteins and sarcomeric thin filament proteins, KLHL40 was subsequently shown to stabilise LMOD3 and NEB and thereby explained their related disease phenotypes[48]. It is predicted, and proven, that additional congenital myopathy genes may be uncovered by investigating genes involved in the same pathway or that are known to interact with a known disease gene[2,49,50].

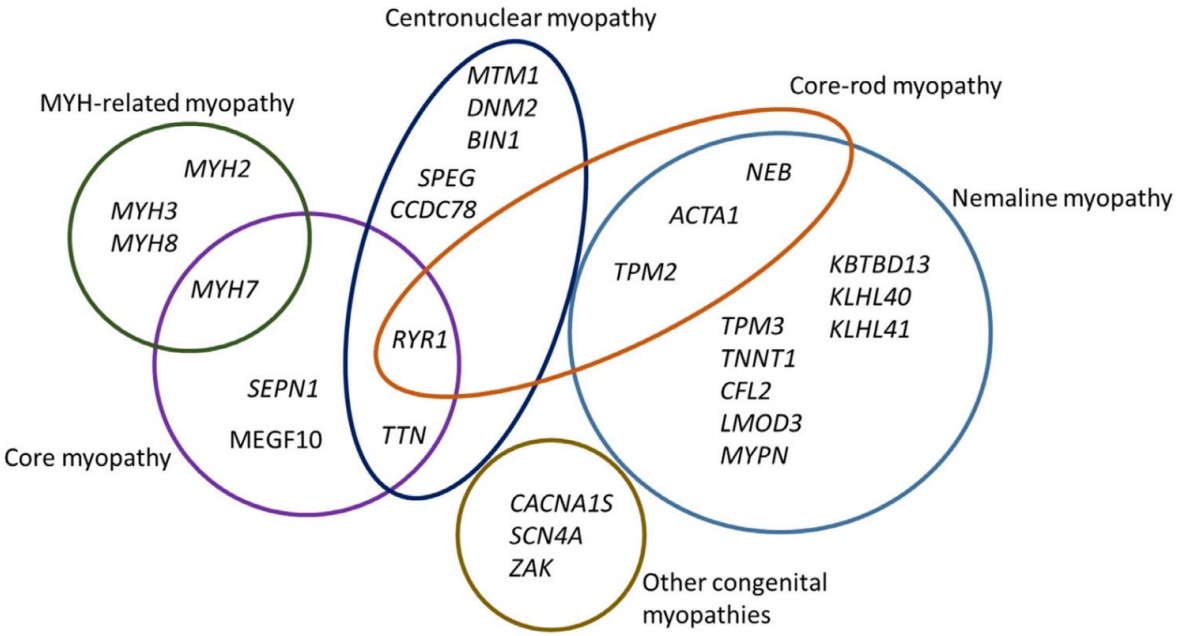


Figure 1.1 Diagrammatic representation of 27 genes implicated in various congenital myopathies[4].

1.3.3 Genotypic and phenotypic overlap with other neuromuscular disorders

Congenital myopathies can also show significant genotypic and phenotypic overlap with other neuromuscular disorders and can thus challenge classical neuromuscular disease nosology. Some overlapping disorders include muscular dystrophies, metabolic myopathies, and some neuropathies[, 5,8,51-55]. Due to the clinical variability, the concepts of many neuromuscular disorders have been challenged as "pure" disease entities but rather, a continuum dependent on the implicated gene and variants[14].

The blurring of disease entities is particularly being observed between congenital myopathies and congenital muscular dystrophies[4]. Congenital muscular dystrophies differ from congenital myopathies by their typically progressive and dystrophic manifestations and are often accompanied by elevated serum creatine kinase levels[51]. Similar to congenital myopathies however, they tend to manifest early in life and involve hypotonia, respiratory failure and feeding difficulties[56]. The phenotypic overlap can add to the difficulty in making a differential diagnosis for some patients, particularly when assessed on the basis of clinical manifestations alone[57,58]. In some cases, the overlap between these disorders is explained by the implicated gene[56]. For example, *SELENON*, initially associated with congenital muscular dystrophy[26] is also implicated in several congenital myopathies including core myopathy and congenital

myopathy with fibre type disproportion[25,56]. Given the challenge to accurately diagnose SELENON-related disorders and other congenital myopathies using classical nosology[4,56,57,59,60], a gene-centric classification for myopathies has been increasingly used to provide a clearer and unbiased diagnosis[4]. Other examples include actinopathies[14] and titinopathies[61].

1.4 Next generation sequencing for molecular diagnostics and disease gene discovery

The shift to a gene-centric naming of congenital myopathies largely stems from the increased applications of NGS, a high-throughput massively parallel sequencing technology that can comprehensively map genetic variation and identify underlying genetic causes of rare diseases[5,12].

1.4.1 NGS approaches

The three common NGS approaches; targeted gene panel sequencing, whole exome sequencing (WES), and whole genome sequencing (WGS) have dramatically expanded the number of variants and genes implicated in neuromuscular diseases, including congenital myopathies[62].

1.4.1.1 Targeted gene panel sequencing

Targeted gene panels involve selective screening for variants in known disease genes and are useful for identifying coding variants in congenital myopathy-associated genes[63]. Targeted panels are routinely used in diagnostic facilities such as PathWest (Perth, Australia)[63] and commercial diagnostic sequencing providers. For example, the PathWest neurogenetic disease gene panels capture over 500 neuromuscular disease genes and is routinely updated to include novel neuromuscular disease genes[63,64] .

1.4.1.2 Whole exome sequencing

WES involves screening the coding regions (i.e., 1-2%) of the genome, which are estimated to comprise 85% of pathogenic variants implicated in Mendelian disease[65]. WES has been applied both clinically and in research settings and has facilitated the discovery of numerous novel human disease genes[13,62,66,67]. Of note, the gene coverage captured by WES is generally lower than the targeted panel approach[62]. This may limit identification of some pathogenic changes such as copy number variants, and variants

in poorly covered coding regions such as regions with high GC content[62,65,68].

1.4.1.3 Whole genome sequencing

WGS can detect variants throughout the genome that may be uncaptured by WES and targeted gene sequencing. These include deep intronic variants and structural variants such as copy number variants, short tandem repeat expansions and other complex genomic events, e.g. inversions, rearrangements[69,70]. Despite its utility, WGS is typically performed for WES-negative cases on a research basis and as a last resort due to its comparatively higher costs and relatively greater challenges associated with variant interpretation and prioritisation[70-72]. Regardless of the approach, NGS findings often require validation by approaches such as Sanger sequencing, particularly when in regions of low coverage[62,65].

1.4.1.4 Short read sequencing

Standard NGS approaches are performed by short-read sequencing methods[70]. This involves amplification and sequencing of short, broken DNA fragments (150–300 bp) that are subsequently assembled by alignment to a reference genome. This method is prone to alignment errors, poor coverage and sequencing artifacts[70]. Consequently, this can limit the identification of variants in highly repetitive regions, high GC content regions, large, complex structural variants and tandem repeat expansions, and limit pseudogene discrimination, and allele phasing[70,73]. Consequently, patients with such variants may remain undiagnosed following standard NGS[73].

1.4.1.5 Long-read sequencing

Long-read sequencing (LRS) technologies such as the Oxford Nanopore have recently emerged as promising diagnostic tools that can identify pathogenic variants that have not been discovered by short-read sequencing[70]. The main advantages of LRS stem from the real-time sequencing of long reads (>10 kb) from a single DNA molecule[74]. This not only reduces alignment and mapping errors but also overcomes the need for PCR amplification, which enables direct detection of epigenetic modifications such as methylation on native DNA[74,75]. Additionally, LRS can enable sequencing of whole transcripts without requirements for assembly[74]. Thus far, LRS has been primarily used to investigate rare disorders with previous knowledge of the suspected loci. For example, Mangin *et al*.[76] recently demonstrated the use of long-read targeted sequencing in detecting mosaicism and complex repeat expansions in the myotonic

dystrophy protein kinase (*DMPK*) gene implicated in myotonic dystrophy. Despite its potential, LRS is in its infancy and is accompanied by several limitations including relatively higher costs, higher error rates and more labour-intensive library preparation compared to standard NGS[73,77]. Additionally, data handling and variant interpretation pipelines are less developed than standard NGS aproaches[77].

1.4.2 NGS variant analysis approaches

1.4.2.1 Variant prioritisation platforms

Due to the vast genetic data captured by NGS, identification of the genetic cause of disease is likened to searching for a "needle in a haystack"[65,78]. To reduce the considerable number of variants identified by NGS, computational and data sharing platforms for variant prioritisation have been developed[79]. Software frameworks such as seqr (https://seqr.broadinstitute.org/) are designed with user friendly interfaces for variant calling, annotation, filtering and interpretation[80]. These platforms integrate data from secondary databases including population genetic databases, gene expression databases and *in silico* variant pathogenicity prediction tools to aid in variant filtration, analysis, and prioritisation[81].

1.4.2.2 Variant population databases

Analysis of allele frequencies in large-scale population datasets of natural genetic variation can help distinguish between variants that are common and polymorphic, or rare and deleterious in the population, this is particularly powerful in the context of rare and early-onset disease such as congenital myopathies[80]. The Genome Aggregation Database (gnomAD)[82] is particularly useful for this purpose as it contains over 240 million small variants, structural variants and non-coding variants captured from 125,748 exomes and 15,708 whole genomes from healthy individuals[83]. For analysis of congenital myopathies and other rare disorders, variants with a minor allele frequency (MAF) $\leq 1\%$ in the healthy population are principally prioritised for further investigation as they support a rare manifestation of disease[84].

Large-scale population data can provide useful information about gene intolerance to variation[82,85]. The proportion of observed missense and loss of function (LoF) variants in a gene compared to the statistically predicted number of variants can be used to determine gene constraint metrics. For example, observed-to-expected ratio (OE), Z scores, and probability of loss of function intolerance (pLI) are available on gnomAD

and can help evaluate the likelihood that variants in a candidate gene will cause disease[82,85]. However, this approach is not suitable for evaluating constraint of short genes, as the scores may be subject to length bias[86,85]. In addition, sequencing artefacts can lead to false annotations of variation and confound gene constraint values[85,87]. Therefore, such scores should be interpreted with reason.

1.4.2.3 Data sharing

Identifying multiple patients with matching phenotypes and variants in the same novel disease gene candidate provides strong evidence of gene causality[88]. This can be achieved by researching and sharing candidate genes and variants in human disease databases for example the Leiden Open Variation Database (LOVD)[89] and the ClinVar database[72,90]. In addition, matches can be identified by personal communications and via disease gene matchmaking platforms[5,78]. *Matchbox*[81], GeneMatcher[91] and MatchMaker Exchange[78] are examples of such platforms used to connect researchers worldwide to facilitate matching and support novel disease gene candidates[78]. In addition, data sharing can facilitate collaborations between groups with broader expertise and allow for implementation of functional investigations and model organism studies to achieve an accurate molecular and clinical diagnosis[92].

1.4.2.4 Variant segregation analysis

The diagnostic efficiency of NGS can be significantly improved when NGS is performed for patients and their parents (i.e., trio sequencing) and other affected or unaffected family members[93]. This approach enables refining and filtering of variants that co-segregate with disease in the family. Trio analysis can be useful for identifying *de novo* and compound heterozygous pathogenic variants that are often overlooked when a proband is sequenced in isolation[88,94].

1.4.2.5 Gene expression databases

Gene expression analyses can be useful for prioritising candidate genes not previously associated with myopathy[84,87]. Expression of candidate genes can be explored using available data in expression databases from various control human tissues. Examples of expression databases include GTEx (Genotype-Tissue Expression project)[95] and FANTOM5 (Functional Annotation of the Mammalian genome consortium)[96]. Variants in genes that are enriched in disease-relevant tissue e.g. skeletal muscle can provide stronger justification for being associated with the disease under investigation,

e.g. congenital myopathy[84,97]. However, it is noteworthy that some ubiquitously expressed genes can also cause a muscle-specific pathology[97-99]. For example, *SELENON* is detected in all human tissues but is specifically associated with congenital myopathy and congenital muscular dystrophy[100]. Thus, prioritisation of candidate genes according to specific tissue enrichment should be approached with caution.

1.4.2.6 In silico pathogenicity tools

Various *in silico* tools are available for predicting variant pathogenicity. These include SIFT (Sorting Intolerant from Tolerant)[101], PolyPhen-2 (Polymorphism Phenotyping)[102], CADD (Combined Annotation Dependent Depletion)[103], PROVEAN (Protein Variation Effect Analyzer)[104] and MutationTaster[105]. Such tools predict pathogenicity by assessing distinct parameters, for example sequence conservation (e.g. SIFT), variant position in relation to structural and functional protein features (e.g. PolyPhen-2) as well as combined genomic features including epigenetic, evolutionary, and functional annotations (e.g. CADD). Nevertheless, most *in silico* tools require *a priori* knowledge of gene features and function to accurately predict pathogenicity[12,106]. Additionally, most current tools do not integrate data relating to gene constraint, tissue-enrichment or ethnicity and may thus lack complete "genotype awareness" or "phenotype awareness"[12,84,107]. As such, *in silico* prediction tools can be prone to type I and type II predictive errors and respectively report benign variants as pathogenic, and pathogenic variants as benign[84,107]. Assessment of multiple prediction tools is recommended to increase predictive accuracy, though conflicting results between tools can obscure such results[12,72,108]. Meta-predictor tools such as MetaSVM (meta-analytic support vector machine)[109] provide a consensus variant pathogenicity prediction using machine learning by integrating results from multiple *in silico* tools. Such tools have been shown in some cases to improve prediction accuracy[12].

1.4.2.7 In silico modelling tools

Various resources and computational tools are also available for visually exploring disease gene and protein candidates. In particular, the Protein Data Bank (PDB; www.rcsb.org)[110] provides open access to experimentally derived data on the 3D structures and interactions of protein, DNA and RNA. The 3D structures can be readily

downloaded and further analysed by molecular dynamic programs such as the PyMOL molecular graphics system (Schrödinger) to explore the effects of substitutions on protein structure prior to performing functional work. In addition, various molecular dynamic programs are available that can simulate interactions and predict the effects of mutations on the free binding energy, conformation, and biochemical properties of candidate proteins. For example, Mohajer *et al*.[111] recently used molecular simulations to predict the effects of ACTA1 substitutions implicated in CFTD. Their simulations suggested that changes in the hydrogen bonding capacity, hydrophobicity and conformation of mutant actin resulted in abnormalities in thin filament assembly[111]. Of note, candidate proteins without experimentally derived structural data are difficult to investigate by such tools[112]. This limitation is being addressed by development of novel machine learning approaches such as AlphaFold which can accurately predict protein structures using deep learning algorithms[112].

1.4.2.8 Variant classification guidelines

To standardise the classification and reporting of variants prioritised from NGS, variant classification guidelines are available to assist with variant interpretation. The American College of Medical Genetics and Genomics and Association of Molecular Pathology (ACMG-AMP) guidelines classify variants according to five classes ranging from pathogenic to benign[88]. The classifications are dependent on factors including variant allele frequency in healthy and disease population databases, variant co-segregation, *in silico* predictions and functional evidence[72]. However, the ACMG-AMP guidelines are not suitable for the interpretation of novel disease gene candidates identified by NGS. In their frameworks, genes not previously associated with disease are referred to as "genes of uncertain significance"[88]. It is tempting to predict that such frameworks can obstruct novel congenital myopathy genes from being further investigated and validated given their uncertain significance status. Thus, recommendations for the interpretation and reporting of genes of unknown significance are needed to increase the clinical utlility and diagnostic efficiency of NGS[113,114].

1.4.2.8.1 Variants of unknown significance

Despite the utility of NGS and the development of variant prioritisation strategies, a considerable number of patients with congenital myopathies remain without a genetic

diagnosis following screening[12,17,53,115-118]. Particularly, the bottleneck appears to lie in the interpretation of variants identified by NGS and distinguishing pathogenic variants from variants that are benign or unrelated to the disease of interest[12,113,119,120]. NGS has identified a significant number of variants in known disease genes with uncertain functional effect or insufficient evidence to infer pathogenicity. Such variants cannot be classified as pathogenic or benign according to the ACMG-AMP variant classification guidelines and are thus classed as variants of unknown significance (VUS)[121]. Common VUS include missense variants in large genes such as *RYR1*, *TTN* and *NEB*[122]. VUS are predicted to account for ~40% of variants captured by NGS[121] and the number is expected to rise with the increased quantity and complexity of genetic data captured[5,88,120,121,123,124]. The interpretation and reporting of VUS can be discordant between different investigators and sequencing providers and thus hinder an accurate genetic diagnosis[125,126,127]. Thus, improved frameworks and functional approaches are needed to reclassify VUS as likely benign or pathogenic in the context of disease[121].

1.5 Functional approaches to validating disease variants and genes

Functional assays can provide strong support for variant pathogenicity and are often required when the mechanism of a variant is not well understood or previously characterised[72,84]. Numerous assays have been described in the literature that capture various disease mechanisms[118,128]. These broadly fall under investigations of gene expression, protein abundance, structure and/or function, as well as cell-based and model organism-based investigations.

1.5.1 Functional approaches for gene expression analyses

1.5.1.1 Variants affecting splicing

Variants predicted to affect splicing can be investigated by reverse transcriptase PCR using cDNA studies from available patient tissue[124,129]. Where patient tissue is unavailable, *in vitro* assays such as minigene splicing assays can be performed[130-132]. Minigene assays are often performed in cells that are relatively easy to transfect and culture, for example human embryonic kidney 293 (HEK293) cells[132]. However, consideration must be taken into the cell type used given that splicing regulation can be cell-specific and may not reflect what is occurring in affected patient tissue[131].

1.5.1.2 Variants in untranslated regions

Variants in untranslated regions (UTRs) can have various implications on transcript expression, localisation and translation[133]. Pathomechanisms include aberrant splicing[134], removal of RNA protein binding sites and polyadenylation sites[133], as well as formation of unstable secondary structures[133] and illegitimate microRNA sites[134,135]. Some of these mechanisms can be captured by *in vitro* expression assays and reporter assays[135,136]. For example, Dusl *et al*.[135] used a dual-luciferase assay to show that a 3' UTR variant in *GFPT1* (the glutamine-fructose-6-phosphate transaminase 1 gene) in patients with unsolved congenital myasthenic syndrome created an illegitimate microRNA binding site for miR-206*. Additionally, they validated the 3' UTR mechanism in patient-derived myoblasts by showing that administration of miR-206* inhibitors restored GFPT1 levels[135]. Given the crucial roles of UTRs in RNA metabolism, it is expected that UTR variants are underrecognised causes of disease in many patients[133,135,137]. While UTR variants are often detected by NGS of patients with neuromuscular disorders, they are often overlooked[135]. Performing functional assays such as those described by Dusl *et al*.[135] to characterise recurrent UTR variants may clarify the unsolved disorders affecting some patients[134,136].

1.5.1.3 RNA-seq

Complementary transcriptomic approaches such as RNA sequencing (RNA-seq) are emerging as valuable tools (where patient samples are available) for the identification, interpretation, and validation of genetic candidates that affect gene expression or cause aberrant splicing[138]. RNA-seq utilises NGS platforms to quantify transcript abundance in patient-derived cells or tissue and has been useful for investigating variants that are difficult to identify or interpret by NGS alone[139-141]. For example, using RNA-seq, Cummings *et al*.[139] discovered a deep intronic splice variant in the *COL6A1* (collagen type VI alpha 1 chain) gene which was undetected by targeted gene sequencing and WES. Notably, this variant was highly recurrent in patients with unsolved Ullrich congenital muscular dystrophy[139]. This suggests that many patients with unsolved disorders may similarly harbour hidden deep intronic variants that are identifiable by RNA-seq[139]. While RNA-seq is clinically applied for the diagnosis of myopathies with *a priori* knowledge about the pathomechanisms of disease, the utility of RNA-seq as a research-based method for identifying novel mechanisms remains in its infancy[139].

1.5.2 Protein investigations

1.5.2.1 Variants affecting protein abundance

Western blotting is commonly used to investigate variants predicted to affect protein stability and abundance[17]. Western blotting can be performed using patient-derived tissue or from cell-based protein expression assays[17,135,142]. However, the dependency of this approach on validated antibodies (which can be difficult to obtain) can make it ineffective and time consuming[143]. Proteomic approaches such as quantitative mass spectrometry are emerging as high-throughput antibody-free techniques that may eventually replace western blotting. Quantitative mass spectrometry can accurately quantify numerous proteins simultaneously in cells and tissues of interest[144] This method can be useful for investigating the muscle proteome and facilitate identification of novel congenital myopathy candidates[145]. However quantitative mass spectrometry data can be complicated by signal interference. In addition, the high running costs can make it relatively more impractical than western blotting[143,146].

1.5.2.2 Variants affecting protein structure, stability, and/or function

Recombinant wildtype and mutant protein can be produced from bacterial and mammalian expression systems for a variety of structural and functional analyses[98,147,148]. Techniques such as circular dichroism, X-ray crystallography and nuclear magnetic resonance can be used to investigate whether substitutions affect protein structure and thermal stability[147-150]. Kinetic assays can be performed *in vitro* to investigate whether substitutions affect enzyme activity[98]. Additionally, yeast two hybrid assays and isothermal titration calorimetry can be used to investigate protein-protein interactions[48,147]. However, one consideration to recombinant protein characterisation is how representative such proteins are to those *in vivo*. In particular, expression systems may lack the relevant post translational modification machinery that ensure proper protein folding, function and protein-protein interactions[151]. Of note, recombinant studies can be challenging for investigating large proteins such as TTN[36]. The entire protein sequence for such proteins may be difficult to clone and express in a single construct and thus may need to be "dissected" into several constructs of key functional domains for analysis[36,124].

1.5.3 Cell-based investigations

Cell-based reporter assays are also available for investigating specific implications of

a variant or gene knockdown on cell function. Assays include measuring cell oxidative stress[152], assessing membrane repair responses[98], investigating protein stability and localisation[64,153] and measuring abundance of metabolites for example retinoic acid[128].

1.5.4 Model organisms

Cell and animal models have been essential for investigating the genetic mechanisms of congenital myopathies *in vitro* and *in vivo*[142,154]. Novel disease gene candidates can be knocked out by gene editing technologies such as CRISPR/Cas9 to confirm relevant genotype-phenotype associations[155]. While CRISPR/Cas9 editing of cells is commonly used to investigate functional effects of variants in rare diseases[118,156], it remains a challenge to perform such investigations on skeletal muscle cells given their difficulty to transfect and survive gene editing[155].

Induced pluripotent stem cells (IPSCs) are emerging as useful *in vitro* disease models for rare disease investigation[157]. IPSCs can be reprogrammed from patient-derived cells and differentiated into relevant cell types[157]. The differentiated cells retain the patient's genetic disorder and can thus be used as a relevant biological system to explore molecular pathomechanisms and potential treatments and therapies[158,159]. While there are few reports of IPSC models for congenital myopathies, Clayton *et al.*[158] and Houweling *et al.*[160] recently demonstrated the utility of reprogramming patient-derived cells into IPSCs for investigating pathomechanisms of ACTA1-related diseases[158,160].

Various animal models have been used for neuromuscular disease research, ranging from simple invertebrates such as yeast (e.g. *C. elegans*) to more advanced models such as fruit fly (*D. melanogaster*), zebrafish, (*D. rerio*), mouse (*M. musculus*) and dog (*C. lupus*)[154,161]. Zebrafish models have been particularly useful in investigating functions of novel congenital myopathy disease genes for example *MYL1*[13] and *KLHL40*[17]. The embryos of zebrafish are transparent and contain a relatively high proportion of skeletal muscle[162]. The embryos develop *ex utero* and can thus be closely studied throughout all stages of development[154,162]. Of note, the anatomical and biological differences between zebrafish and human may be argued as a limitation when interpretating the resulting phenotypes of a gene defect[154]. Mammalian models such as mouse can resemble human disease more accurately and have also been used to investigate causes and potential treatments for congenital myopathies[98,161,163,164].

However, some genes are challenging to characterise in mammalian models given that knockdown of some genes leads to embryonic lethality due to placental defects and secondary malformations such as cardiovascular deficits[162,163,165,166].

1.5.5 Guidelines for functional approaches

According to ACMG-AMP guidelines, functional approaches must be "well-established" to qualify as evidence for variant pathogenicity[88]. Attributes to a well-established assay include validity, reproducibility, and biological relevance of the assay. The type of assay performed depends on the known function of the gene, the proposed mode of inheritance and type of variant under investigation[88,126]. Some functional assays may not be suitable in capturing the mechanism of a variant and are thus less reliable predictors of variant pathogenicity. Another important consideration includes how closely an assay reflects the biological environment[88]. Generally, assays involving affected patient samples provide stronger evidence than *in vitro* studies, though they may not always be available for analysis[13,17,88,167]. Of note, while the ACMG-AMP guidelines provide a general framework to guide functional genomic investigations, they lack clarifications regarding assay thresholds and necessary controls required to prove an assay valid[126]. To assist with the evaluation, frameworks have been developed by panels including the Variant Curation Expert Panel (VCEP), which have a list of suggested functional assays per gene and variant type[168]. However, these frameworks are currently mostly suited for the specific diseases and genes listed and are not suitable for investigating novel disease genes. Evaluating the functional data for novel candidates may require expert judgement with knowledge of the assay method, gene ontology and disease mechanism under investigation[168].

1.5.6 Collaborative functional genomics

Functional genomic collaborations can accelerate investigations of novel disease genes and development of therapies[78,169,170]. Global data sharing programs such as the Canadian Rare Diseases Models and Mechanisms (RDMM) Network facilitate global collaborations between gene discovery researchers and functional genomics and model organism research teams[170]. There is now also an Australian Functional Genomics Network (AFGN; https://www.functionalgenomics.org.au/), which facilitates collaborations between Australian rare disease investigators with clinical and model research teams.

The International Mouse Phenotyping Consortium (IMPC) is an ongoing global effort that generates and phenotypes knockout mouse lines in aims of elucidating the function of all mammalian protein-coding genes[163]. The resulting phenotype data is freely accessible on the IMPC website (https://www.mousephenotype.org/) and can assist with the prioritisation and validation of novel disease gene candidates identified by NGS[163]. Interestingly, the current data from these experiments suggest that ~23% of genes are associated with embryonic lethality[163]. Some of these genes have not yet been implicated in human disease yet and can thus be flagged during NGS analysis of patients with severe and early-onset disease[163,171].

1.6 Significance of a genetic diagnosis

While most congenital myopathies are incurable, a genetic diagnosis can significantly improve patient outcomes particularly when identified early[2,3,118]. Identification of the underlying genetic cause of disease can provide closure for families, clarity to improve family planning, and prevent disease-recurrence through preconception, prenatal or carrier screening[5]. In some instances, a molecular diagnosis facilitates treatment, as is the case for some congenital myasthenic syndromes[172]. Moreover, knowledge of the precise molecular mechanism of muscular dysfunction can open new avenues for the development of tailored therapies and personalised treatments[3,5,173].

1.7 Conclusion

The genetic revolution accompanying NGS is shifting the nosology of congenital myopathies and enabling earlier disease detection and accurate diagnosis[53]. However, a significant proportion of patients with congenital myopathy remain without a genetic diagnosis despite comprehensive genetic screening. The bottleneck in NGS diagnostics appears largely in the functional and clinical validation of novel variants and genes identified[118]. Validation of genetic candidates by functional genomics can improve the diagnostic efficiency of NGS[113]. A definitive diagnosis will offer crucial insights that can influence disease management and prognosis for families worldwide[5].

1.8 Research aims and thesis layout

The purpose of this study was to identify and functionally investigate genetic

mechanisms of congenital myopathies in patients without a genetic diagnosis. The overarching aim of this project was to identify the genetic cause of disease in patients with congenital myopathies. This thesis is distilled into a general methods chapter (Chapter 2), two results chapters and a general discussion chapter. Chapter 3 describes the identification and investigation of biallelic *KLHL40* variants in a patient with mild NEM8, including a novel 3' UTR single nucleotide variant. Chapter 4 details the identification and characterisation of HMGCS1-related rigid spine syndrome. The thesis closes with a chapter comprising a general discussion and future directions.

1.9 References

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

1. Cassandrini D, Trovato R, Rubegni A, et al. Congenital myopathies: clinical phenotypes and new diagnostic tools. *Ital J Pediatr.* 2017;43(1):101.

2. Gonorazky HD, Bönnemann CG, Dowling JJ. The genetics of congenital myopathies. *Handb Clin Neurol.* 2018;148:549-564.

3. Beecroft SJ, Lamont PJ, Edwards S, et al. The Impact of Next-Generation Sequencing on the Diagnosis, Treatment, and Prevention of Hereditary Neuromuscular Disorders. *Mol Diagn Ther.* 2020.

4. Pelin K, Wallgren-Pettersson C. Update on the Genetics of Congenital Myopathies. *Semin Pediatr Neurol.* 2019;29:12-22.

5. Ravenscroft G, Davis MR, Lamont P, Forrest A, Laing NG. New era in genetics of early-onset muscle disease: Breakthroughs and challenges. *Semin Cell Dev Biol.* 2017;64:160-170.

6. Goutman SA, Callaghan BC, Feldman E. A 2020 centenary perspective on neuromuscular disorders. *J Neurol Neurosurg Psychiatry.* 2020;91(8):791-792.

7. Benarroch L, Bonne G, Rivier F, Hamroun D. The 2021 version of the gene table of neuromuscular disorders (nuclear genome). *Neuromuscul Disord.* 2020;30(12):1008-1048.

8. Neto OA, Tassy O, Biancalana V, Zanoteli E, Pourquié O, Laporte J. Integrative Data Mining Highlights Candidate Genes for Monogenic Myopathies. *PLoS One.* 2014;9(10):e110888.

9. Ghaoui R, Cooper ST, Lek M, et al. Use of Whole-Exome Sequencing for Diagnosis of Limb-Girdle Muscular Dystrophy: Outcomes and Lessons Learned. *JAMA Neurol.* 2015;72(12):1424-1432.

10. Anselme F, Moubarak G, Savouré A, et al. Implantable cardioverter-defibrillators in lamin A/C mutation carriers with cardiac conduction disorders. *Heart Rhythm.* 2013;10(10):1492-1498.

11. Natera-de Benito D, Nascimento A, Abicht A, et al. KLHL40-related nemaline myopathy with a sustained, positive response to treatment with acetylcholinesterase inhibitors. *J Neurol.* 2016;263(3):517-523.

12. Sun H, Yu G. New insights into the pathogenicity of non-synonymous variants through multi-level analysis. *Sci Rep.* 2019;9(1):1667.

13. Ravenscroft G, Zaharieva IT, Bortolotti CA, et al. Bi-allelic mutations in MYL1 cause a severe congenital myopathy. *Hum Mol Genet.* 2018;27(24):4263-4272.

14. Nowak KJ, Ravenscroft G, Laing NG. Skeletal muscle α-actin diseases (actinopathies): pathology and mechanisms. *Acta Neuropathol.* 2013;125(1):19-32.

15. Goebel HH, Piirsoo A, Warlo I, Schofer O, Kehr S, Gaude M. Infantile intranuclear rod myopathy. *J Child Neurol.* 1997;12(1):22-30.

16. Géraud J, Dieterich K, Rendu J, et al. Clinical phenotype and loss of the slow skeletal muscle troponin T in three new patients with recessive TNNT1 nemaline myopathy. *J Med Genet.* 2021;58(9):602-608.

17. Ravenscroft G, Miyatake S, Lehtokari V-L, et al. Mutations in KLHL40 Are a Frequent Cause of Severe Autosomal-Recessive Nemaline Myopathy. *Am J Hum Genet.* 2013;93(1):6-18.

18. Yuen M, Sandaradura SA, Dowling JJ, et al. Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. *J Clin Invest.* 2014;124(11):4693-4708.

19. Wallefeld W, Krause S, Nowak KJ, et al. Severe nemaline myopathy caused by mutations of the stop codon of the skeletal muscle alpha actin gene (ACTA1). *Neuromuscul Disord.* 2006;16(9):541-547.

20. Malfatti E, Böhm J, Lacène E, Beuvin M, Romero NB, Laporte J. A Premature Stop Codon in MYO18B is Associated with Severe Nemaline Myopathy with Cardiomyopathy. *J Neuromuscul Dis.* 2015;2(3):219-227.

21. Nilipour Y, Nafissi S, Tjust AE, et al. Ryanodine receptor type 3 (RYR3) as a novel gene associated with a myopathy with nemaline bodies. *Eur J Neurol.* 2018;25(6):841-847.

22. Romero NB, Clarke NF. Congenital myopathies. *Handb Clin Neurol.* 2013;113:1321-1336.

23. Phadke R. Myopathology of Congenital Myopathies: Bridging the Old and the New. *Semin Pediatr Neurol.* 2019;29:55-70.

24. Colombo I, Scoto M, Manzur AY, et al. Congenital myopathies: Natural history of a large pediatric cohort. *Neurology.* 2015;84(1):28-35.

25. Clarke NF, Kidson W, Quijano-Roy S, et al. SEPN1: associated with congenital fiber-type disproportion and insulin resistance. *Ann Neurol.* 2006;59(3):546-552.

26. Moghadaszadeh B, Petit N, Jaillard C, et al. Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome. *Nat Genet.* 2001;29(1):17-18.

27. Schartner V, Romero NB, Donkervoort S, et al. Dihydropyridine receptor (DHPR, CACNA1S) congenital myopathy. *Acta Neuropathol.* 2017;133(4):517-533.

28. Jungbluth H, Gautel M. Pathogenic mechanisms in centronuclear myopathies. *Front Aging Neurosci.* 2014;6:339.

29. Clarke NF, North KN. Congenital fiber type disproportion--30 years on. *J Neuropathol Exp Neurol.* 2003;62(10):977-989.

30. Laing NG, Clarke NF, Dye DE, et al. Actin mutations are one cause of congenital fibre type disproportion. *Ann Neurol.* 2004;56(5):689-694.

31. Bharucha-Goebel DX, Santi M, Medne L, et al. Severe congenital RYR1-associated myopathy: the expanding clinicopathologic and genetic spectrum. *Neurology.* 2013;80(17):1584-1589.

32. Sewry CA, Laitila JM, Wallgren-Pettersson C. Nemaline myopathies: a current view. *J Muscle Res Cell Motil.* 2019;40(2):111-126.

33. Ferreiro A, Quijano-Roy S, Pichereau C, et al. Mutations of the Selenoprotein N Gene, Which Is Implicated in Rigid Spine Muscular Dystrophy, Cause the Classical Phenotype of Multiminicore Disease: Reassessing the Nosology of Early-Onset Myopathies. *Am J Hum Genet.* 2002;71(4):739-749.

34. Treves S, Jungbluth H, Muntoni F, Zorzato F. Congenital muscle disorders with cores: the ryanodine receptor calcium channel paradigm. *Curr Opin Pharmacol.* 2008;8(3):319-326.

35. Todd EJ, Yau KS, Ong R, et al. Next generation sequencing in a large cohort of patients presenting with neuromuscular disease before or at birth. *Orphanet J Rare Dis.* 2015;10:148-148.

36. Savarese M, Sarparanta J, Vihola A, Udd B, Hackman P. Increasing Role of Titin Mutations in Neuromuscular Disorders. *J Neuromuscul Dis.* 2016;3(3):293-308.

37. Sahni N, Yi S, Taipale M, et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell.* 2015;161(3):647-660.

38. Witherspoon JW, Meilleur KG. Review of RyR1 pathway and associated pathomechanisms. *Acta Neuropathol Commun.* 2016;4(1).

39. Zima J, Eaton A, Pál E, et al. Intrafamilial variability of limb-girdle muscular dystrophy, LGMD1D type. *Eur J Med Genet.* 2020;63(2):103655.

40. Laing NG, Dye DE, Wallgren-Pettersson C, et al. Mutations and polymorphisms of the skeletal muscle alpha-actin gene (ACTA1). *Hum Mutat.* 2009;30(9):1267-1277.

41. Rahit KMTH, Tarailo-Graovac M. Genetic Modifiers and Rare Mendelian Disease. *Genes.* 2020;11(3):239.

42. Petillo R, D'Ambrosio P, Torella A, et al. Novel mutations in LMNA A/C gene and associated phenotypes. *Acta Myol.* 2015;34(2-3):116-119.

43. Tasca G, Ricci E, Penttilä S, et al. New phenotype and pathology features in MYH7-related distal myopathy. *Neuromuscul Disord.* 2012;22(7):640-647.

44. Spitali P, Zaharieva I, Bohringer S, et al. TCTEX1D1 is a genetic modifier of disease progression in Duchenne muscular dystrophy. *Eur J Hum Genet.* 2020.

45. Lamar K-M, McNally EM. Genetic Modifiers for Neuromuscular Diseases. *J Neuromuscul Dis.* 2014;1(1):3-13.

46. Tosch V, Rohde HM, Tronchère H, et al. A novel PtdIns3P and PtdIns(3,5)P2 phosphatase with an inactivating variant in centronuclear myopathy. *Hum Mol Genet.* 2006;15(21):3098-3106.

47. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci U S A.* 2007;104(21):8685-8690.

48. Garg A, O'Rourke J, Long C, et al. KLHL40 deficiency destabilizes thin filament proteins and promotes nemaline myopathy. *J Clin Invest.* 2014;124(8):3529-3539.

49. Gupta VA, Beggs AH. Kelch proteins: emerging roles in skeletal muscle development and diseases. *Skelet Muscle.* 2014;4:11.

50. Gupta VA, Ravenscroft G, Shaheen R, et al. Identification of KLHL41 Mutations Implicates BTB-Kelch-Mediated Ubiquitination as an Alternate Pathway to Myofibrillar Disruption in Nemaline Myopathy. *Am J Hum Genet.* 2013;93(6):1108-1117.

51. Mercuri E, Muntoni F. Muscular dystrophies. *Lancet.* 2013;381(9869):845-860.

52. Voit T, Kutz P, Leube B, et al. Autosomal dominant distal myopathy: further evidence of a chromosome 14 locus. *Neuromuscul Disord.* 2001;11(1):11-19.

53. Chae JH, Vasta V, Cho A, et al. Utility of next generation sequencing in genetic diagnosis of early onset neuromuscular disorders. *J Med Genet.* 2015;52(3):208-216.

54. Bonne G, Rivier F, Hamroun D. The 2019 version of the gene table of neuromuscular disorders (nuclear genome). *Neuromuscul Disord.* 2018;28(12):1031-1063.

55. Udd B. Distal myopathies. *Curr Neurol Neurosci Rep.* 2014;14(3):434.

56. Kazamel M, Milone M. Congenital myopathy with a novel SELN missense mutation and the challenge to differentiate it from congenital muscular dystrophy. *J Clin Neurosci.* 2019;62:238-239.

57. North KN, Wang CH, Clarke N, et al. Approach to the diagnosis of congenital myopathies. *Neuromuscul Disord.* 2014;24(2):97-116.

58. Kajino S, Ishihara K, Goto K, et al. Congenital fiber type disproportion myopathy caused by *LMNA* mutations. *J Neurol Sci.* 2014;340(1):94-98.

59. Poewe W, Willeit H, Sluga E, Mayr U. The rigid spine syndrome--a myopathy of uncertain nosological position. *J Neurol Neurosurg Psychiatry.* 1985;48(9):887-893.

60. Benarroch L, Bonne G, Rivier F, Hamroun D. The 2020 version of the gene table of neuromuscular disorders (nuclear genome). *Neuromuscul Disord.* 2019;29(12):980-1018.

61. Oates EC, Jones KJ, Donkervoort S, et al. Congenital Titinopathy: Comprehensive characterization and pathogenic insights. *Ann Neurol.* 2018;83(6):1105-1124.

62. Jamuar SS, Tan E-C. Clinical application of next-generation sequencing for Mendelian diseases. *Hum Genomics.* 2015;9(1):10.

63. Beecroft SJ, Yau KS, Allcock RJN, et al. Targeted gene panel use in 2249 neuromuscular patients: the Australasian referral center experience. *Ann Clin Transl Neurol.* 2020;7(3):353-362.

64. Ravenscroft G, Clayton JS, Faiz F, et al. Neurogenetic fetal akinesia and arthrogryposis: genetics, expanding genotype-phenotypes and functional genomics. *J Med Genet.* 2021;58(9):609-618.

65. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet.* 2012;20(5):490-497.

66. Ravenscroft G, Nolent F, Rajagopalan S, et al. Mutations of GPR126 are responsible for severe arthrogryposis multiplex congenita. *Am J Hum Genet.* 2015;96(6):955-961.

67. Kurul SH, Oktay Y, Töpf A, et al. High diagnostic rate of trio exome sequencing in consanguineous families with neurogenetic diseases. *Brain.* 2021.

68. Sanghvi RV, Buhay CJ, Powell BC, et al. Characterizing reduced coverage regions through comparison of exome and genome sequencing data across 10 centers. *Genet Med.* 2018;20(8):855-866.

69. Gross AM, Ajay SS, Rajan V, et al. Copy-number variants in clinical genome sequencing: deployment and interpretation for rare and undiagnosed disease. *Genet Med.* 2019;21(5):1121-1130.

70. Mitsuhashi S, Matsumoto N. Long-read sequencing for rare human genetic diseases. *J Hum Genet.* 2020;65(1):11-19.

71. Schwarze K, Buchanan J, Taylor JC, Wordsworth S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genet Med.* 2018;20(10):1122-1130.

72. Yubero D, Natera-de Benito D, Pijuan J, et al. The Increasing Impact of Translational Research in the Molecular Diagnostics of Neuromuscular Diseases. *Int J Mol Sci.* 2021;22(8):4274.

73. Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics. *Front Genet.* 2019;10(426).

74. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends Genet.* 2018;34(9):666-681.

75. Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods.* 2010;7(6):461-465.

76. Mangin A, de Pontual L, Tsai Y-C, et al. Robust Detection of Somatic Mosaicism and Repeat Interruptions by Long-Read Targeted Sequencing in Myotonic Dystrophy Type 1. *Int J Mol Sci.* 2021;22(5):2616.

77. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 2020;21(1):30.

78. Philippakis AA, Azzariti DR, Beltran S, et al. The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery. *Hum Mutat.* 2015;36(10):915-921.

79. Rodenburg RJ. The functional genomics laboratory: functional validation of genetic variants. *J Inherit Metab Dis.* 2018;41(3):297-307.

80. Dashti MJS, Gamieldien J. A practical guide to filtering and prioritizing genetic variants. *BioTechniques.* 2017;62(1):18-30.

81. Arachchi H, Wojcik MH, Weisburd B, et al. matchbox: An open-source tool for patient matching via the Matchmaker Exchange. *Hum Mutat.* 2018;39(12):1827-1834.

82. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv.* 2019:531210.

83. Koch L. Exploring human genomic diversity with gnomAD. *Nat Rev Genet.* 2020;21(8):448-448.

84. Anderson D, Baynam G, Blackwell JM, Lassmann T. Personalised analytics for rare disease diagnostics. *Nat Commun.* 2019;10(1):5274.

85. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.

86. Petrovski S, Gussow AB, Wang Q, et al. The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. *PLoS Genet.* 2015;11(9):e1005492.

87. Cummings BB, Karczewski KJ, Kosmicki JA, et al. Transcript expression-aware annotation improves rare variant interpretation. *Nature.* 2020;581(7809):452-458.

88. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-423.

89. Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat.* 2011;32(5):557-563.

90. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42(Database issue):D980-D985.

91. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat.* 2015;36(10):928-930.

92. Wright CF, Fitzpatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet.* 2018;19(5):253-268.

93. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet.* 2014;46(9):944-950.

94. Oliveira J, Negrao L, Fineza I, et al. New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing. *J Hum Genet.* 2015;60(6):305-312.

95. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585.

96. Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 2015;16(1):22.

97. Lage K, Hansen NT, Karlberg EO, et al. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc Natl Acad Sci U S A.* 2008;105(52):20870-20875.

98. Foley AR, Zou Y, Dunford JE, et al. GGPS1 Mutations Cause Muscular Dystrophy/Hearing Loss/Ovarian Insufficiency Syndrome. *Ann Neurol* 2020;88(2):332-347.

99. Kitsak M, Sharma A, Menche J, et al. Tissue Specificity of Human Disease Module. *Sci Rep.* 2016;6:35241.

100. Castets P, Lescure A, Guicheney P, Allamand V. Selenoprotein N in skeletal muscle: from diseases to function. *J Mol Med.* 2012;90(10):1095-1107.

101. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003;31(13):3812-3814.

102. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.

103. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-315.

104. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One.* 2012;7(10):e46688.

105. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods.* 2014;11(4):361-362.

106. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47(D1):D886-d894.

107. Bope CD, Chimusa ER, Nembaware V, Mazandu GK, de Vries J, Wonkam A. Dissecting in silico Mutation Prediction of Variants in African Genomes: Challenges and Perspectives. *Front Genet.* 2019;10(601).

108. Brnich SE, Rivera-Muñoz EA, Berg JS. Quantifying the potential of functional evidence to reclassify variants of uncertain significance in the categorical and Bayesian interpretation frameworks. *Hum Mutat.* 2018;39(11):1531-1541.

109. Kim S, Jhong J-H, Lee J, Koo J-Y. Meta-analytic support vector machine for integrating multiple omics data. *BioData Min.* 2017;10:2-2.

110. Burley SK, Berman HM, Bhikadiya C, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* 2018;47(D1):D464-D474.

111. Sadat Mohajer F, Parvizpour S, Razmara J, Shahir Shamsir M. The two mutations of actin–myosin interface and their effect on the dynamics,

structures, and functions of skeletal muscle actin. *J Biomol Struct Dyn.* 2019;37(2):372-382.

112. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583-589.

113. Lek M, MacArthur D. The Challenge of Next Generation Sequencing in the Context of Neuromuscular Diseases. *J Neuromuscul Dis.* 2014;1(2):135-149.

114. Nykamp K, Anderson M, Powers M, et al. Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet Med.* 2017;19(10):1105-1117.

115. Agrawal PB, Strickland CD, Midgett C, et al. Heterogeneity of nemaline myopathy cases with skeletal muscle alpha-actin gene mutations. *Ann Neurol.* 2004;56(1):86-96.

116. Cavallin M, Mine M, Philbert M, et al. Further refinement of COL4A1 and COL4A2 related cortical malformations. *Eur J Med Genet.* 2018;61(12):765-772.

117. Jang W, Kim Y, Han E, et al. Chromosomal Microarray Analysis as a First-Tier Clinical Diagnostic Test in Patients With Developmental Delay/Intellectual Disability, Autism Spectrum Disorders, and Multiple Congenital Anomalies: A Prospective Multicenter Study in Korea. *Ann Lab Med.* 2019;39(3):299-310.

118. Hartin SN, Means JC, Alaimo JT, Younger ST. Expediting rare disease diagnosis: a call to bridge the gap between clinical and functional genomics. *Mol Med.* 2020;26(1):117.

119. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature.* 2014;508(7497):469-476.

120. Nigro V, Piluso G. Next generation sequencing (NGS) strategies for the genetic testing of myopathies. *Acta Myol.* 2012;31(3):196-200.

121. Federici G, Soddu S. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. *J Exp Clin Cancer Res.* 2020;39(1):46.

122. Lehtokari V-L, Kiiski K, Sandaradura SA, et al. Mutation Update: The Spectra of Nebulin Variants and Associated Myopathies. *Hum Mutat.* 2014;35(12):1418-1426.

123. Savarese M, Di Fruscio G, Mutarelli M, et al. MotorPlex provides accurate variant detection across large muscle genes both in single myopathic patients and in pools of DNA samples. *Acta Neuropathol Commun.* 2014;2(1).

124. Savarese M, Maggi L, Vihola A, et al. Interpreting Genetic Variants in Titin in Patients With Muscle Disorders. *JAMA Neurol.* 2018;75(5):557-565.

125. Pont-Kingdon G, Gedge F, Wooderchak-Donahue W, et al. Design and analytical validation of clinical DNA sequencing assays. *Arch Pathol Lab Med.* 2012;136(1):41-46.

126. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* 2019;12(1):3.

127. Harrison SM, Dolinsky JS, Knight Johnson AE, et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet Med.* 2017;19(10):1096-1104.

128. Beecroft SJ, Ayala M, McGillivray G, et al. Biallelic hypomorphic variants in ALDH1A2 cause a novel lethal human multiple congenital anomaly syndrome encompassing diaphragmatic, pulmonary, and cardiovascular defects. *Hum Mutat.* 2021;42(5):506-519.

129. Bryen SJ, Ewans LJ, Pinner J, et al. Recurrent TTN metatranscript-only c.39974-11T>G splice variant associated with autosomal recessive arthrogryposis multiplex congenita and myopathy. *Hum Mutat.* 2020;41(2):403-411.

130. Yang YM, Yan K, Liu B, et al. Comprehensive genetic diagnosis of patients with Duchenne/Becker muscular dystrophy (DMD/BMD) and pathogenicity analysis of splice site variants in the DMD gene. *J Zhejiang Univ Sci B.* 2019;20(9):753-765.

131. Gaildrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol.* 2010;653:249-257.

132. Osborn DPS, Emrahi L, Clayton J, et al. Autosomal recessive cardiomyopathy and sudden cardiac death associated with variants in MYL3. *Genet Med.* 2021;23(4):787-792.

133. Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell.* 2009;101(5):251-262.

134. Liaqat K, Chiu I, Lee K, et al. Novel missense and 3'-UTR splice site variants in LHFPL5 cause autosomal recessive nonsyndromic hearing impairment. *J Hum Genet.* 2018;63(11):1099-1107.

135. Dusl M, Senderek J, Müller JS, et al. A 3'-UTR mutation creates a microRNA target site in the GFPT1 gene of patients with congenital myasthenic syndrome. *Hum Mol Genet.* 2015;24(12):3418-3426.

136. Jin C, Jiang J, Wang W, Yao K. Identification of a MIP mutation that activates a cryptic acceptor splice site in the 3' untranslated region. *Mol Vis.* 2010;16:2253-2258.

137. Bicknell AA, Cenik C, Chua HN, Roth FP, Moore MJ. Introns in UTRs: Why we should stop ignoring them. *BioEssays.* 2012;34(12):1025-1034.

138. Gonorazky H, Liang M, Cummings B, et al. RNAseq analysis for the diagnosis of muscular dystrophy. *Ann Clin Transl Neurol.* 2015;3(1):55-60.

139. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med.* 2017;9(386):1-11.

140. Hoeijmakers WA, Bartfai R, Stunnenberg HG. Transcriptome analysis using RNA-Seq. *Methods Mol Biol.* 2013;923:221-239.

141. Mele M, Ferreira PG, Reverter F, et al. Human genomics. The human transcriptome across tissues and individuals. *Science.* 2015;348(6235):660-665.

142. Ravenscroft G, Nowak KJ, Jackaman C, et al. Dissociated flexor digitorum brevis myofiber culture system--a more mature muscle culture system. *Cell Motil Cytoskeleton.* 2007;64(10):727-738.

143. Aebersold R, Burlingame AL, Bradshaw RA. Western blots versus selected reaction monitoring assays: time to turn the tables? *Mol Cell Proteomics.* 2013;12(9):2381-2382.

144. Stroud DA, Surgenor EE, Formosa LE, et al. Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature.* 2016;538(7623):123-126.

145. Schessl J, Bach E, Rost S, et al. Novel recessive myotilin mutation causes severe myofibrillar myopathy. *Neurogenetics.* 2014;15(3):151-156.

146. Pappireddi N, Martin L, Wühr M. A Review on Quantitative Multiplexed Proteomics. *ChemBioChem.* 2019;20(10):1210-1224.

147. Rudloff MW, Woosley AN, Wright NT. Biophysical characterization of naturally occurring titin M10 mutations. *Protein Sci.* 2015;24(6):946-955.

148. Kirwan JP, Hodges RS. Transmission of stability information through the N-domain of tropomyosin is interrupted by a stabilizing mutation (A109L) in the hydrophobic core of the stability control region (residues 97-118). *J Biol Chem.* 2014;289(7):4356-4366.

149. Louis-Jeune C, Andrade-Navarro MA, Perez-Iratxeta C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins.* 2012;80(2):374-381.

150. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc.* 2006;1(6):2876-2890.

151. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol.* 2014;5:172.

152. Olive M, Engvall M, Ravenscroft G, et al. Myoglobinopathy is an adult-onset autosomal dominant myopathy with characteristic sarcoplasmic inclusions. *Nat Commun.* 2019;10(1):1396.

153. Freed AS, Schwarz AC, Brei BK, et al. CHRNB1-associated congenital myasthenia syndrome: Expanding the clinical spectrum. *Am J Med Genet A.* 2021;185(3):827-835.

154. Pappalardo A, Pitto L, Fiorillo C, Alice Donati M, Bruno C, Santorelli FM. Neuromuscular Disorders in Zebrafish: State of the Art and Future Perspectives. *NeuroMolecular Med.* 2013;15(2):405-419.

155. Goullée H, Taylor RL, Forrest ARR, Laing NG, Ravenscroft G, Clayton JS. Improved CRISPR/Cas9 gene editing in primary human myoblasts using low confluency cultures on Matrigel. *Skelet Muscle.* 2021;11(1).

156. Cabrera-Serrano M, Coote DJ, Azmanov D, et al. A homozygous UBA5 pathogenic variant causes a fatal congenital neuropathy. *J Med Genet.* 2020;57(12):835-842.

157. Ye L, Swingen C, Zhang J. Induced pluripotent stem cells and their potential for basic and clinical sciences. *Curr Cardiol Rev.* 2013;9(1):63-72.

158. Clayton JS, Scriba CK, Romero NB, et al. Generation of two isogenic induced pluripotent stem cell lines from a 4-month-old severe nemaline myopathy patient with a heterozygous dominant c.553C > A (p.Arg183Ser) variant in the ACTA1 gene. *Stem Cell Res.* 2021;53:102273.

159. Moretti A, Fonteyne L, Giesert F, et al. Somatic gene editing ameliorates skeletal and cardiac muscle failure in pig and human models of Duchenne muscular dystrophy. *Nat Med.* 2020;26(2):207-214.

160. Houweling PJ, Coles CA, Tiong CF, et al. Generating an iPSC line (with isogenic control) from the PBMCs of an ACTA1 (p.Gly148Asp) nemaline myopathy patient. *Stem Cell Res.* 2021;54:102429.

161. Fusto A, Moyle LA, Gilbert PM, Pegoraro E. Cored in the act: the use of models to understand core myopathies. *Dis Model Mech.* 2019;12(12).

162. Lin YY. Muscle diseases in the zebrafish. *Neuromuscul Disord.* 2012;22(8):673-684.

163. Dickinson ME, Flenniken AM, Ji X, et al. High-throughput discovery of novel developmental phenotypes. *Nature.* 2016;537(7621):508-514.

164. Ochala J, Ravenscroft G, Laing NG, Nowak KJ. Nemaline myopathy-related skeletal muscle α-actin (ACTA1) mutation, Asp286Gly, prevents proper strong myosin binding and triggers muscle weakness. *PLoS One.* 2012;7(9):e45923-e45923.

165. Weninger WJ, Geyer SH, Martineau A, et al. Phenotyping structural abnormalities in mouse embryos using high-resolution episcopic microscopy. *Dis Model Mech.* 2014;7(10):1143-1152.

166. Perez-Garcia V, Fineberg E, Wilson R, et al. Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature.* 2018;555(7697):463-468.

167. de Winter JM, Molenaar JP, Yuen M, et al. KBTBD13 is an actin-binding protein that modulates muscle kinetics. *J Clin Invest.* 2020;130(2):754-767.

168. Kanavy DM, McNulty SM, Jairath MK, et al. Comparative analysis of functional assay evidence use by ClinGen Variant Curation Expert Panels. *Genome Med.* 2019;11(1):77.

169. Lassuthova P, Rebelo AP, Ravenscroft G, et al. Mutations in ATP1A1 Cause Dominant Charcot-Marie-Tooth Type 2. *Am J Hum Genet.* 2018;102(3):505-514.

170. Boycott KM, Campeau PM, Howley HE, et al. The Canadian Rare Diseases Models and Mechanisms (RDMM) Network: Connecting Understudied Genes to Model Organisms. *Am J Hum Genet.* 2020;106(2):143-152.

171. Dickerson JE, Zhu A, Robertson DL, Hentges KE. Defining the role of essential genes in human disease. *PLoS One.* 2011;6(11):e27368.

172. Witting N, Vissing J. Pharmacologic Treatment of Downstream of Tyrosine Kinase 7 Congenital Myasthenic Syndrome. *JAMA Neurol.* 2014;71(3):350-354.

173. Tsoumpra MK, Fukumoto S, Matsumoto T, Takeda Si, Wood MJA, Aoki Y. Peptide-conjugate antisense based splice-correction for Duchenne muscular dystrophy and other neuromuscular diseases. *EBioMedicine.* 2019;45:630-645.

Chapter 2

General methodology

2.1 Preface

In this chapter, I describe the general approach used to identify and prioritise variants from whole exome sequencing (WES) data of patients within the congenital myopathy cohort. In addition, I describe some of the methods used for the functional investigation of the two identified candidate genes: *KLHL40* and *HMGCS1*. A summary of the methodology is illustrated in Figure 2.1.

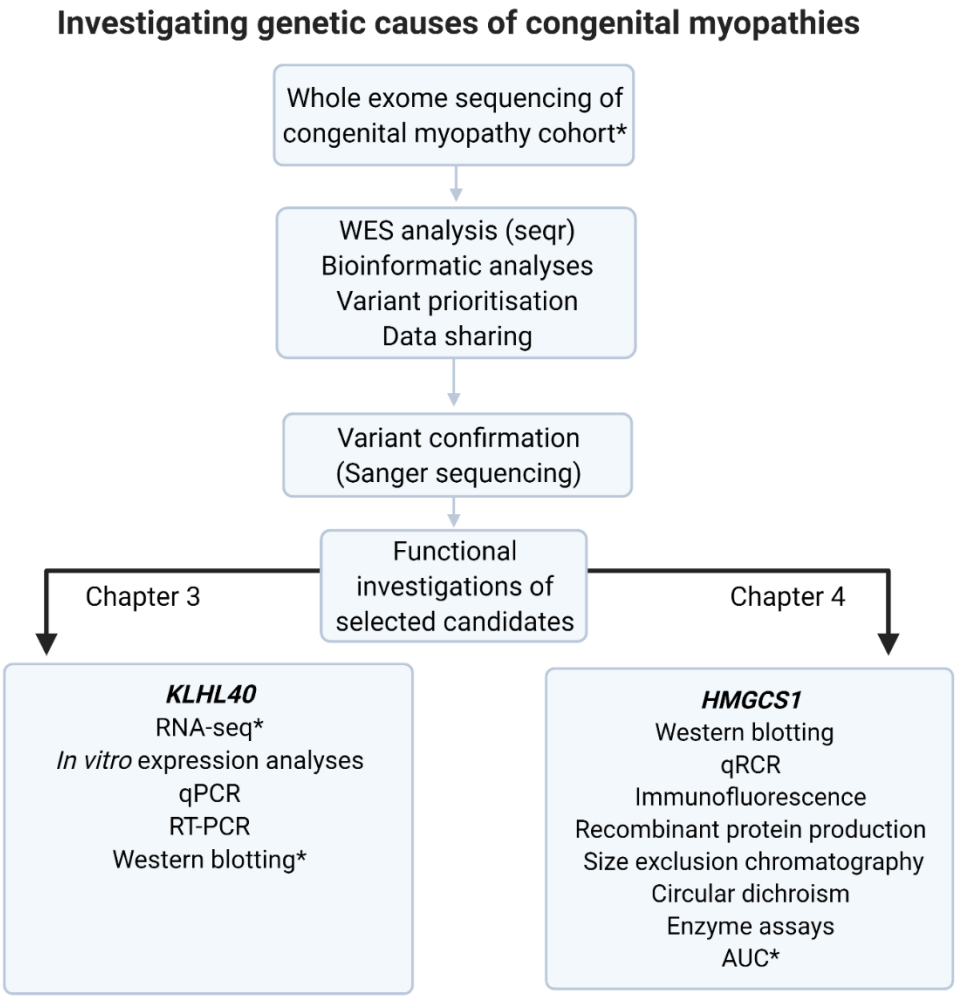


Figure 2.1 Summary of the approach used for whole exome sequencing and functional analyses in this thesis. Abbreviations include WES: whole exome sequencing, RNA-seq: RNA sequencing, qPCR: quantitative PCR, RT-PCR: reverse transcriptase PCR, AUC: analytical ultracentrifugation. Asterisks indicate methods performed by collaborators.

2.2 Cohort details

Patients and available family members were recruited through the rare disease gene discovery study approved by the UWA Human Research Ethics Committee (RA/4/20/1008). Informed consent was provided by patients and/or families. The cohort initially consisted of nine affected patients and available family members from Turkey, Spain, and Australia (n = 21). The patients were clinically diagnosed with a congenital myopathy based on manifestations and histopathological features observed in muscle biopsy (where assessed). For example, presence of nemaline bodies and intranuclear rods. The genetic basis of disease in each of these patients was unresolved despite most cases undergoing screening for known disease genes using diagnostic gene panels.

2.3 Whole exome sequencing data analysis

Whole exome sequencing was performed at the Broad Institute of Harvard and MIT (Boston, USA). WES data from affected patients and available family members were mapped, variant-called and transferred to the Broad institute seqr platform (https://seqr.broadinstitute.org/) for singleton or trio analysis.

2.3.1 Variant filtration and prioritisation

Whole exome sequencing data was analysed following recommended bioinformatic pipelines[1-5]. Single-nucleotide variants and copy number variants were filtered according to read depths and genotype qualities (>20) as well as variant frequencies in healthy population databases (minor allele frequency; MAF <0.01), mainly gnomAD (Genome Aggregation Database v2.1.1; https://gnomad.broadinstitute.org/)[6]. Where familial WES data was available, variants were filtered according to segregation and mode of inheritance using seqr inheritance filters. These include de novo/dominant filters as well as homozygous, compound heterozygous and X-linked recessive filters. Analyses were initially limited to variants in known congenital myopathy genes listed in the 2020 gene table of neuromuscular disorders (http://www.musclegenetable.fr/)[7]. Where filtration strategies yielded no promising results, stringency of seqr filters were reduced in an iterative process.

2.3.2 Analysis of variants using disease gene databases

Candidate genes and variants were interrogated in online human disease databases including ClinVar[8] (https://www.ncbi.nlm.nih.gov/clinvar/), LOVD[9] (Leiden Open Variation Database; https://www.lovd.nl/), and OMIM[10] (Online Mendelian Inheritance in Man; https://omim.org/).

2.3.3 Variant interpretation and classification

Variants in known myopathy genes with relevant genotype-phenotype manifestations were analysed in accordance with the 2015 American College of Medical Genetics and Genomics (ACMG) and Association of Molecular Pathology (AMP) guidelines[11]. Variants in genes not previously associated with disease were explored in the literature and further prioritised using bioinformatic approaches (Section 2.4) to assess the likelihood of causing a myopathic phenotype.

2.4 Bioinformatic analyses

A wholistic approach was taken for the prioritisation of disease gene candidates and variants, considering a number of parameters including population frequency, gene expression, gene constraint information and pathogenicity prediction tools.

2.4.1 Analysis of candidate gene expression from published data

Expression of candidate disease genes was assessed using available data from expression databases such as GTEx[12] (Genotype-Tissue Expression portal; https://gtexportal.org/), NCBI BioProject (PRJEB4337)[13] and FANTOM5[14] (Functional Annotation of the Mammalian genome consortium; https://fantom.gsc.riken.jp/5/). Given the patients had a myopathic phenotype, priority was given to disease gene candidates detected or enriched in skeletal muscle[15].

2.4.2 Analysis of gene intolerance to variation

Gene constraint to loss of function (LoF) and missense variants was assessed using gnomAD constraint metrics scores, following standard interpretation guidelines[16,17]. Predicted loss of function intolerance (pLI) was taken as a predictor for gene constraint to LoF. Genes with pLI ≥ 0.9 were considered to be likely intolerant and under selection against LoF variation[6]. In addition, Z scores[18] (deviation from expectation) were also considered as predictors of gene intolerance to missense and synonymous

variants. Z scores ranged from -5 to 10, with higher (more positive) scores suggesting greater gene intolerance to missense and synonymous variation[16].

To assess gene and protein sequence conservation, Swiss-Prot reference sequences from orthologues and available paralogues were obtained from UniProt (https://www.uniprot.org/)[19]. Multisequence alignments were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

2.4.3 Variant effect prediction tools

Webtools for variant effect prediction included CADD (combined annotation dependent depletion; https://cadd.gs.washington.edu/)[20,21], SIFT (Sorting Intolerant from Tolerant; http://sift.jcvi.org/), POLYPHEN-2 (Polymorphism Phenotyping; http://genetics.bwh.harvard.edu/pph2/)[22] and MutationTaster (http://www.mutationtaster.org/). Variants were also submitted to Varsome (https://varsome.com/) which compiled and compared scores from other pathogenicity prediction tools.

Splice region variants, non-coding variants and synonymous variants were investigated using SpliceAI[23] (https://spliceailookup.broadinstitute.org/) and Human Splicing Finder[24] (http://umd.be/Redirect.html) to predict any impacts on splicing. Variants in 5' untranslated regions (5' UTR) and 3' untranslated regions (3' UTR) were assessed using the RegRNA2.0[25] tool (http://regrna2.mbc.nctu.edu.tw/) and the splicing prediction tools. The Strvctvre[26] (Structural Variant Classifier Trained on Variants Rare and Exonic) tool was used to predict pathogenicity of copy number variants.

Concordance between multiple prediction tools was considered stronger evidence for pathogenicity. Given that the list of genes with rare coding variants returned for each family analysis was small (less than 30 in total for both AR and de novo dominant modes of inheritance), all candidate genes returned were manually curated. Thus, candidates with prediction scores below the thresholds recommended by the respective tools were also investigated and prioritised when other factors such as high gene constraint and enrichment in skeletal muscle suggested potential for pathogenecity[15].

2.4.4 *In silico* protein analysis

Candidate proteins were searched using Uniprot[19] (https://uniprot.org), PDB (Protein Data Bank; www.rcsb.org)[27], PDBe-KB (PDB in Europe Knowledge Base; https://www.ebi.ac.uk/pdbe/pdbe-kb) and the HOPE Web server[28] (http://www.cmbi.ru.nl/hope/). To predict whether variants localise near an interaction site, the ISPRED4[29] webtool (Interaction Site PREDictor version 4; http://ispred4.biocomp.unibo.it) was used. The iPTM[30] (integrated post-translational modifications; https://research.bioinformatics.udel.edu/iptmnet/) tool was used to assess whether variant sites were subject to known post-translational modifications.

2.5 PCR and Sanger confirmation of candidate variants

Candidate single-nucleotide variants were validated by polymerase chain reaction (PCR), and bidirectional Sanger sequencing (AGRF Perth, Australia). Primers of standard length (18-25bp) and GC% content (40-60%) were designed manually using Benchling (https://benchling.com). Primer specificity was assessed using MFEprimer-3.1[31] (https://mfeprimer3.igenetech.com/spec) and the UCSC *in silico* PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr).

Reactions were performed in a final volume of 25 µL and contained a 1X GoTaq G2 master mix (Promega, Madison, WI [#M7832]), 0.4 µM forward primer, 0.4 µM reverse primer and 10 ng DNA. For target regions with high GC-content, 5% v/v DMSO (Sigma-Aldrich) was added. Optimised conditions and primers for PCR reactions are described in the relevant chapters.

Products were resolved on a 1.5% w/v agarose gel in 1X Tris-acetate-EDTA (TAE) buffer containing 0.5 µg/mL EtBr. Products were purified using a MinElute PCR purification kit (QIAGEN; #28004). DNA was quantified on a NanoDrop One spectrophotometer (Thermo Scientific) and bidirectionally sequenced at AGRF. Sequence reactions were prepared in a total volume of 12 µL and consisted of 10 pmol forward or reverse primer and the recommended amount of DNA as per AGRF guidelines.

The FinchTV 1.4.0 software (http://www.geospiza.com) was used to check the quality of ABI sequence chromatograms and correct N-called lower quality bases. Sequences

were aligned to the Ensembl reference sequences for *KLHL40* (ENST00000287777) or *HMGCS1* (ENST00000325110) using the MAFFT algorithm in benchling (https://benchling.com).

2.6 Searching for additional individuals with matching aetiologies

To search for matching individuals with overlapping phenotypes and variants in the same candidate gene, candidates were submitted to the MatchMaker Exchange platform using the *matchbox*[32] tool in seqr. Additionally, candidates were discussed with global clinical networks to determine whether the candidate genes had been previously identified or reported in other families with similar disease features.

2.7 Functional experimental methods

2.7.1 Quantitative PCR

Quantitative PCR was performed following a previously described method[33] to examine expression of *KLHL40 and HMGCS1* in human cells and tissue. Samples included a human embryonic kidney cell line (HEK293), a primary human fibroblast cell line and cultured skeletal muscle myoblasts and myotubes from three donors (Cook Myosite; differentiation method described in section 2.7.3). Tissue included cortex and cauda equina (Victorian Brain Bank), fetal skeletal muscle (Perinatal Pathology, KEMH) and healthy adult skeletal muscle from *in vitro* contracture testing (IVCT).

RNA was extracted from cell pellets using a Qiagen Rneasy mini kit and from tissue using the Qiagen Rneasy Fibrous Tissue Mini Kit. Complementary DNA (cDNA) was synthesised using the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, Waltham, MA).

Quantitative PCR reactions were performed in 10 µL volumes containing a 2X Rotor-Gene SYBR Green PCR master mix (Qiagen; 204076), 1 µL diluted cDNA and 0.8 µM of forward and reverse primer (see Chapters 3 and 4 for primer sequences). A standard curve from serially diluted positive control cDNA was generated to validate efficiency of primers used for each gene. Cycling was performed on a Rotor-Gene Q real-time PCR cycler (Qiagen) using the following conditions: 95°C (5 min), 45 cycles

of 95°C (10 s) and 60°C (15 s), followed by melt curve analysis (60 – 95°C; +1°C per cycle). Data were normalised using the delta Ct method[34], compared to the geometric mean from two reference genes (*TBP* and *EEF2*).

2.7.2 Western blotting

Lysate preparations were performed as previously described, with slight adjustments[35]. Tissue samples were cryo-sectioned at -20°C into 10 µM slices to aid mechanical homogenisation. Tissue and frozen cell pellets were suspended in lysis buffer containing 2% w/v SDS, 125 mM Tris (pH 6.8), and 1x PIC (Halt™ Protease Inhibitor Cocktail; Thermo Scientific). Suspensions were freeze thawed three times in dry ice before sonicating at 15 amps for 3 s twice, resting on wet ice in between. Lysates were centrifuged at 18,000 g for 3 min, and supernatants retained. Protein concentrations were determined using a BCA protein assay (Pierce; #23227). Protein extracts (5-40 µg) and recombinant protein (5 ng) were prepared in loading buffer (10% w/v SDS, 312.5 mM Tris, pH 6.8, 50% v/v glycerol, bromophenol blue saturated solution, 50 mM DTT and 1x PIC). Samples were separated on a 4-12% w/v Bis-Tris NuPAGE Novex polyacrylamide gel at 150 V for 2 h. Proteins were transferred to polyvinylidene difluoride transfer membranes (Thermo Fisher Scientific) for 2 h at 300 mA and blocked with PBS + 0.1% v/v Tween-20 and 5% w/v skim milk. Membranes were probed with antibodies described in Chapters 3 and 4 and incubated overnight at 4°C. Membranes were washed thrice in wash buffer (PBS + 0.1% v/v Tween-20) for 1 h and incubated with a secondary goat anti-rabbit horseradish peroxidase antibody (Sigma; #A0545) for 1 h at room temperature. Membranes were rewashed twice as above and imaged by chemiluminescent detection (Pierce ECL Plus kit; #32132) using the Invitrogen iBright FL1000 imaging system (Thermo Fisher Scientific). Membranes were re-probed with an anti-GAPDH antibody (Sigma; #G8795; [1:10,000]) as a loading control. Relative protein abundance between samples was measured using ImageJ[36].

2.7.3 Differentiation of primary human skeletal muscle cells

To measure gene expression and protein abundance of candidates in myoblasts and myotubes, primary human skeletal muscle cells (skMDCs; Cook MyoSite) were cultured from two healthy female donors (32F and 42F) and a healthy male donor (18M). Cells were reanimated in MyoTonic Basal Medium supplemented with 1X

Myotonic Growth Supplement (Cook MyoSite), 10% v/v fetal bovine serum (FBS) and 1% w/v penicillin/streptomycin (Gibco). At 60% confluency, myoblasts were plated into a 6-well plate (2.5 x 10^5 cells/well) coated with Matrigel and regrown to ~90% confluency. Differentiation to myotubes was induced by changing the media to DMEM/F12 (HPMR) supplemented with 1X GlutaMAX supplement (Gibco), 1X N2 supplement (Gibco), 1X ITS-X (Gibco), and 1% w/v penicillin/streptomycin (Gibco). Cells were maintained under standard conditions, with a media change every two days. Myoblasts and myotubes were harvested at timepoints 0, 2, 4, 6, 8 and 12.

2.7.4 *In vitro* expression analyses

Mammalian and bacterial expression vectors encoding the coding sequences for genetic candidates were purchased from Genscript (https://www.genscript.com/). These were used for *in vitro* expression analyses and recombinant protein investigations as described in Chapters 3 and 4.

2.8 Statistical analyses and figures

GraphPad Prism 9.0.1 (La Jolla, USA) was used for graphical and statistical analyses. Schematic figures were designed using BioRender (BioRender.com).

2.9 References

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

1. Cassandrini D, Trovato R, Rubegni A, et al. Congenital myopathies: clinical phenotypes and new diagnostic tools. *Ital J Pediatr.* 2017;43(1):101.

2. Gonorazky HD, Bönnemann CG, Dowling JJ. The genetics of congenital myopathies. *Handb Clin Neurol.* 2018;148:549-564.

3. North KN, Wang CH, Clarke N, et al. Approach to the diagnosis of congenital myopathies. *Neuromuscul Disord.* 2014;24(2):97-116.

4. Nykamp K, Anderson M, Powers M, et al. Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet Med.* 2017;19(10):1105-1117.

5. Lek M, MacArthur D. The Challenge of Next Generation Sequencing in the Context of Neuromuscular Diseases. *J Neuromuscul Dis.* 2014;1(2):135-149.

6. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.

7. Benarroch L, Bonne G, Rivier F, Hamroun D. The 2020 version of the gene table of neuromuscular disorders (nuclear genome). *Neuromuscul Disord.* 2019;29(12):980-1018.

8. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42(Database issue):D980-D985.

9. Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat.* 2011;32(5):557-563.

10. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* 2019;47(D1):D1038-d1043.

11. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-423.

12. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585.

13. Fagerberg L, Hallström BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics.* 2014;13(2):397-406.

14. Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 2015;16(1):22.

15. Anderson D, Baynam G, Blackwell JM, Lassmann T. Personalised analytics for rare disease diagnostics. *Nat Commun.* 2019;10(1):5274.

16. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-291.

17. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv.* 2019:531210.

18. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet.* 2014;46(9):944-950.

19. Consortium TU. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2018;47(D1):D506-D515.

20. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-315.

21. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47(D1):D886-d894.

22. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;76(1):7.20.21-27.20.41.

23. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell.* 2019;176(3):535-548.e524.

24. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 2009;37(9):e67.

25. Chang T-H, Huang H-Y, Hsu JB-K, Weng S-L, Horng J-T, Huang H-D. An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics.* 2013;14(2):S4.

26. Sharo AG, Hu Z, Brenner SE. StrVCTVRE: A supervised learning method to predict the pathogenicity of human structural variants. *bioRxiv.* 2020:2020.2005.2015.097048.

27. Burley SK, Berman HM, Bhikadiya C, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* 2018;47(D1):D464-D474.

28. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics.* 2010;11:548.

29. Savojardo C, Fariselli P, Martelli PL, Casadio R. ISPRED4: interaction sites PREDiction in protein structures with a refining grammar model. *Bioinformatics.* 2017;33(11):1656-1663.

30. Huang H, Arighi CN, Ross KE, et al. iPTMnet: an integrated resource for protein post-translational modification network discovery. *Nucleic Acids Res.* 2018;46(D1):D542-D550.

31. Wang K, Li H, Xu Y, et al. MFEprimer-3.0: quality control for PCR primers. *Nucleic Acids Res.* 2019;47(W1):W610-W613.

32. Arachchi H, Wojcik MH, Weisburd B, et al. matchbox: An open-source tool for patient matching via the Matchmaker Exchange. *Hum Mutat.* 2018;39(12):1827-1834.

33. McNamara EL, Taylor RL, Clayton JS, et al. Systemic AAV8-mediated delivery of a functional copy of muscle glycogen phosphorylase (Pygm) ameliorates disease in a murine model of McArdle disease. *Hum Mol Genet.* 2020;29(1):20-30.

34. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002;3(7):research0034.0031.

35. Ravenscroft G, Nowak KJ, Jackaman C, et al. Dissociated flexor digitorum brevis myofiber culture system--a more mature muscle culture system. *Cell Motil Cytoskeleton.* 2007;64(10):727-738.

36. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9(7):671-675.

Chapter 3

A *KLHL40* 3' UTR splice-altering variant causes milder NEM8

3.1 Preface

A manuscript arising from this chapter is in preparation for submission to *Human Molecular Genetics*, as an article titled *A KLHL40 3' UTR splice-altering variant causes milder NEM8*. As such, this chapter has been formatted for *Human Molecular Genetics*.

An authorship attribution statement is provided in Appendix I.

This chapter also resulted in an abstract that was published in *Neuromuscular Disorders* as part of a poster presentation for World Muscle Society (Dofash *et al*., 2021, DOI:10.1016/j.nmd.2021.07.059). A copy of this abstract can be found in Appendix II.

3.2 Abstract

Nemaline myopathy 8 (NEM8) is typically a severe autosomal recessive disorder associated with variants in the kelch-like family member 40 gene (*KLHL40*). To date, only protein-altering pathogenic variants in *KLHL40* have been implicated in NEM8. Common features include fetal akinesia/hypokinesia, fractures, contractures, dysphagia, respiratory failure, and neonatal death. Here, we describe a 26-year-old man with relatively mild NEM8. He presented with hypotonia and bilateral femur fractures at birth, later developing bilateral Achilles' contractures, scoliosis, and elbow and knee contractures. He had walking difficulties throughout his childhood and became wheelchair bound from age 13 after prolonged immobilization. Muscle MRI at age 13 indicated prominent fat replacement in his pelvic girdle, posterior compartments of thighs, vastus intermedius and legs. Muscle biopsy revealed nemaline bodies and intranuclear rods. RNA sequencing and western blotting of patient skeletal muscle indicated significant reduction in *KLHL40* mRNA and protein respectively. Using gene panel screening, exome sequencing and RNA sequencing, we identified compound heterozygous variants in *KLHL40*; a truncating 10.9 kb deletion *in trans* with a likely pathogenic variant (c.*152G>T) in the 3' untranslated region (UTR). Computational tools SpliceAI and Introme predicted the c.*152G>T variant created a cryptic donor splice site. RNA-seq and *in vitro* analyses indicated that the c.*152G>T variant induces multiple splicing events, predominantly a 78 bp cryptic intron in the 3' UTR. The introduction of 3' UTR introns likely degrades *KLHL40* mRNA by nonsense mediated decay, a mechanism that may be underrecognised in Mendelian disease. We encourage consideration of this mechanism during disease gene screening.

Keywords: 3' UTR, nonsense mediated decay, RNA sequencing, *KLHL40*, nemaline myopathy

Abbreviations: IGV = Integrative Genomics Viewer; NEM8 = nemaline myopathy 8; RT-PCR = reverse transcriptase polymerase chain reaction

3.3 Introduction

Nemaline myopathies (NEM) are a clinically and genetically heterogenous group of congenital myopathies typically characterised by the presence of nemaline bodies within muscle fibres[1]. Clinical symptoms include muscular hypotonia and weakness, often accompanied by respiratory insufficiency[2]. The age of onset, disease progression and severity are variable and largely depend on the underlying genetic cause[2]. To date, 13 genes have been associated with NEM[3].

One common severe subtype of NEM is nemaline myopathy 8 (NEM8; OMIM# 615348), an autosomal recessive disorder associated with biallelic variants in the kelch-like family member 40 gene (*KLHL40*)[4]. Almost all patients with NEM8 present with severe disease[4,5]. Characteristic features include fetal akinesia or hypokinesia, fractures, contractures, facial involvement, dysphagia, respiratory failure, and neonatal death (average age at death in a NEM8 cohort is 5 months of age)[4]. To date, only two patients have been reported with relatively mild NEM8, one of whom responded beneficially to acetylcholinesterase inhibitors[6,7].

KLHL40 is a member of the kelch-repeat-containing protein superfamily and is involved in a range of interactions that regulate skeletal muscle myogenesis and promote skeletal muscle maintenance[8,9]. KLHL40 is reported to bind the E2F1-DP1 complex to regulate myogenesis[9]. In addition, KLHL40 binds NEB and LMOD3 to stabilise the thin filament[8]. To date, pathogenic variants that change the *KLHL40* open reading frame have been implicated in NEM8[4,5]. These include protein truncating variants including frameshift, essential splice site and nonsense variants[4]. Pathogenic missense variants have also been reported and are predicted to destabilise KLHL40 by disrupting intramolecular interactions[4]. KLHL40 deficiency consequently destabilises the thin filament and causes nemaline myopathy[8]. NEM8 patients typically show reduced or absent KLHL40 in skeletal muscle by western blotting[4-6].

In this study, we describe a Spanish patient with mild NEM8 who had diminished *KLHL40* transcript and protein abundance. We report a novel 3' UTR splice variant in *KLHL40* which introduces an intron 152 bp downstream the termination codon and likely induces nonsense mediated decay. This mechanism may be underrecognised in Mendelian disease.

3.4 Patient and methods

This project was approved by the University of Western Australia Human Research Ethics Committee (RA/4/20/1008) and the Curtin University Human Research Ethics Office (HRE2019-0566). Written informed consent was provided by the family.

3.4.1 Patient details

The patient was clinically diagnosed with nemaline myopathy on the basis of histopathological findings on skeletal muscle biopsy. Clinical assessments, MRI, muscle biopsy analysis and western blotting of patient muscle biopsy were performed by collaborators at the Biomedicine Institute of Sevilla and the Department of Neurology at Virgen del Rocío University Hospital in Sevilla, Spain.

3.4.2 Methods for histology and electron microscopy

Muscle biopsy was obtained from the quadriceps muscle, immediately frozen by standard methods, and processed following the standard procedures. Routine histochemical techniques were performed on 7 μM transverse sections of frozen muscle, including hematoxylin and eosin, modified Gomori trichrome, oil red O, periodic acid-Schiff, nicotinamide adenine dinucleotide-tetrazolium reductase (NADH-TR), succinate dehydrogenase (SDH), cytochrome C oxidase (COX), and adenosine triphosphatase (ATPase) pH 9.4, 4.6, and 4.3.

For ultrastructural studies a small fragment of the muscle was fixed in 2.5% w/v glutaraldehyde solution, postfixed in 1% w/v osmium tetroxide, and embedded in epoxy resin. Semithin sections were stained with 1% toluidine blue. Ultrathin sections were mounted on copper grids and examined with a Zeiss Libra 120 transmission electron microscope (Carl Zeiss NTS GmbH, Oberkochen, Germany).

3.4.3 Western blotting

Frozen muscle samples were homogenized in RIPA buffer (20 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% v/v IGEPAL, 0.1% w/v SDS) containing protease inhibitor cocktail (Roche). The lysates were centrifuged at 13,000 rpm at 4^{o}C for 20 min. The supernatant was collected. The protein lysates were separated on 10% w/v SDS-PAGE gels and transferred onto PDVF membranes (Millipore). Western blot analysis of equal-protein loading was performed with the following primary antibodies: rabbit polyclonal anti-KLHL40 (Sigma-Aldrich; #HPA024463; [1:500])

and rabbit polyclonal anti-GAPDH (Sigma-Aldrich; [1:2000]). Immunoreactivity was detected with secondary antibodies conjugated to horseradish peroxidase (1:5000; Jackson Immuno Research) and developed with SuperSignal West Femto (Thermo Fisher Scientific) using an ImageQuant LAS 4000 MiniGold System (GE Healthcare Life Sciences).

3.4.4 Genetic analyses

3.4.4.1 Neuromuscular disease gene panel

DNA from the proband was sequenced on the neuromuscular disease gene panel at PathWest using Illumina sequencing chemistry as previously described[10,11]. Base calling, mapping, and variant calling were performed as previously described[10,11].

3.4.4.2 Whole exome sequencing

Illumina whole exome sequencing (WES) was performed at the Broad Institute of Harvard and MIT as previously described[11]. Exome data were analysed in seqr[12] (https://seqr.broadinstitute.org).

3.4.4.3 Variant interpretation and confirmation

Variants were analysed and interpreted following standard bioinformatic approaches[13,14]. Candidate single nucleotide variants were validated by bidirectional Sanger sequencing (Australian Genome Research Facility (AGRF), Perth)[10]. Sanger chromatograms were aligned to a reference sequence in benchling (benchling.com) using the MAFFT algorithm.

3.4.4.4 In silico variant prediction analysis

Introme[15] (https://github.com/CCIACB/introme) and SpliceAI[16] (https://spliceailookup.broadinstitute.org/) were used to predict the effect of variants on splicing.

3.4.4.5 RNA sequencing

RNA extraction from patient skeletal muscle and RNA sequencing were performed at the Department of Diagnostic Genomics (PathWest, Perth) as previously described[17]. RNA-seq data were visualised with the Integrative Genomics Viewer (IGV)[18]. Data were also assembled and analysed using MINTIE[19], a reference-free RNA-seq analysis pipeline that maps reads *de novo* to identify novel structural and splice variants[20]. Muscle RNA-seq FASTQ data from five additional patients were used as controls for MINTIE.

3.4.5 *In vitro* expression analyses

3.4.5.1 Expression constructs

Mammalian *pcDNA3.1* expression constructs (pcDNA3.1) were synthesised by Genscript. These constructs contained the *KLHL40* coding sequence (ENST00000287777) with the wild-type or mutant 3' UTR encoding either of two variants: c.*152G>T or c.*151_*228del.

3.4.5.2 Cell culture

Human embryonic kidney cell lines (HEK293FT) were maintained in Dulbecco's Modified Eagle Medium (DMEM; Gibco) supplemented with 10% v/v fetal bovine serum (FBS) and 1% w/v Penicillin/Streptomycin (Gibco). At 50-60% confluency, expression constructs were transfected following the Viafect kit protocol (Promega) at a 4:1 ratio (Viafect transfection reagent: DNA). Cultures were maintained at 37°C with 5% CO_2. A pmaxGFP vector was transfected alongside to estimate transfection efficiency. Cells were harvested at 72 h post-transfection and pellets snap frozen and stored at -80°C until required for RNA and protein extraction.

3.4.5.3 RNA extraction and quantitative PCR

Total RNA was extracted from HEK293FT pellets using the Rneasy mini kit protocol (Qiagen). Complementary DNA (cDNA) was synthesised from 1 µg RNA using the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, Waltham, MA). To investigate relative transcript expression levels of wildtype *KLHL40*, *KLHL40*$^{c.*152G>T}$ and *KLHL40*$^{c.*151\_*228del}$, qPCR was performed as previously described[21]. Reactions were performed in 10 µL volumes containing a 2X Rotor-Gene SYBR Green PCR master mix (Qiagen; 204076), 1 µL diluted cDNA and 0.8 µM of the following forward and reverse primers: KLHL40_FD5_6 (5' GGAGGTATAACGAGGAGGAGAA-3', bridges the junction of exon 5 and exon 6), and KLHL40_RV6 (5'-CTGAGCTGGTCACATCTTAGTC-3'). Reactions were performed in technical quadruplicates for each biological replicate (n = 3) and Ct values were averaged. Expression was normalised using the delta Ct method[22], comparing to the geometric mean of two reference genes (*TBP* and *EEF2*)[21]. Statistical significance was measured using a two-tailed unpaired t-test (n = 3) on GraphPad Prism 9.0.1 (La Jolla, USA).

3.4.5.4 Reverse transcriptase PCR

RT-PCR was performed to confirm the occurrence of mis-splicing events in the 3' UTR region. Reactions were performed in 25 µL volumes containing a 1X GoTaq G2 master mix (Promega, Madison, WI [#M7832]), 0.4 µM forward primer (KLHL40_UTR_F; 5'-CCAGCTCAGGCAGACTGAAC-3'), 0.4 µM reverse primer (KLHL40_UTR_R; 5- GACACCAGATGGAGAGCAGAG-3') and 7.5 ng cDNA. Touchdown PCR cycling conditions were: 5 min at 98°C; 15 cycles of 15 s at 98°C, 10 s at 65°C-57.5°C (-0.5°C per cycle), and 15 s at 72°C; 30 cycles of 15 s at 98°C, 10 s at 57°C, and 15 s at 72°C; and a final cycle of 5 min at 72°C. Products were resolved on 2% w/v agarose gel in 1X TAE at 95 V for 1 h. PCR products were gel excised and purified using the QIAquick gel and PCR purification kit. DNA was quantified on a NanoDrop One spectrophotometer (Thermo Scientific) and Sanger sequenced at AGRF (Perth, Australia). Sanger chromatograms were aligned to a reference sequence in benchling (benchling.com) using the MAFFT algorithm.

3.4.5.5 Protein extraction and western blotting

Frozen cell pellets were suspended in lysis buffer containing 2% w/v SDS, 125 mM Tris (pH 6.8), and 1x PIC (Halt™ Protease Inhibitor Cocktail; Thermo Scientific) and homogenised by sonication. Protein concentrations were determined by a BCA protein assay (Pierce; #23227). Lysates were prepared in loading buffer containing 10% w/v SDS, 312.5 mM Tris (pH 6.8), 50% v/v glycerol, bromophenol blue saturated solution, 50 mM DTT and 1x PIC. Western blotting was performed as previously described[4]. Lysates (500 ng) were separated on 4-12% w/v Bis-Tris NuPAGE Novex polyacrylamide gels and transferred to polyvinylidene difluoride transfer membranes (Thermo Fisher Scientific). After 1 h blocking with PBS + 0.1% v/v Tween-20 and 5% w/v skim milk, membranes were incubated with a Human Protein Atlas rabbit polyclonal KLHL40 (KBTBD5) antibody (Sigma-Aldrich; #HPA024463; [1:2500]) overnight at 4°C and a secondary goat anti-rabbit horseradish peroxidase antibody (Sigma; #A0545) at room temperature for 1 h. Membranes were imaged by chemiluminescent detection (Pierce ECL Plus kit; #32132) using the Invitrogen iBright FL1000 imaging system (Thermo Fisher Scientific). Membranes were also blotted with an anti-GAPDH antibody (Sigma; #G8795; [1:10,000]) as a loading control. Relative KLHL40 abundance was quantified using ImageJ[23].

3.5 Results

3.5.1 Clinical features and pathology

The patient is a male born to non-consanguineous parents, with no relevant family history of disease. He was born hypotonic with bilateral femur fractures. He walked at the age of 30 months, but never jumped or ran. He lost independent ambulation at age 13, after prolonged immobilization after Achilles tendon lengthening. From childhood he had scoliosis, a myopathic face and ogival palate, and weakness of bilateral orbicular oculi muscles. He was last examined at age 24, being able to stand with bilateral support. At last examination, he had weakness of proximal and distal muscles of upper and lower limbs and showed obvious atrophy of intrinsic hand muscles. Deep tendon reflexes were absent, with bilateral Achilles tendon contractures as well as elbow contractures and cervical and dorsal rigid spine.

A muscle MRI showed widespread fat replacement of pelvic girdle muscles, thighs, and lower legs (Figure 3.1A-C). At the thigh level, the fat replacement was particularly severe involving posterior compartments as well as *rectus anterioris*.

Light microscopy examination revealed moderate and patchy fat replacement, without relevant fibrosis, with mild variability in fibre size. Necrotic or regenerating fibres were not seen. Numerous eosinophilic birefringent structures, rod or round shaped were seen forming clusters in the centre of the sarcoplasm or subsarcolemmal regions (Figure 3.1D). There were internal nuclei. Some intranuclear rods were identified on the semithin sections (Figure 3.1F). Gomori trichrome confirmed the presence of abundant rods with irregular distribution in the sarcoplasm (not shown). Histoenzymatic techniques showed an irregular intermyofibrillar pattern, with presence of pseudo-lobulated fibres and some core-like areas devoid of oxidative activity harbouring the clusters of rods. ATPase stains showed 100% of the fibres were type 1 (Figure 3.1E).

Electron microscopy confirmed the presence of numerous electron-dense rod-shaped structures corresponding to nemaline rods in most muscle fibres (Figure 3.1G). Rods were irregularly distributed throughout the fibre, clustered in centre, forming subsarcolemmal clusters in periphery. They often ran parallel to the long axis of the sarcomere and showed continuity with Z-disks.

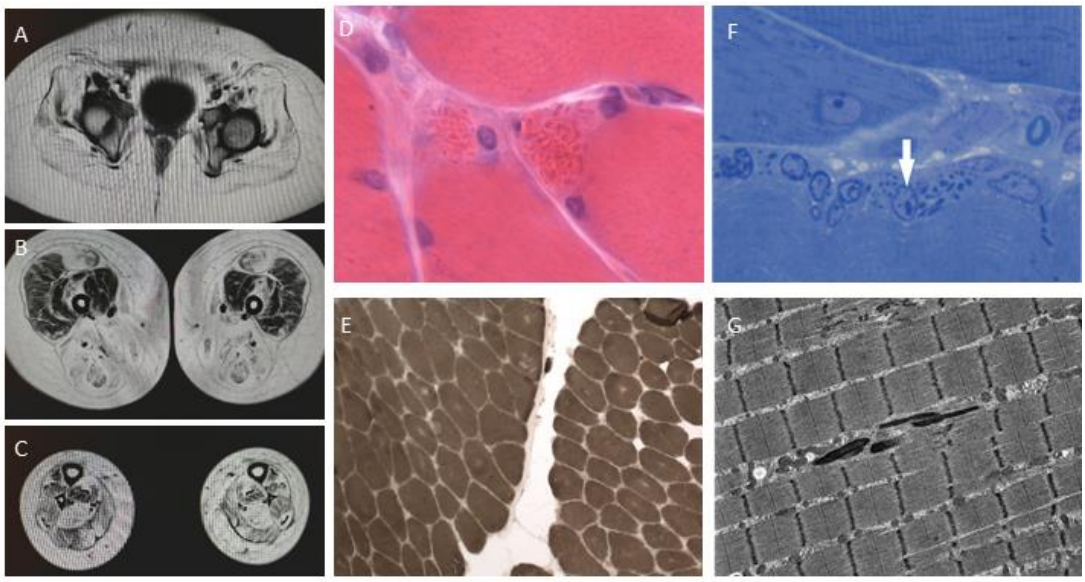


Figure 3.1 Muscle pathology identified in the patient with a milder form of NEM8. (A-C) Magnetic resonance imaging shows extensive fatty replacement of muscles of pelvic girdle (A), thighs (B), with most prominent involvement of posterior compartment (B) and legs (C). (D-G) Muscle biopsy sections: haematoxylin and eosin stain showing clusters of nemaline rods (D), ATPase 4.3 stain demonstrating all fibres in the sample are type 1 (E), Semithin slide stained with toluidine blue showing an intranuclear rod (F, arrow), and Electron microscopy displaying electron-dense rod-shaped structures that correspond to nemaline bodies (G).

3.5.2 Genomic, transcriptomic and protein investigations

Gene panel sequencing and WES identified a heterozygous 10.9 kb deletion in *KLHL40* (hg19; chr3:42729537-42740458) in the patient (Figure 3.2A). This deletion spanned exons 2-6 of the six exons of *KLHL40* (ENSG00000157119) and results in a complete loss of function allele. The deletion was absent in gnomAD (v2.1) and was only present in this patient in the 10,000 alleles in the Broad structural variant callset.

Western blotting indicated very low abundance of KLHL40 in patient skeletal muscle compared to healthy controls (Figure 3.2B). Although deficiency of KLHL40 is sufficient for diagnosis of NEM8, a second variant *in trans* would be expected given the recessive nature of KLHL40-related nemaline myopathy[4]. Following re-analysis of WES data to include rare non-coding variants, a single nucleotide variant (c.*152G>T) was identified in the 3' UTR of *KLHL40 in trans* with the deletion (Figure 3.2A). The variant appeared homozygous given the complete absence of the

3’ UTR in the other allele with the exon 2-6 deletion. The c.*152G>T variant was present in gnomAD (v3.1.2) at a frequency of 3.94 $x10^{-5}$ (6/152204 alleles; no homozygotes). Five of six alleles were from the European (non-Finnish) population. The sixth allele was from the Latino/Admixed American population. The c.*152G>T variant was confirmed by bidirectional Sanger sequencing (Figure 3.2C). Parental DNA was not available.

The SpliceAI tool[16] predicted that the c.*152G>T variant creates a weak donor splice site which results in a 255 bp cryptic intron in the 3’ UTR of *KLHL40*. The prediction score (0.17) was below the Δ score cut-off (0.2) suggested in the SpliceAI publication[16]. The Introme tool[15] also predicted the variant creates a weak donor splice site and a 255 bp intron with a score above the threshold (Introme score 0.76, threshold 0.54).

Analysis of patient RNA-seq data showed significantly lower levels of *KLHL40* transcripts compared to tissue-matched controls (Figure 3.2D). Patient RNA-seq coverage ranged between 1-38 reads (Figure 3.2D). Control RNA-seq coverage ranged between 5-927 reads (Figure 3.2D). The low coverage of *KLHL40* appeared specific in the patient sample, as IGV analyses indicated comparable coverage of 10,670 other genes to controls. Normal splicing of *KLHL40* exons 1-6 was observed at low frequency (Figure 3.2D). While splicing of the predicted 255 bp cryptic intron was not detected by RNA-seq, a distinct drop in coverage of 78 bp was identified neighbouring the 3’ UTR variant site (Figure 3.2E,F) suggesting the formation of a 78 bp cryptic intron which would result in a secondary variant in the 3’ UTR of *KLHL40*; c.*151_*228del. MINTIE alignment of the RNA-seq reads confirmed the 78 bp deletion (hg19; chr3:42733635-42733714) and indicated the presence of a new exon junction with a variant allele frequency of 0.94. This implies that 94% of the reads contained the cryptic 78 bp deletion. Some RNA-seq reads also suggested the formation of other cryptic deletions, however these were not detected by MINTIE alignment.

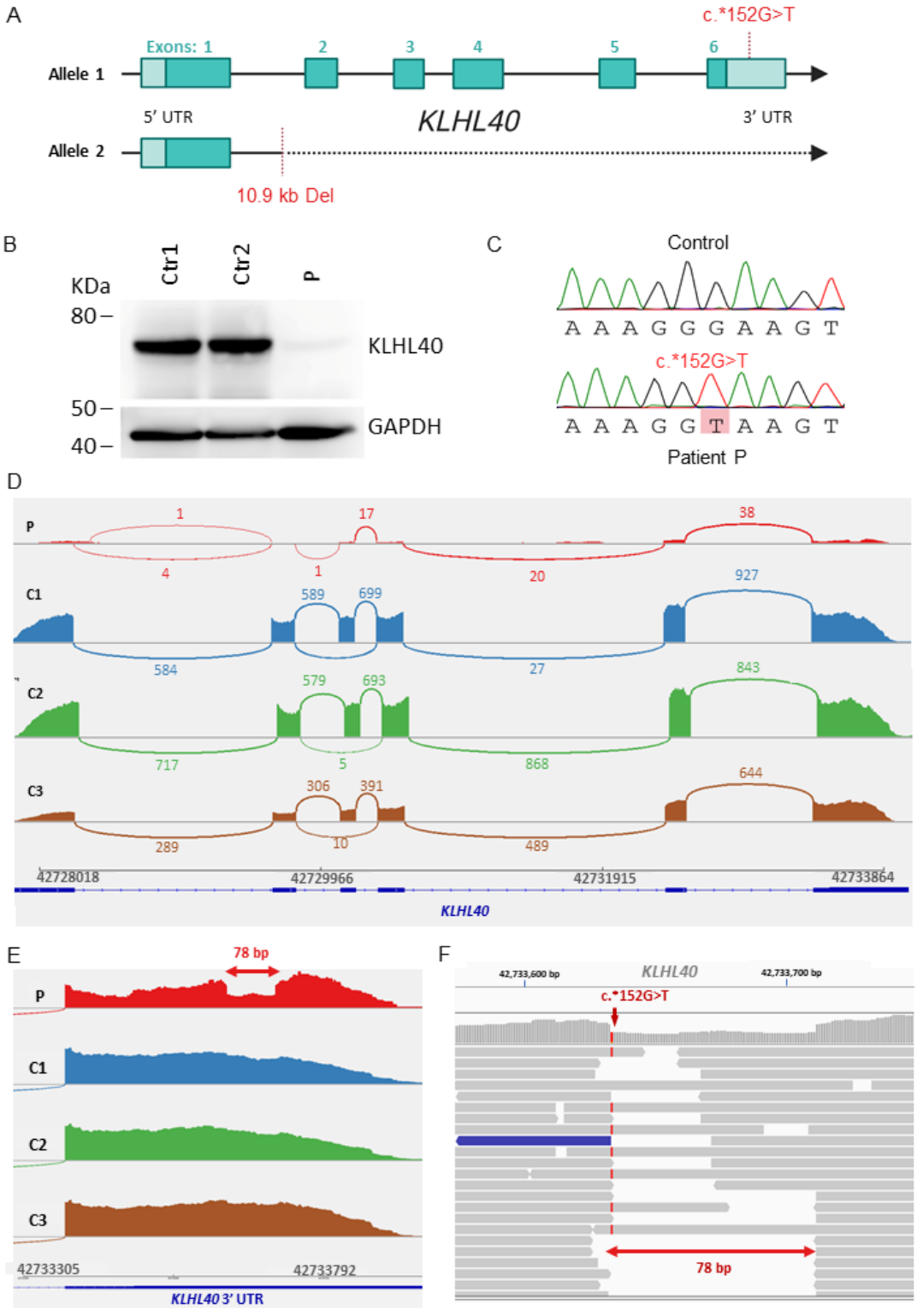


Figure 3.2 Alterations of *KLHL40* at the DNA, mRNA, and protein level of the patient. (A) Schematic representation of the *KLHL40* gene with the 5' and 3' UTRs (light blue), 6 coding exons (cyan) and the compound heterozygous variants in the patient. Figure designed with BioRender.com. (B) Western blotting for KLHL40 (top panel) from patient (P) and control (Ctrl1-2) muscle biopsies. Western blotting for

GAPDH (bottom panel) as a loading control. (C) Sanger confirmation of the c.*152G>T variant in Patient P (bottom) compared to an unaffected control (top), the variant appears homozygous due to deletion of the other allele. (D-F) RNA-seq data of muscle RNA from patient P and controls C1-C3 aligning to *KLHL40*. (D) Sashimi plots for patient and controls. (E) Close up of Sashimi plots at the 3' UTR region of *KLHL40*. (F) Patient RNA-seq reads spanning the 3' UTR region with the c.*152G>T variant. Sequences viewed using the integrative Genomics Viewer (IGV).

3.5.3 *In vitro KLHL40* expression analyses

To functionally demonstrate the effect of the c.*152G>T variant on *KLHL40* expression and abundance, we transiently expressed *KLHL40* with either the wild-type (WT) or mutant (c.*152G>T) 3' UTR in HEK293FT cells. We also generated a 78bp deletion mutant (c.*151_*228del) to mimic the cryptic intron identified by RNA-seq.

qPCR analysis indicated significantly lower expression of the $KLHL40^{c.*152G>T}$ transcript (52 ± 3) relative to wildtype $KLHL40^{WT}$ (182 ± 47) and $KLHL40^{c.*151\_*228del}$ (119 ± 37) confirming the association of *KLHL40* c.*152G>T with reduced *KLHL40* mRNA levels (Figure 3.3A). The difference in expression between $KLHL40^{WT}$ and $KLHL40^{c.*152G>T}$ was statistically significant using a two-tailed unpaired t-test (P = 0.0086). Interestingly, the relative expression of $KLHL40^{c.*151\_*228del}$ was significantly higher than $KLHL40^{c.*152G>T}$ (P = 0.0351; Figure 3.3A) despite harbouring the 78 bp deletion in the 3' UTR. The difference in expression between $KLHL40^{WT}$ and $KLHL40^{c.*151\_*228del}$ was not significant.

Western blotting data from *in vitro* expression studies similarly indicated that levels of $KLHL40^{c.*152G>T}$ were significantly lower than $KLHL40^{c.*151\_*228del}$ (21.7%, P = 0.0242) and $KLHL40^{WT}$ (15.3%, P = 0.0348) thus confirming an association of *KLHL40* c.*152G>T with reduced KLHL40 abundance (Figure 3.3B,C).

RT-PCR analysis of $KLHL40^{c.*152G>T}$ revealed three prominent products (Figure 3.3D). The largest product (a) was of comparable size to $KLHL40^{WT}$ (551 bp) and corresponds to the un-spliced transcript with the single nucleotide variant. The second product (b) was of comparable size to $KLHL40^{c.*151\_*228del}$ (473 bp) and likely corresponds to a spliced transcript containing the 78 bp deletion (Figure 3.3D). Interestingly, the smallest product (c; ~300 bp) appears to correspond to the cryptic

transcript predicted by SpliceAI and Introme. Sequencing alignments indicated this product contained a 255 bp deletion spanning positions c.*151 to c.*405 (Supplementary Fig 3.1). This suggests that in addition to the 78 bp intron, the c.*152G>T variant may also induce splicing of the 255 bp cryptic intron which results in a second deletion; c.*151_*405del.

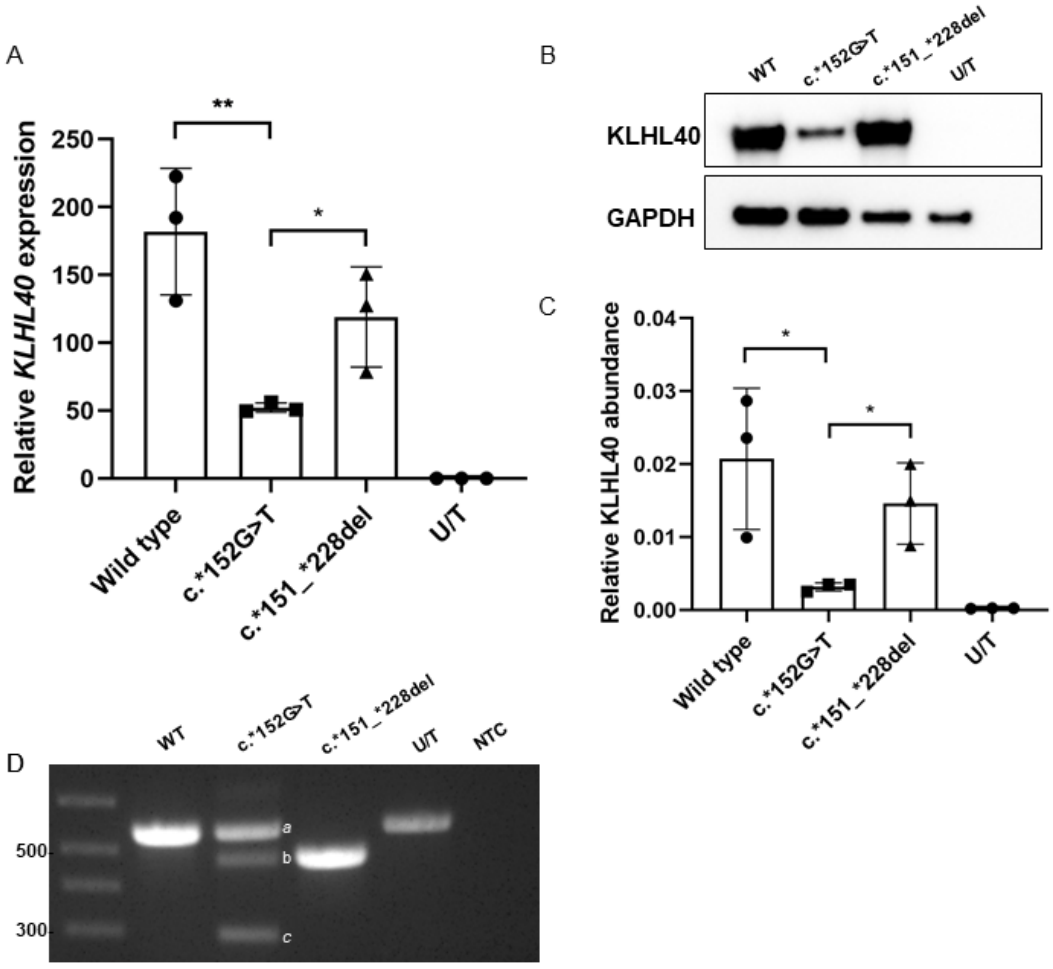


Figure 3.3 *In vitro KLHL40* expression studies. (A) qPCR data from Hek293FT cDNA expressing pcDNA3.1 encoding *KLHL40* wildtype (WT), c.*152G>T, and c.*151_*228del variant transcripts. Expression compared with endogenous *KLHL40* from untreated Hek293 cells (U/T). Reactions performed in biological triplicates (*n* = 3). Expression normalised to *EEF2* and *TBP*. Error bars indicate ± SD. (B) Representative western blot for KLHL40 (top panel) from HEK293FT cells expressing wildtype (WT) and mutant (c.*152G>T, and c.*151_*228del) KLHL40 protein. Western blotting for GAPDH (bottom panel) as a loading control. (C) Relative abundance of KLHL40 wildtype and mutant proteins quantified by ImageJ from KLHL40 western blots (*n* = 3). Relative abundance normalised to GAPDH. (D) RT-

PCR and agarose gel electrophoresis of *KLHL40* 3' UTR products. Significance for (A) and (C) was determined by a two-tailed unpaired t-test (*$P < 0.05$, **$P < 0.01$).

3.6 Discussion

We describe a patient with NEM8 with two novel variants in *KLHL40,* a 10.9 kb deletion spanning exons 2-6 and a single nucleotide variant (c.*152G>T) in the 3' UTR. Together, these biallelic variants fit with the recessive inheritance pattern of NEM8[4]. Notably, the patient's clinical phenotype is milder than most NEM8 cases described in the literature[4,5,24]. He represents one of very few cases who survived into adolescence[4] and remained ambulant, albeit with walking difficulties, until 13 years of age. The findings in his skeletal muscle biopsy, including nemaline rods supported his NEM8 diagnosis. It is noteworthy that intranuclear rods were also observed on biopsy, given this feature has not been previously reported in NEM8 cases. The patient's skeletal muscle revealed significantly low levels of KLHL40 compared to controls, thus supporting *KLHL40* as the underlying genetic cause of his disease.

The exon 2-6 deletion is expected to generate a non-functional transcript and result in complete loss of KLHL40 from the affected allele. Given the biallelic nature of KLHL40-related disease[4], we believe the c.*152G>T variant in *KLHL40* is the second underlying cause of nemaline myopathy in the patient based on the following evidence. This variant was detected *in trans* with the pathogenic deletion and was predicted by SpliceAI and Introme to create a cryptic donor splice site in the 3' UTR. This rare mechanism has been previously suggested by the literature to reduce transcript stability and abundance[25,26]. As predicted, RNA-seq and western blotting analyses of patient skeletal muscle as well as *in vitro* expression assays indicated a significant reduction in *KLHL40* mRNA and protein, thus supporting the pathogenic effects of the c.*152G>T variant. Interestingly, this variant appeared to induce multiple splicing events in the 3' UTR. The 78 bp cryptic intron was apparent in both RNA-seq and *in vitro* analyses (Figure 3.2F, Figure 3.3D) and appears to form a second deletion variant; c.*151_*228del that replaces the c.*152G>T variant (Figure 3.2F). In contrast, the 255 bp cryptic intron, which was predicted by SpliceAI and Introme, was captured in our *in vitro* studies (Figure 3.3D) but not detected by RNA-seq. We predict that the 255 bp deletion may prevent maturation of the resulting transcript, which would explain why it is undetected by RNA-seq. RT-PCR analyses of patient skeletal

muscle RNA may help confirm whether the transcript with the 255 bp deletion is present in the patient.

We initially postulated that the deletion(s) in the 3' UTR would reduce *KLHL40* stability by removing recognition sites for regulatory elements such as polyadenylation proteins[27,28], or by introducing illegitimate microRNA sites[29,30]. However, there were no remarkable annotations for this region in 3' UTR databases. Moreover, our *in vitro* expression analyses suggested that the 78 bp deletion itself does not significantly affect expression of *KLHL40* (Figure 3.3A-C). While mRNA and protein abundance following expression of *KLHL40*$^{c.*152G>T}$ were significantly lower than wildtype *KLHL40* ($P < 0.01$, Figure 3.3A), mRNA and protein abundance following expression of *KLHL40*$^{c.*151\_*228del}$ appeared comparable to the wildtype despite it containing the 78 bp deletion. This suggested that the reduction in transcript levels was directly influenced by splicing of the c.*152G>T variant. We propose the pathomechanism of c.*152G>T may be attributed to a rare, underrecognised mechanism involving nonsense mediated decay stimulated by 3' UTR intron splicing[25,31].

Nonsense mediated decay is an RNA surveillance mechanism implicated in various genetic disorders[32]. In most reports, nonsense mediated decay is triggered by variants that result in protein pretermination, including frameshift, nonsense, and splice variants. A rarer mechanism of nonsense mediated decay involves splicing in the 3' UTR[25]. It has been suggested that intronic insertions into most human 3' UTRs at a distance >50–55 bp downstream of the termination codon will stimulate nonsense mediated decay[26]. This mechanism was functionally demonstrated by introduction of a spliceable intron 62 bp into the 3' UTR of wild-type β-Globin mRNA which reduced protein abundance to 24% of the wildtype[33]. Notably, while 3' UTR-splice activated nonsense mediated decay has been demonstrated by *in vitro* studies[33-35], to our knowledge, this mechanism has not been reported in Mendelian disease to date.

The mechanism of 3' UTR-splice activated nonsense mediated decay may explain the patient's relatively milder disease presentation. Given the c.*152G>T cryptic donor splice site is weakly spliced, a proportion of transcripts are expected to evade aberrant splicing and proceed to translation of functional KLHL40. Indeed, some apparently normal RNA-seq reads were present as seen by IGV alignment (Figure 3.2F) and

MINTIE alignment. Additionally, our *in vitro* experiments detected some expression of *KLHL40*$^{c.*152G>T}$ mRNA and protein (Figure 3.3).

Mechanistically, 3' UTR splice-activated nonsense mediated decay involves deposition of RNA binding complexes, such as the human UPF protein complex and the exon junction complex (EJC) at the 3' UTR[31,34]. Following translation termination, interactions between the 3' UTR-bound protein complexes and the terminating ribosome signals activation of nonsense mediated decay[25,32,36]. Based on our data, we postulate that the cryptic introns introduced by the c.*152G>T variant may influence recruitment of such complexes to the 3' UTR and thus trigger nonsense mediated decay (Figure 3.4)[31]. This is supported by the findings from our *in vitro* studies, whereby the *KLHL40*$^{c.*151\_*228del}$ construct containing the 78 bp deletion without a splice site did not show a significant reduction in transcript expression. Further experiments including a nonsense mediated decay inhibitor in our *in vitro* KLHL40 expression system may confirm whether this mechanism indeed drives the reduction in *KLHL40* transcript and protein levels[37].

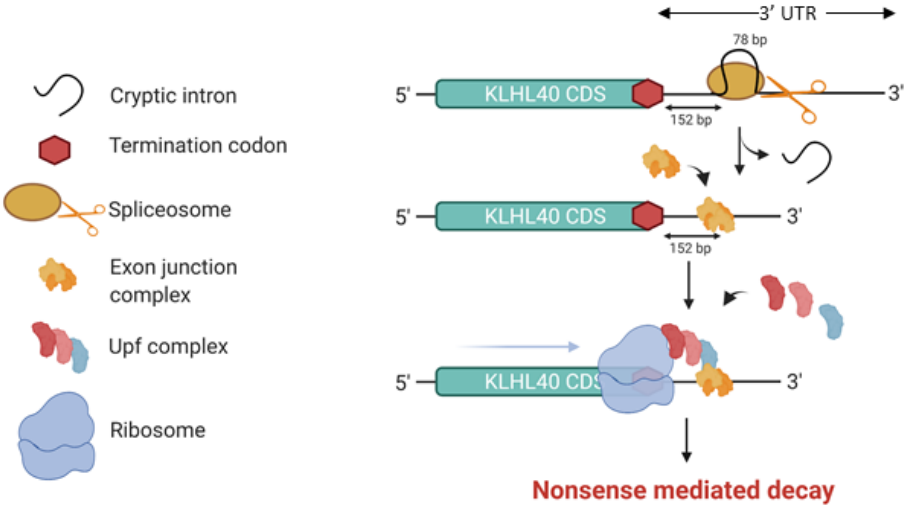


Figure 3.4 Schematic representation of the predicted 3' UTR splice-activated nonsense mediated decay mechanism in *KLHL40*. Figure designed with BioRender.com.

Despite the crucial roles of 3' UTRs in mRNA metabolism and surveillance[25,34,38], 3' UTR variants identified during disease gene screening are often overlooked or reluctantly assessed[29]. This is partly due to the limited understanding of 3' UTR pathomechanisms, as well as limited functional assays available to assess the clinical

significance of such variants[29]. There are a few reports on pathogenic 3' UTR variants mediating aberrant splicing[29,30]. However, none explicitly describe splice-activated nonsense mediated decay as the driving mechanism. One example involves a 3' UTR splice variant in the lipoma HMGIC fusion partner-like 5 gene (*LHFPL5*) associated with hearing impairment (OMIM# 610265). The variant: c.*16+1G>A is predicted to create a cryptic splice site leading to a longer 3′ UTR containing cryptic regulatory motifs that destabilise the *LHFPL5* transcript[29]. Another example involves a c.607–1G>A variant in the major intrinsic protein (*MIP*) gene, implicated in autosomal dominant congenital cataracts (OMIM# 615274)[39]. This variant activates a cryptic acceptor splice site in the 3' UTR, which impacts sequences that encode the critical distal and terminal MIP domains[39]. Additionally, 3' UTR (CTG)n expansions in the DM1 Protein Kinase *(DMPK)* gene are implicated in myotonic dystrophy (OMIM# 160900). There are several speculations as to how the expansions reduce levels of *DMPK* mRNA and protein, one of which involves aberrant splicing[28,40]. It is tempting to suggest that the underlying mechanism of the latter example involves 3' UTR splice-activated nonsense-mediated decay. Overall, such reports indicate that aberrant 3' UTR-splicing may underlie causes of rare diseases that remain undiagnosed.

Our data supports that the biallelic c.*152G>T variant and 10.9 kb deletion in *KLHL40* are the underlying causes of *KLHL40* deficiency in the patient. This finding expands the genotypic and phenotypic spectrum of NEM8. Of note, the low frequency of the c.*152G>T variant in the healthy population suggests that this variant could become a recurrent cause of NEM8. Therefore, we encourage consideration of this variant during nemaline myopathy gene screening.

3.7 Conclusions

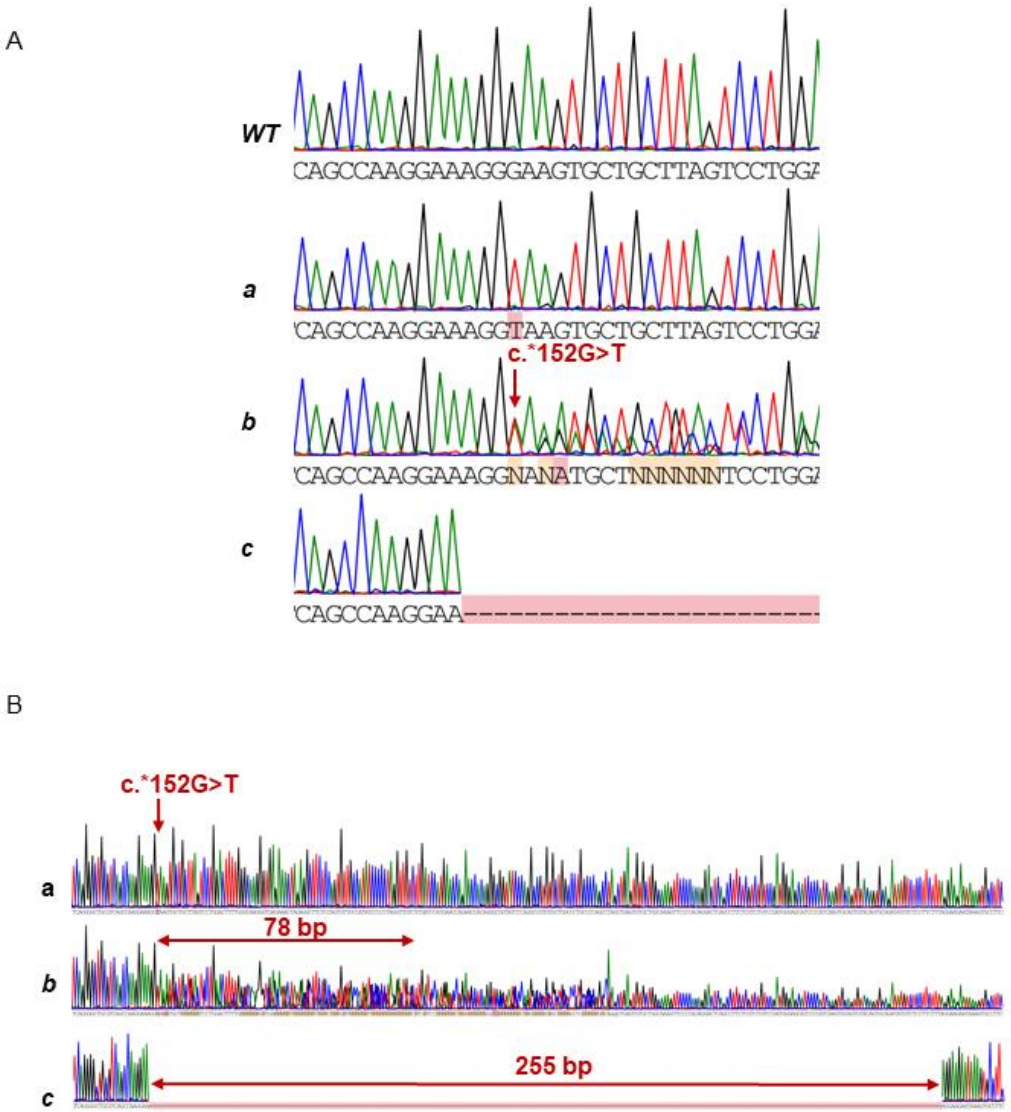
This study indicates a likely pathogenic novel splice variant in the 3' UTR of *KLHL40* (c.*152G>T) in a patient with a milder presentation of NEM8. Using next generation sequencing, RNA-seq, and *in vitro* studies, we provide the first evidence of a pathogenic variant destabilising the *KLHL40* mRNA by 3' UTR splice-activated nonsense mediated decay. This may be an underrecognised mechanism of Mendelian disease. Our study highlights the utility of RNA-seq in 1) detecting transcript deficiency from patient muscle biopsy, and 2) identifying cryptic splice variants in

non-coding genomic regions that may evade detection by targeted gene panel and whole exome sequencing[41]. Inclusion of RNA-seq into the diagnostic toolkit for rare disorders can thus enable identification of pathogenic variants that may be substantially contributing to the cases remaining undiagnosed following genetic screening[41]. Our findings shed light into an often-overlooked pathway and encourage 3' UTR screening and complimentary RNA-seq in patients with unsolved disease[30]. As 3' UTRs play a crucial role in mRNA metabolism and surveillance, we urge investigations of 3' UTR variants in disease gene screening and suggest they underlie a proportion of unsolved genetic disorders.

3.8 Acknowledgements

We would like to express our gratitude to the patient and family for their involvement in this study. We thank Associate Professor David Groth and Dr. Danielle Dye for their inputs into this study. We gratefully acknowledge the funding bodies for their valuable support. LD is supported by an Australian Government Research Training Program (RTP) Scholarship. GR (Investigator Grant, APP2007769) and NGL (Fellowship APP1117510) are supported by the Australian NHMRC. This work is funded by NHMRC Ideas Grant (APP2002640).

3.9 Supplementary materials



Supplementary Fig 3.1 Sanger sequencing of RT-PCR products from *in vitro* expression of *KLHL40*^{WT} and *KLHL40*$^{c.*152G>T}$. (A) Close up view and (B) expanded view of sequence chromatograms. Corresponding gel products (a-c) are shown in Figure 3.3D

3.10 References

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged

1. Nowak KJ, Davis MR, Wallgren-Pettersson C, Lamont PJ, Laing NG. Clinical utility gene card for: Nemaline myopathy. *Eur J Hum Genet.* 2012;20(6):713-713.

2. Gonorazky HD, Bönnemann CG, Dowling JJ. The genetics of congenital myopathies. *Handb Clin Neurol.* 2018;148:549-564.

3. Benarroch L, Bonne G, Rivier F, Hamroun D. The 2021 version of the gene table of neuromuscular disorders (nuclear genome). *Neuromuscul Disord.* 2020;30(12):1008-1048.

4. Ravenscroft G, Miyatake S, Lehtokari V-L, et al. Mutations in KLHL40 Are a Frequent Cause of Severe Autosomal-Recessive Nemaline Myopathy. *Am J Hum Genet.* 2013;93(1):6-18.

5. Chen TH, Tian X, Kuo PL, Pan HP, Wong LC, Jong YJ. Identification of KLHL40 mutations by targeted next-generation sequencing facilitated a prenatal diagnosis in a family with three consecutive affected fetuses with fetal akinesia deformation sequence. *Prenat Diagn.* 2016;36(12):1135-1138.

6. Seferian AM, Malfatti E, Bosson C, et al. Mild clinical presentation in KLHL40-related nemaline myopathy (NEM 8). *Neuromuscul Disord.* 2016;26(10):712-716.

7. Natera-de Benito D, Nascimento A, Abicht A, et al. KLHL40-related nemaline myopathy with a sustained, positive response to treatment with acetylcholinesterase inhibitors. *J Neurol.* 2016;263(3):517-523.

8. Garg A, O'Rourke J, Long C, et al. KLHL40 deficiency destabilizes thin filament proteins and promotes nemaline myopathy. *J Clin Invest.* 2014;124(8):3529-3539.

9. Gong W, Gohla RM, Bowlin KM, Koyano-Nakagawa N, Garry DJ, Shi X. Kelch Repeat and BTB Domain Containing Protein 5 (Kbtbd5) Regulates Skeletal Muscle Myogenesis through the E2F1-DP1 Complex. *J Biol Chem.* 2015;290(24):15350-15361.

10. Ravenscroft G, Clayton JS, Faiz F, et al. Neurogenetic fetal akinesia and arthrogryposis: genetics, expanding genotype-phenotypes and functional genomics. *J Med Genet.* 2021;58(9):609-618.

11. Beecroft SJ, Yau KS, Allcock RJN, et al. Targeted gene panel use in 2249 neuromuscular patients: the Australasian referral center experience. *Ann Clin Transl Neurol.* 2020;7(3):353-362.

12. Pais L, Snow H, Weisburd B, et al. seqr: a web-based analysis and collaboration tool for rare disease genomics. *medRxiv.* 2021:2021.2010.2027.21265326.

13. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-423.

14. Dashti MJS, Gamieldien J. A practical guide to filtering and prioritizing genetic variants. *BioTechniques.* 2017;62(1):18-30.

15. Sullivan P, Mayoh C, Wong-Erasmus M, et al. NEW GENES AND DISEASES / NGS & RELATED TECHNIQUES: P.334 Introme identifies non-canonical splice-altering variants in neuromuscular patients resulting in multiple new genetic diagnoses. *Neuromuscul Disord.* 2020;30:S144.

16. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell.* 2019;176(3):535-548.e524.

17. Bryen SJ, Ewans LJ, Pinner J, et al. Recurrent TTN metatranscript-only c.39974-11T>G splice variant associated with autosomal recessive arthrogryposis multiplex congenita and myopathy. *Hum Mutat.* 2020;41(2):403-411.

18. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178-192.

19. Cmero M, Schmidt B, Majewski IJ, Ekert PG, Oshlack A, Davidson NM. MINTIE: identifying novel structural and splice variants in transcriptomes using RNA-seq data. *bioRxiv.* 2020:2020.2006.2003.131532.

20. Cmero M, Schmidt B, Majewski IJ, Ekert PG, Oshlack A, Davidson NM. MINTIE: identifying novel structural and splice variants in transcriptomes using RNA-seq data. *Genome Biol.* 2021;22(1):296.

21. McNamara EL, Taylor RL, Clayton JS, et al. Systemic AAV8-mediated delivery of a functional copy of muscle glycogen phosphorylase (Pygm) ameliorates disease in a murine model of McArdle disease. *Hum Mol Genet.* 2020;29(1):20-30.

22. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002;3(7):research0034.0031.

23. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9(7):671-675.

24. Todd EJ, Yau KS, Ong R, et al. Next generation sequencing in a large cohort of patients presenting with neuromuscular disease before or at birth. *Orphanet J Rare Dis.* 2015;10:148-148.

25. Bicknell AA, Cenik C, Chua HN, Roth FP, Moore MJ. Introns in UTRs: Why we should stop ignoring them. *BioEssays.* 2012;34(12):1025-1034.

26. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci.* 1998;23(6):198-199.

27. Nourse J, Spada S, Danckwardt S. Emerging Roles of RNA 3'-end Cleavage and Polyadenylation in Pathogenesis, Diagnosis and Therapy of Human Disorders. *Biomolecules.* 2020;10(6).

28. Frisch R, Singleton KR, Moses PA, et al. Effect of Triplet Repeat Expansion on Chromatin Structure and Expression of DMPK and Neighboring Genes, SIX5 and DMWD, in Myotonic Dystrophy. *Mol Genet Metab.* 2001;74(1):281-291.

29. Liaqat K, Chiu I, Lee K, et al. Novel missense and 3'-UTR splice site variants in LHFPL5 cause autosomal recessive nonsyndromic hearing impairment. *J Hum Genet.* 2018;63(11):1099-1107.

30. Dusl M, Senderek J, Müller JS, et al. A 3'-UTR mutation creates a microRNA target site in the GFPT1 gene of patients with congenital myasthenic syndrome. *Hum Mol Genet.* 2015;24(12):3418-3426.

31. Giorgi C, Yeo GW, Stone ME, et al. The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. *Cell.* 2007;130(1):179-191.

32. Kurosaki T, Maquat LE. Nonsense-mediated mRNA decay in humans at a glance. *J Cell Sci.* 2016;129(3):461-467.

33. Thermann R, Neu-Yilik G, Deters A, et al. Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.* 1998;17(12):3484-3494.

34. Lykke-Andersen J, Shu MD, Steitz JA. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell.* 2000;103(7):1121-1131.

35. Carter MS, Li S, Wilkinson MF. A splicing-dependent regulatory mechanism that detects translation signals. *EMBO J.* 1996;15(21):5965-5975.

36. Le Hir H, Gatfield D, Izaurralde E, Moore MJ. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* 2001;20(17):4987-4997.

37. Martin L, Grigoryan A, Wang D, et al. Identification and characterization of small molecules that inhibit nonsense-mediated RNA decay and suppress nonsense p53 mutations. *Cancer Res.* 2014;74(11):3104-3113.

38. Preussner M, Gao Q, Morrison E, et al. Splicing-accessible coding 3′UTRs control protein stability and interaction networks. *Genome Biol.* 2020;21(1):186.

39. Jin C, Jiang J, Wang W, Yao K. Identification of a MIP mutation that activates a cryptic acceptor splice site in the 3' untranslated region. *Mol Vis.* 2010;16:2253-2258.

40. Carango P, Noble JE, Marks HG, Funanage VL. Absence of myotonic dystrophy protein kinase (DMPK) mRNA as a result of a triplet repeat expansion in myotonic dystrophy. *Genomics.* 1993;18(2):340-348.

41. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med.* 2017;9(386):1-11.

Chapter 4

Biallelic variants in *HMGCS1* are a novel cause of rare rigid spine syndrome

4.1 Preface

A manuscript arising from this chapter is currently in preparation for submission to *Brain*, as an article titled *Biallelic variants in HMGCS1 are a novel cause of rare rigid spine syndrome*. As such, this chapter has been formatted for *Brain.*

An authorship attribution statement is provided in Appendix I.

4.2 Abstract

Rigid spine syndrome is a severe and rare inherited myopathy that resembles some congenital myopathies and muscular dystrophies. Key features include scoliosis, contractures of the neck and spine, hypotonia and respiratory failure. Onset typically occurs during the first decade of life and can be static or slowly progressive. Common genetic causes of rigid spine syndrome include biallelic variants in *SELONON* and *RYR1*. However, the genetic cause in some patients remains to be identified.

In this study, we performed whole exome sequencing or whole genome sequencing of DNA from patients with severe rigid spine syndrome of unknown genetic aetiology.

We identified biallelic variants in the 3-hydroxy-3-methylglutaryl-coenzyme A synthase gene (*HMGCS1*) in four patients with rigid spine syndrome from three unrelated families. These include four missense variants and a frameshift variant. Patients presented with a consistent phenotype involving spinal rigidity and scoliosis primarily affecting the cervical region, and respiratory insufficiency. Clinical course worsened with infection and febrile illnesses; two patients died from respiratory failure following infection.

Western blotting for HMGCS1 of skeletal muscle biopsy of a patient homozygous for the p.(Ser447Pro) substitution suggests that the variant does not reduce HMGCS1 abundance. Using recombinant human wildtype and mutant ($HMGCS1^{S447P}$) protein, we performed size exclusion chromatography, circular dichroism, and enzymatic assays respectively to determine whether the p.(Ser447Pro) variant disrupts HMGCS1 dimerisation, secondary structure, thermal stability, and enzyme activity. We show that $HMGCS1^{S447P}$ maintains its 3D structure and dimerized state and retains ~92% of the wildtype's thermal stability. Enzymatic analyses indicated moderate dysregulation of enzyme activity, with $HMGCS1^{S447P}$ retaining ~56% of the wildtype activity when saturated with acetoacetyl-CoA. The clinical significance of this dysregulation remains uncertain.

HMGCS1 encodes a key enzyme in the mevalonate pathway and has not previously been associated with human disease. It presents as a suitable disease gene candidate given that rigid spine manifestations have been reported in disorders of the mevalonate

pathway. Altogether, our analyses suggest that biallelic variants in *HMGCS1* are pathogenic and act through a hypomorphic mechanism yet to be elucidated. Here, we report a novel cause of rigid spine syndrome which we have termed *HMGCS1*-myopathy and suggest *HMGCS1* should be screened in patients presenting with an unresolved rigid spine phenotype.

Keywords: rigid spine syndrome; HMGCS1; mevalonate pathway; enzymopathy; neuromuscular disease.

Abbreviations: AcAc-CoA = acetoacetyl coenzyme A; Ac-CoA = acetyl coenzyme A; AUC = analytical ultracentrifugation; CD = circular dichroism; HMG-CoA = 3-hydroxy-3-methylglutaryl coenzyme A; K_m = Michaelis constant; TPM = transcripts per million; V_{max} = maximum velocity; WES = whole exome sequencing; WGS = whole genome sequencing

4.3 Introduction

Inherited myopathies are a clinically and genetically heterogeneous group of debilitating disorders characterised by skeletal muscle dysfunction and weakness.[1] Traditional subgroups include congenital myopathies, muscular dystrophies, mitochondrial myopathies, and metabolic myopathies. However, the clinical overlap between these entities can make it difficult to attain an accurate diagnosis.[1,2]

Rigid spine syndrome is a severe disorder described across the various myopathy subgroups.[2-4] Key features include contractures of the limb and spinal joints, limitation of flexion of the neck and trunk, weakness of the cervical and dorsolumbar spine muscles and progressive scoliosis.[2,3,5-9] Muscle pathology is variable and can include multiminicores, central cores, core-rods, rimmed vacuoles, and fibre-type predominance.[3,4,10] The syndrome typically manifests during the first decade of life and can be slowly progressive or non-progressive and may be accompanied by elevated serum creatine kinase levels.[11] Patients are susceptible to recurrent infections and respiratory insufficiency which may be fatal.[3,5-9] Rigid spine syndrome can manifest as a primary disorder, for example SELONON-related myopathy[12] or present as a secondary consequence of other primary congenital myopathies and muscular dystrophies.[13] Diseases associated with rigid spine syndrome include rigid spine muscular dystrophy (OMIM# 602771), congenital myopathy with fibre-type disproportion (OMIM# 255310), central core disease (OMIM# 117000), and minicore myopathy with external ophthalmoplegia (OMIM# 255320). For the purposes of this manuscript, the term rigid spine syndrome will be used as an umbrella description for the rigid spine-associated disorders described above.

The genetic mechanisms and pathways that underlie rigid spine syndrome are diverse.[2,6,9,14,15] Biallelic variants in the gene (*SELONON*) encoding selenoprotein N are the most frequently reported cause of rigid spine manifestations.[7,9,12] Other associated genes include *RYR1*[16], *TOR1AIP1*[6] *TPM3*[17], *ACTA1*[8] and more recently, *GGPS1*.[10] However, the genetic cause of some patients remains unknown despite screening for known disease-associated genes.[2,13] This suggests that additional genes are yet to be associated with rigid spine syndrome.

In this study, we investigated genetic causes of rigid spine syndrome in patients

remaining without a genetic diagnosis using next generation sequencing and functional genomic approaches. In four patients from three unrelated families, we identified biallelic variants in *HMGCS1*, a gene of the mevalonate pathway encoding 3-hydroxy-3-methylglutaryl-coenzyme A synthase. To our knowledge, *HMGCS1* has not been previously associated with disease, though rigid spine manifestations have been associated with disorders of the mevalonate pathway.[10,18] Here, we report biallelic variants in *HMGCS1* as a novel cause of rigid spine syndrome which we have termed *HMGCS1*-myopathy. This report adds a new puzzle piece to the roles of the mevalonate pathway in skeletal muscle physiology and encourages screening of *HMGCS1* in patients with unsolved rigid spine myopathy.

4.4 Materials and methods

4.4.1 Patients and ethics

The study included four patients from three unrelated families from Spain, Japan, and the USA. Patients P1 and P2 (Family 1) were clinically and morphologically investigated at the Biomedicine institute in Sevilla, Spain. Written informed consent was provided by the families and approved by the University of Western Australia Human Research Ethics Committee (2019/RA/4/20/1008) and the Curtin University Human Research Ethics Office (HRE2019-0566).

Patient P3 (Family 2) was identified by personal communication with collaborators from the National Institute of Neuroscience and National Center of Neurology and Psychiatry in Tokyo, Japan. Patient P4 (Family 3) was identified via the Broad Institute's *matchbox* by collaborators from the Center for Mendelian Genomics (Broad Institute of Harvard and MIT, Massachusetts, USA). All patients were clinically diagnosed with a rigid spine syndrome of unknown genetic aetiology.

4.4.2 Genetic studies

4.4.2.1 Whole exome sequencing and variant validation

For Family 1, Illumina whole exome sequencing was performed for the affected siblings and unaffected parents at the Center for Mendelian Genomics (Broad Institute of Harvard and MIT, Boston, USA) as previously described.[19] Data were mapped, variant called and analysed using seqr (https://seqr.broadinstitute.org/).[20] Variant interpretation and prioritisation were performed following standard bioinformatic

approaches.[21-25] This included confining the data to rare variants (allele frequency < 0.01; gnomAD) that segregated with disease in the family. Variants were confirmed by PCR and bi-directional Sanger sequencing (AGRF Perth, Australia).

4.4.3 Modelling of the HMGCS1 substitutions

The HMGCS1 substitutions were mapped to the HMGCS1 gene (ENSG00000112972) and protein structure (PDB: 2P8U).[26] The crystal structure for HMGCS1 was accessed from the Protein Data Bank (PDB; www.rcsb.org)[27] and analysed using the PyMOL Molecular Graphics System, Version 2.4.1 (Schrödinger). Annotations including residues of the HMGCS1 active sites, the CoA binding sites, and the salt bridge were obtained from the UniProt database (Q01581) and the ICn3D (I-see-in-3D) molecular structure viewer.[28] For conservation analysis, HMGCS1 amino acid sequences were obtained from UniProt and aligned by Clustal omega alignment (https://www.ebi.ac.uk/Tools/msa/clustalo/).

4.4.4 *HMGCS1* expression, HMGCS1 protein abundance and localisation

4.4.4.1 Gene expression databases

RNA-seq data for *HMGCS1* (ENST00000325110) was obtained from the Genotype-Tissue Expression (GTEx) portal (https://gtexportal.org/) and the NCBI (BioProject PRJEB4337). Additionally, CAGE (Cap Analysis Gene Expression) data was obtained from the Functional Annotation of the Mammalian genome consortium (FANTOM5; https://fantom.gsc.riken.jp/5/).[29]

4.4.4.2 Quantitative PCR

RNA extraction, cDNA synthesis and quantitative real-time PCR (qPCR) were performed following previously described methods.[30,31] Samples for *HMGCS1* expression analysis included: healthy adult skeletal muscle samples from *in vitro* contracture testing (IVCT), cultured skeletal muscle myoblasts and myotubes from a healthy male donor (18M) and two healthy female donors (32F and 42F, Cook Myosite), cauda equina and cortex (Victorian Brain Bank), human embryonic kidney cell line (HEK293FT), fibroblast cell lines, and fetal skeletal muscle tissue samples (Perinatal Pathology, KEMH). In brief, 10 µL reactions were performed using 2X Rotor-Gene SYBR Green PCR master mix (Qiagen; 204076) and contained 1 µL diluted cDNA and 0.8 µM of forward and reverse primers (Supplementary Table 4.1). A standard curve from positive control cDNA was generated to validate primer

efficiency. Cycling was performed on a Rotor-Gene Q real-time PCR cycler (Qiagen) following the described conditions: 95°C (5 min), 45 cycles of 95°C (10 s) and 60°C (15 s), and subsequent melt curve analysis (60- 95°C; +1°C per cycle). Data were normalised using the delta Ct method[32] by comparing to the geometric mean of two reference genes (*TBP* and *EEF2*).[31]

4.4.4.3 Western blotting

Protein extraction and western blotting was performed as previously described.[30] Frozen cell pellets and mechanically shredded tissue were suspended in lysis buffer containing 2% w/v SDS, 125 mM Tris (pH 6.8), and 1x PIC (Halt™ Protease Inhibitor Cocktail; Thermo Scientific) and homogenised by sonication. Protein concentrations were determined by a BCA protein assay (Pierce; #23227). Samples were prepared in loading buffer (10% w/v SDS, 312.5 mM Tris (pH 6.8), 50% v/v glycerol, bromophenol blue saturated solution, 50 mM DTT and 1x PIC). Extracts (12 µg) and recombinant HMGCS1 (5 ng) were separated on 4-12% w/v Bis-Tris NuPAGE Novex polyacrylamide gels and transferred to polyvinylidene difluoride transfer membranes (Thermo Fisher Scientific). Membranes were blocked with PBS + 0.1% v/v Tween-20 and 5% w/v skim milk for 1 h and probed with an anti-HMGCS1 rabbit polyclonal antibody (MyBioSource; #MBS2026097; 1:800 incubated overnight at 4°C) and, subsequently, a secondary goat anti-rabbit horseradish peroxidase antibody (Sigma; #A0545) for 1 h. Membranes were imaged by chemiluminescent detection (Pierce ECL Plus kit; #32132) using the Invitrogen iBright FL1000 imaging system (Thermo Fisher Scientific). An anti-GAPDH antibody (Sigma; #G8795; 1:10,000) was used as a loading control. Relative HMGCS1 abundance was calculated using ImageJ.[33]

4.4.4.4 Quantification of HMGCS1 in primary human myoblasts and myotubes by Mass spectrometry

Quantitative mass spectrometry was performed as previously described[34] to quantify HMGCS1 peptides in cultured skeletal muscle myoblasts and myotubes (day 2 and day 8 of differentiation) from 18M, 32F and 42F donors (Cook Myosite).[35]

4.4.4.5 Analysis of HMGCS1 localisation in skeletal muscle

Immunofluorescence was performed as previously described.[31] Skeletal muscle cross sections of 10 µM thickness were fixed in 2% w/v paraformaldehyde for 10 min and blocked in 10% v/v fetal calf serum and 10% v/v bovine serum albumin (10 mg/ml) in

phosphate buffered saline for 1 h at room temperature. Sections were incubated with a HMGCS1 rabbit polyclonal antibody (diluted 1:10, MyBioSource [MBS2026097]), coimmunostained with a mouse monoclonal myosin heavy chain (fast) antibody (Novocastra [NCL-MHCF], diluted 1:20) overnight at 4°C. Recombinant wild type HMGCS1 (produced as per methods below) was used as a blocking peptide control to assess background signal. Recombinant HMGCS1 was denatured at 95°C for 20 min with occasional vortexing. Control skeletal muscle sections were incubated with 0.096 µg/uL of recombinant peptides (corresponding to 3x the concentration of HMGCS1 rabbit polyclonal antibody) overnight at 4°C. All sections were subsequently incubated with a secondary anti-rabbit IgG (diluted 1:500; AlexaFluor 555, Life Technologies) and a donkey anti-mouse antibody (diluted 1:500; AlexaFluor 488, Life Technologies) for 1 h. Sections were imaged with a fluorescence microscope (model IX-71, Olympus) and processed using the Olympus cellSens software.

4.4.5 Recombinant protein production and functional analyses

4.4.5.1 Recombinant protein expression

Bacterial pET-30a(+) expression constructs were synthesised by Genscript (https://www.genscript.com/). These encoded the coding DNA sequences for human HMGCS1 (ENST00000325110) wildtype ($HMGCS1^{WT}$), HMGCS1 p,Ser447Pro ($HMGCS1^{S447P}$), or ACAT2 (ENST00000367048). The constructs were codon-optimised for *E. coli* expression and encoded a six-histidine tag at the N-terminus. Wild-type and mutant constructs were confirmed by Sanger sequencing (AGRF, Perth, Australia). Constructs were transformed into BL21-DE3 *E. coli* following a standard heat-shock method.[36] Cultures were grown at 37°C in Luria-Bertani (LB) broth containing 35 µg/mL kanamycin. Expression was induced with 0.1 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) with overnight growth at 16°C. Protein extraction and purification was performed as previously described.[37] Cultures were lysed by sonication in 50 mM Tris (pH 8.0), 250 mM sodium chloride, 10% v/v glycerol and 0.1% v/v Triton X-100. The soluble fraction containing recombinant protein was incubated overnight at 4°C with resins (profinity IMAC, #156-0123) charged with 0.2 M Nickel (II) sulfate hexahydrate (Sigma). Proteins were eluted by washing in buffer containing 50 mM Tris (pH 8.0), 150 mM NaCl, and 5% v/v glycerol and increasing concentrations of imidazole (0 to 300 mM). Eluants were concentrated to 1 M in buffer containing 50 mM Tris (pH 8.14), 150 mM NaCl and 5% v/v glycerol by centrifuging

at 3,000 g, 4°C using 30 kDa Amicon ultra centrifugal filter units (Sigma).

4.4.5.2 Structural and functional analyses

4.4.5.2.1 Size exclusion chromatography

Size exclusion chromatography was performed as previously described.[37] Protein concentrates were loaded onto a XK16/60 Superdex 200 column or 10/300 GL Superdex 200 column (GE Healthcare Life Sciences) for purification and dimerisation analysis. Columns were equilibrated in 150 mM NaCl, 50 mM Tris (pH 8.14), 5% v/v glycerol. Fractions were collected and tested for protein of the right size by SDS-PAGE. Bands were excised for validation of protein identity and purity by mass spectrophotometry (Proteomics International, Perth). The concentration of recombinant protein was determined by NanoDrop 2000 V1.5 (Thermo Scientific) at 280 nm. The exclusion coefficient and molecular weight of protein were calculated using the Expasy ProtParam webtool[38] (https://web.expasy.org/protparam/). The extinction coefficient and molecular weight (excluding His-tags) respectively for HMGCS1 were 65.39 $mM^{-1}cm^{-1}$ and 57.29 kDa, and for ACAT2 were 28.46 $mM^{-1}cm^{-1}$ and 43.03 kDa, respectively.

4.4.5.2.2 Circular dichroism

Circular dichroism (CD) analyses were performed as previously described.[37] $HMGCS1^{WT}$ and $HMGCS1^{S447P}$ were buffer-exchanged to 10 mM sodium phosphate (pH 8.0) using 30 kDa Amicon filters. In technical replicates, proteins were diluted to 0.1 mg/mL for CD wavelength spectra analysis and 0.5 mg/ml for melt curve analysis. Proteins (200 µL) were loaded into a 0.1 cm quartz cuvette and analysed by a Jasco J-810 CD spectrometer with a 1 nm bandwidth. To correct for background signals, the spectrophotometer was blanked in 10 mM phosphate buffer. For CD spectra analysis, samples were measured at 25°C (195-260 nm) at a rate of 1 nm/s. Spectral data were collected using the JASCO Spectra manager software (V1). For melt curve analyses, upon equilibration of the instrument to 25°C, samples were measured from 25-95°C (220 nm) at a rate of +1°C/min. Data were collected using the JASCO Spectra manager (V2). $HMGCS1^{WT}$ and $HMGCS1^{S447P}$ melt curves were plotted to a sigmoidal four parameter logistic regression fit to determine the IC50 value (point of inflection).[37] The IC50 was taken as a measure of protein melting temperature (Tm; temperature at which 50% of protein is denatured).[37]

4.4.5.2.3 Enzyme assays

HMGCS1 catalyses the condensation reaction between acetyl coenzyme A (Ac-CoA) and acetoacetyl coenzyme A (AcAc-CoA) to form 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA).[39] To investigate whether the p.Ser447Pro substitution alters HMGCS1 activity, we adapted an assay described by Lowe and Tubbs (1985)[39] using recombinant human wildtype and mutant HMGCS1. The standard assay system contained 100 mM Tris/HCI, pH 8, 40 mM $MgCl_2$ in a total volume of 200 µl. Variable acetyl-CoA and acetoacetyl-CoA concentrations were used, ranging between 0-200 µM. Reactions were performed in a UV-star 96 well microplate (Greiner) at 25°C and initiated with 1 µg of enzyme. The amount of AcAc-CoA consumed per reaction was determined in relation to a standard curve. Data were fitted to a Michaelis-Menten curve on GraphPad Prism 9.0.1 (La Jolla, USA).

To determine the Ac-CoA-associated catalytic parameters of $HMGCS1^{WT}$ and $HMGCS1^{S447P}$, we performed assays at various concentrations of Ac-CoA (0-200 µM) at a fixed concentration of AcAc-CoA (25 µM). To investigate the effect of AcAc-CoA on catalytic activity, $HMGCS1^{WT}$ and $HMGCS1^{S447P}$ activity were measured at increasing concentrations of AcAc-CoA (0-300 µM) and at a fixed concentration of Ac-CoA (25 µM). The resultant data were fit to Michaelis-Menten plots to determine catalytic parameters including maximum velocity (V_{max}) and Michaelis constant (K_m).

4.4.5.2.4 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) was performed to investigate whether $HMGCS1^{WT}$ or $HMGCS1^{S447P}$ complexed with ACAT2 *in vitro*. Protein production, purification and quantification were performed following methods described above.[37] AUC was performed at La Trobe University (Victoria, Australia) following methods previously described.[40-42] Further details are provided in the Supplementary material.

4.5 Data availability

The data and constructs generated by this study are available upon request.

4.6 Results

4.6.1 Clinical descriptions of the rigid spine syndrome cohort

Four patients from three families were clinically examined at centres in Spain, Japan, and the USA. All individuals were reported with features resembling rigid spine syndrome. Features included spinal rigidity affecting the cervical (4/4) and dorsolumbar (3/4) regions, scoliosis (4/4) and restrictive pulmonary function (4/4). In addition, some patients were reported with proximal limb weakness (2/4), scapular winging (2/4) and bulbar weakness (1/4). Creatine kinase levels were elevated in all cases and ranged between 190-7890 IU/L (normal ~70-170 IU/L[43]). The age of onset for all patients was two years of age. Muscle pathology findings included multiminicore-like structures and occasional central nuclei (patient P1, Figure 4.1) and rimmed vacuoles (patient P3). The disease severity for patients P2 (Family 1) and P3 (Family 2) worsened episodically during intercurrent disease and was associated with increased muscular weakness and/or respiratory difficulty. Patients P2 and P3 succumbed at 15 years of age and 36 years of age, respectively. Patients P1 (Family 1) and P4 (Family 3) were alive when last seen at 27 years of age and 45 years of age, respectively. A detailed clinical history for Family 1 can be found in the Supplementary material.

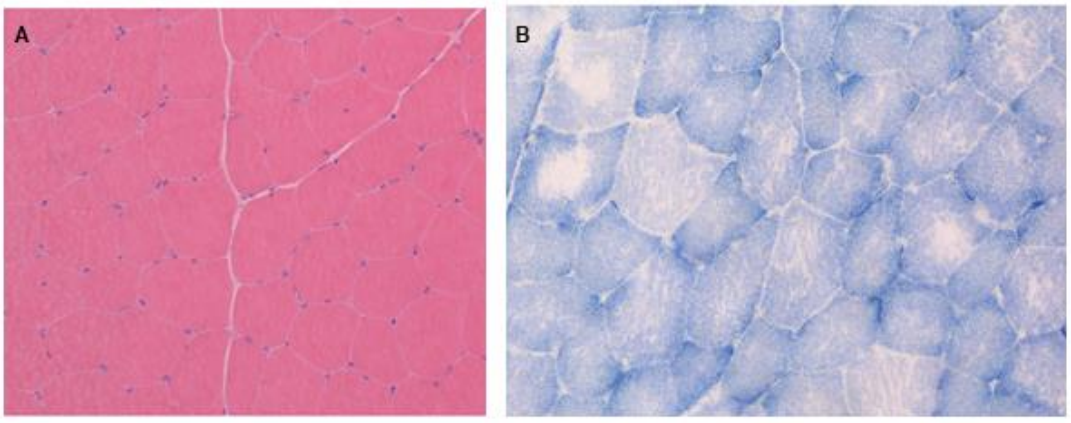


Figure 4.1 Muscle pathology of HMGCS1-myopathy. Staining of muscle sections from patient P1 reveal (A) occasional central nuclei by haematoxylin and eosin staining, and (B) multiminicore features by succinate dehydrogenase staining.

4.6.2 Genetic and *in silico* analyses of *HMGCS1* variants

In all families, screening of genes associated with rigid spine syndrome and congenital myopathy genes listed in the neuromuscular disease gene table[44] identified no obvious candidate variant/s. A total of five biallelic *HMGCS1* variants were subsequently identified by next generation sequencing in four affected individuals from three unrelated families. These included four missense variants and a frameshift variant.

In Family 1, whole exome sequencing analysis revealed a homozygous missense variant c.1339T>C, p.(Ser447Pro) in *HMGCS1* (ENSG00000112972) of the two affected siblings (P1, P2). The variant was absent in gnomAD data (v2.1.1) and heterozygous in the unaffected parents. Segregation was confirmed by PCR and Sanger sequencing (Figure 4.2B). The S447 residue was conserved in multiple species including the HMGCS2 paralogue (Figure 4.2C). The p.S447P substitution was predicted to be tolerated by multiple *in silico* prediction tools (Supplementary Table 4.2).

Additional rigid spine syndrome cases with *HMGCS1* variants were ascertained via international collaborations and gene-matching platforms, including The Broad Institute's *matchbox*. In Family 2, genetic screening for the isolated proband (P3) revealed compound heterozygous variants in *HMGCS1*: c.86A>T, p.(Gln29Leu) and c.344_345del, p.(Ser115Trp*fs**12). Both variants were absent from gnomAD, and the Q29 residue was conserved across vertebrate and invertebrate species (Figure 4.2C). The Q29L substitution was predicted to be deleterious by Provean but was tolerated by SIFT and PolyPhen (Supplementary Table 4.2).

In Family 3, trio whole genome sequencing for the isolated proband (P4) revealed compound heterozygous variants in *HMGCS1*: c.209T>C, p.(Met70Thr) and c.803G>C, p.(Cys268Ser). The c.209T>C, p.(Met70Thr) was absent from gnomAD and predicted to be deleterious by most *in silico* tools (Supplementary Table 4.2). Interestingly, M70 was not conserved among orthologues (Figure 4.2C). The second variant, c.803G>C, p.(Cys268Ser) was present in 4 of 251,086 alleles in gnomAD (allele frequency 1.59×10^{-5}) although only in the heterozygous state. This variant was predicted by most tools to be deleterious (Supplementary Table 4.2). Conservation analysis indicated that C268 is highly conserved (Figure 4.2C).

The c.209T>C, p.(Met70Thr) variant was also identified via gene matching in an additional apparently unrelated proband with severe rigid spine syndrome. This variant was heterozygous in the patient and inherited from the unaffected father. Whole genome sequencing and RNA sequencing of patient fibroblasts did not identify a second likely pathogenic *HMGCS1* variant *in trans* for this patient. Thus, it remains inconclusive to whether *HMGCS1* is the underlying genetic cause of rigid spine syndrome for this patient.

Three of the five variants (p.Q29L, p.M70T and p.Ser115Trp*fs**12) occur in exon 3 of *HMGCS1* (Figure 4.2A). The two remaining variants (p.C268S and p.S447P) occur in exons 6 and 10, respectively (Figure 4.2A). The C268S substitution is adjacent to a coenzyme A (CoA) binding residue, K269 (Figure 4.2A; Figure 4.3A, C). Positions of the remaining substitutions were distant from residues of the CoA active sites (E95, C129, H264), binding sites (S221, K269, K273) and residues of the dimer salt bridge (D119, E121, R194, D208, K239, K461 and H462). Substitutions p.M70T, p.S447P and p.Q29L appear to cluster towards the surface of HMGCS1 (Figure 4.3). Interestingly, substitutions p.Q29L (Family 2) and p.S447P (Family 1) were <5 Å apart in 3D space (Figure 4.3).

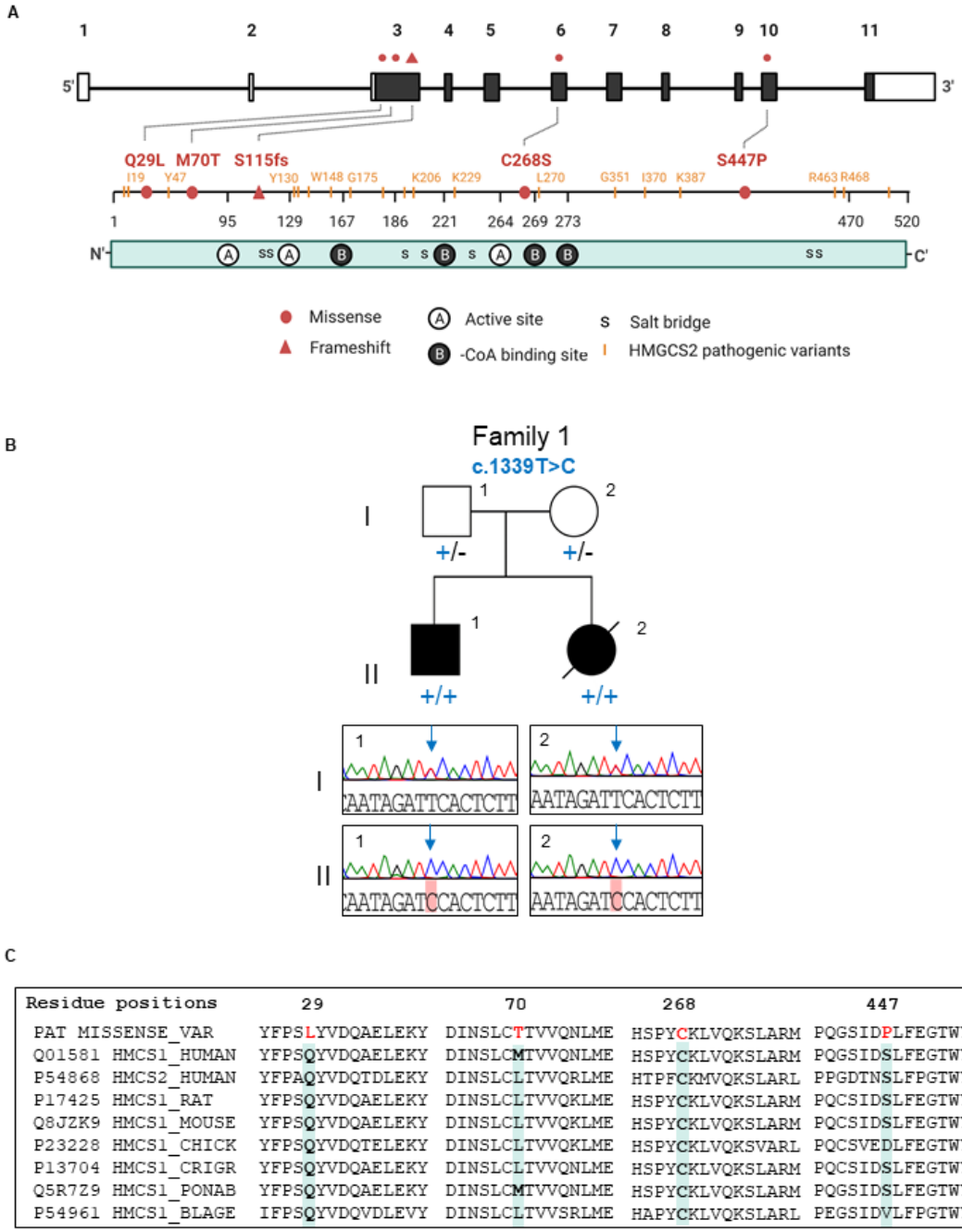


Figure 4.2 *HMGCS1* variants in the rigid spine syndrome cohort. (A) Schematic representations of the *HMGCS1* gene (top) and protein (bottom). The gDNA structure of *HMGCS1* includes non-coding exons (empty boxes) and coding exons (filled boxes). The red dots and the triangle respectively represent positions of the missense ([c.86A>T; p.Q29L], [c.209T>C; p.M70T], [c.803G>C; p.C268S], [c.1339T>C; p.S447P]) and truncating (c.344_345del; p.Ser115Wfs*12) variants identified in patients with rigid spine myopathy. Letters A, B, and s in the HMGCS1 protein representation (bottom) respectively represent residues of the active site (E95, C129, H264), coenzyme A binding site (N167, S221, K269, K273) and salt bridge (D119,

E121, R194, D208, K239, K461 and H462). Orange lines correspond to representative pathogenic HMGCS2 variants annotated according to HMGCS1 residue positions. Diagram not to scale. (B) Pedigree and Sanger confirmation of the c.1339T>C variant in Family 1. (C) Conservation of the four missense variant positions in various species and in HMGCS2; the mitochondrial paralogue.

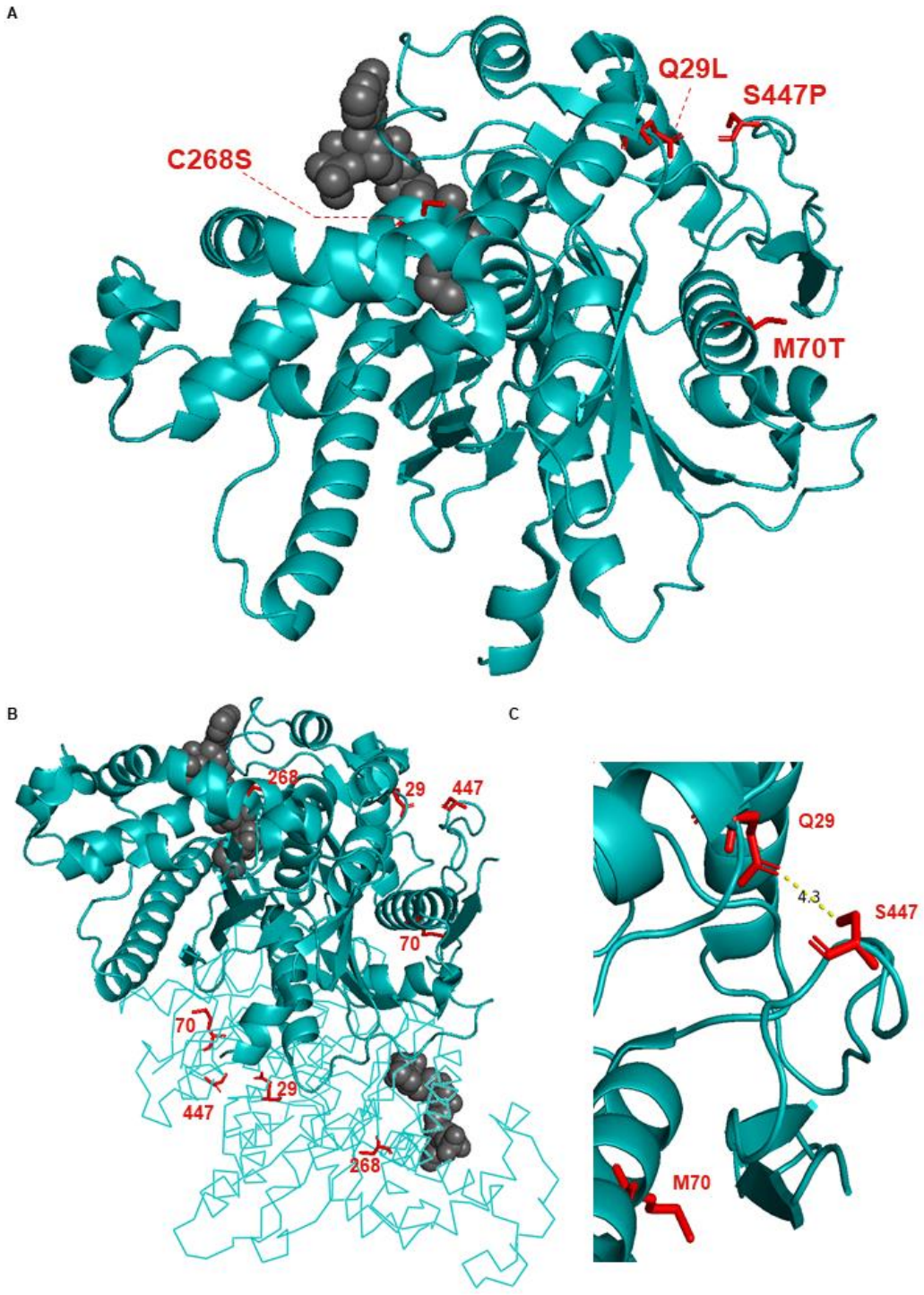


Figure 4.3 Protein modelling of the HMGCS1 missense variants. (A-C) Cartoon representations of the HMGCS1 protein monomer and annotated variant positions

represented as red liquorish. (B) Representation of the HMGCS1 homodimer with chain A (ribbon) and chain B (cartoon). (C) Close-up of substitutions at position Q29, M70 and S447. Minimum distance between S447 and Q29 is ~4.3 Å. Models were visualised with PyMOL.

4.6.3 *HMGCS1* is abundant in various tissues, including developing and adult skeletal muscle

RNA-seq expression data from the NCBI BioProject (PRJEB4337)[45] and the Genotype-Tissue Expression (GTEx) portal[46] indicated that *HMGCS1* is enriched in brain and liver and is expressed at relatively lower levels in various other tissues, including skeletal muscle. Similarly, qPCR analyses indicated enriched expression of *HMGCS1* in human cortex, and relatively lower expression in cells and tissues, including myotubes and adult skeletal muscle (Figure 4.4A). Of the myogenic samples analysed in the FANTOM5 database, *HMGCS1* expression appeared highest in skeletal muscle satellite cells (235-462 transcripts per million: TPM) relative to fetal-derived skeletal muscle cells (3-326 TPM), myotubes (0-240 TPM), and myoblasts (31-50 TPM). *HMGCS1* expression in cerebral cortex astrocytes ranged between 189-477 TPM.

Western blotting for HMGCS1 in various human tissues and cells revealed a band at ~57 kDa, which matched the product size detected when using recombinant HMGCS1 protein (Figure 4.4B-E). We show that HMGCS1 is enriched in cortex, consistent with the transcriptomic data and the proteomic data from the Human Protein Atlas[47] (Figure 4.4B). In addition, HMGCS1 was detected in adult human skeletal muscle, unlike the proteomic data from the Human Protein Atlas.[47] HMGCS1 was also detected in human myoblasts and myotubes from days 0 to 12 of differentiation Figure 4.4E, F) but was poorly detected in human fetal skeletal muscle (Figure 4.4B). HMGCS1 was present at similar levels in mouse EDL (*extensor digitorum longus*) and soleus muscles (Figure 4.4D), suggesting HMGCS1 is similarly abundant in type I and II myofibres.

HMGCS1 was similarly abundant in skeletal muscle of P1 (Family 1) and control skeletal muscle biopsies (Figure 4.4C). Thus, the p.(Ser447Pro) substitution does not appear to reduce protein stability or abundance.

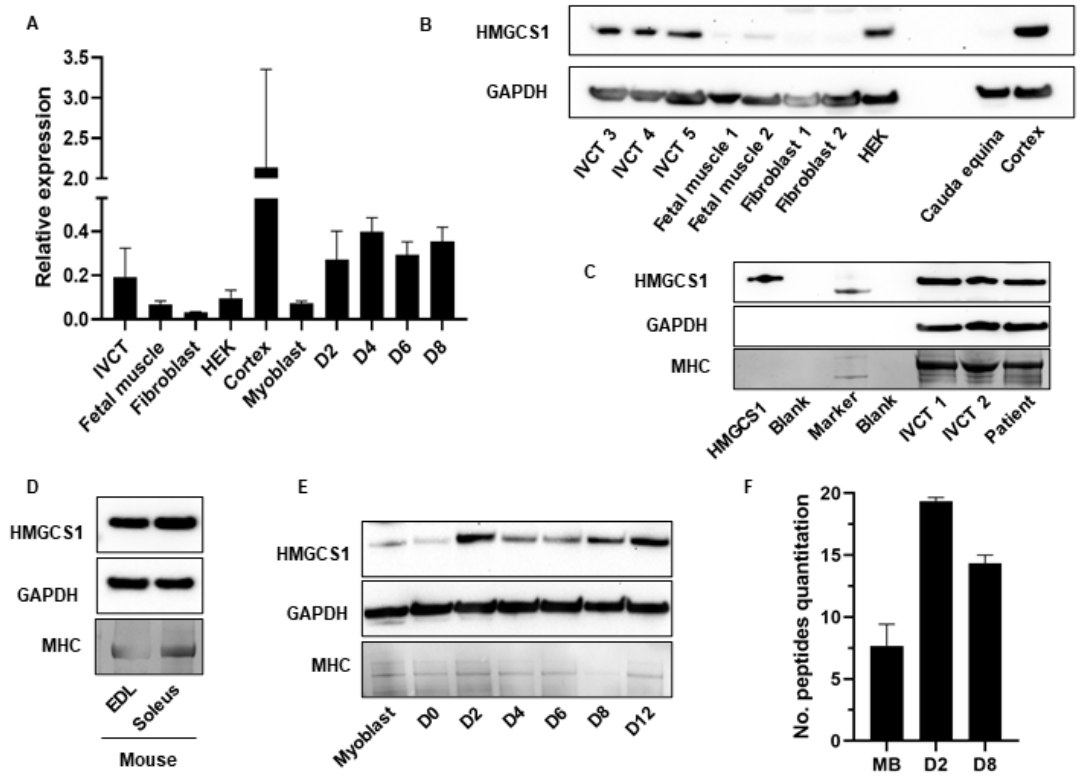


Figure 4.4 *HMGCS1* is expressed in various tissue including patient skeletal muscle. (A) Quantitative real-time PCR analysis using cDNA from healthy human cell lines and tissues. Expression data from human embryonic kidney cells (HEK293FT) and fibroblasts (n = 3 technical), myoblasts, myotubes at days D2, D4, D6, D8 of differentiation, cortex, skeletal muscle controls from *in vitro* contractile testing (IVCT) and fetal muscle (n = 3 biological). Expression normalised to *EEF2* and *TBP* geometric means using the delta Ct method. Data presented as mean ± SEM. (B-E) Western blotting for HMGCS1 and GAPDH in (B) control human tissue and cell lines, (C) patient (P1) and IVCT skeletal muscle, (D) EDL and soleus of mouse, and (E) primary human (Cook Myosite) myoblasts and myotubes from days D0, D2, D4, D6, D8 of differentiation. (C) Western blotting for recombinant wild-type HMGCS1 (left) used as a control for antibody specificity. Western blots for GAPDH (B-E) and the gel-stained myosin heavy chain (MHC; C-E) are shown as loading controls. (F) Graphical presentation of HMGCS1 peptides detected by quantitative mass spectrometry from myoblasts (MB) and myotubes at days 2 (D2) and 8 (D8) of differentiation.

4.6.4 HMGCS1 immunofluorescence in skeletal muscle

Immunostaining with the validated rabbit polyclonal antibody used for western blotting (MyBioSource [MBS2026097]) revealed HMGCS1 to be present within myofibres, with a stronger signal in type II myofibres (Figure 4.5). These signals were

lost when the antibody was preincubated with excess recombinant HMGCS1, suggesting they were specific. There was also strong reactivity and punctate signals within blood vessels and in the periphery of myofibres. However, these signals were not altered by preincubation with recombinant protein, suggesting some non-specific staining. The possibility of non-specific staining by immunofluorescence using the HMGCS1 antibody renders the interpretation of these data inconclusive.

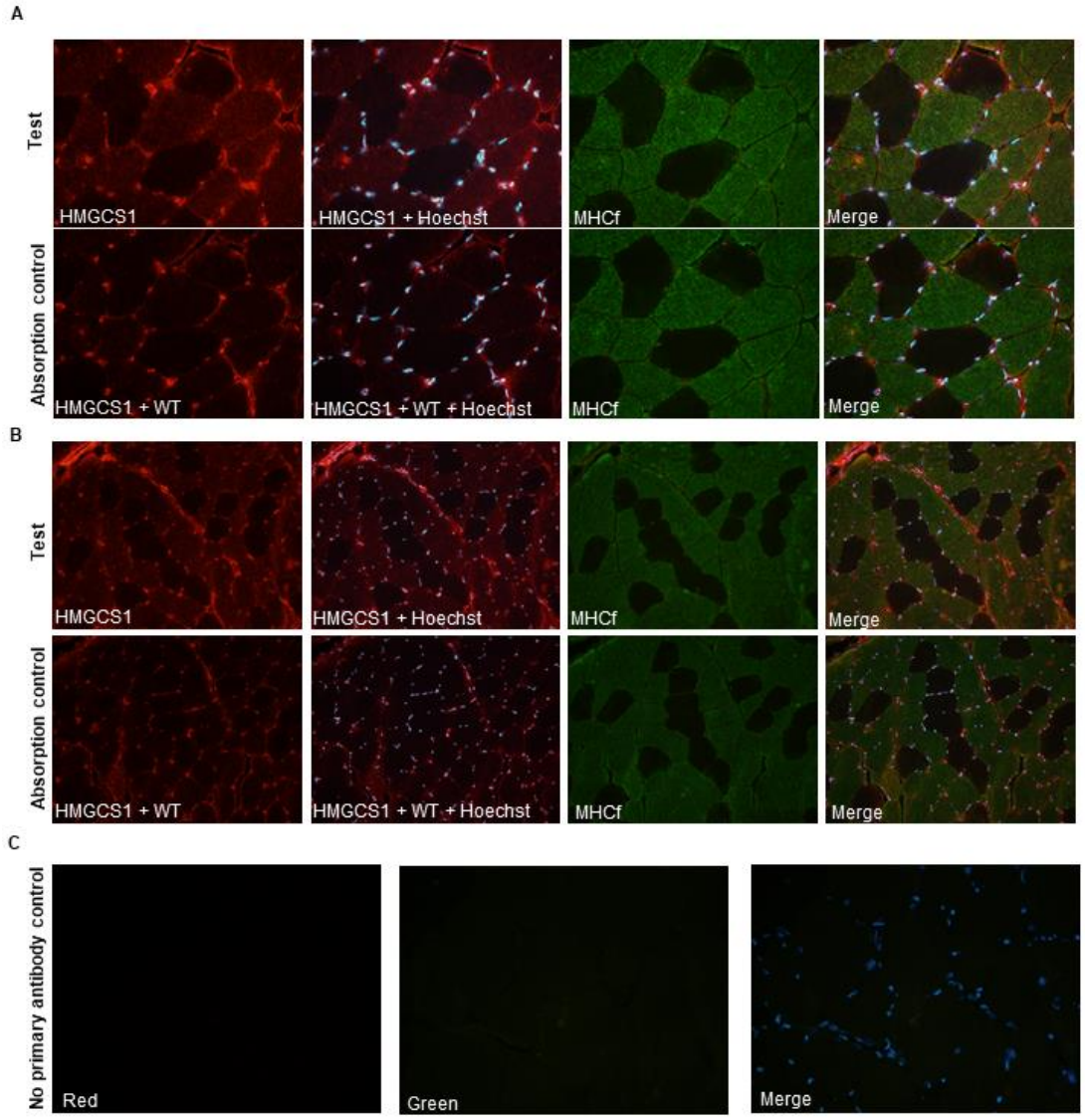


Figure 4.5 HMGCS1 appears to be type II fibre specific. (A,B) Cross sections of adult human skeletal muscle controls from *in vitro* contractile testing (IVCT) stained with a HMGCS1 rabbit polyclonal antibody (red; MBS2026097), a MHCf mouse monoclonal antibody (green; NCL-MHCf) and Hoechst (blue). Absorption controls contain the HMGCS1 antibody incubated with denatured recombinant HMGCS1^{WT} (WT) to assess antibody specificity. (C) No antibody primary controls consist of

sections incubated with blocking solution. Images taken with a 40X (A,C) and 20X (B) objective.

4.6.5 The p.S447P substitution does not affect HMGCS1 conformation

To explore the pathobiology associated with HMGCS1 substitutions, a range of assays were performed using recombinant WT and p.Ser447Pro human HMGCS1 proteins. Studies of the other three substitutions are ongoing. HMGCS1 has been shown to exist as a dimer in nature.[26,48] We initially predicted that the steric hindrance introduced by the p.S447P substitution may affect the secondary and tertiary structures of HMGCS1 and thus destabilise the dimerisation interphase. However, size exclusion chromatography indicated that recombinant wildtype ($HMGCS1^{WT}$) and mutant ($HMGCS1^{S447P}$) proteins eluted at equivalent points corresponding to the expected molecular weight for dimerised HMGCS1 (~160 kDa; Figure 4.6A). There were no indications of monomeric HMGCS1 which suggested that the p.S447P substitution does not interfere with HMGCS1 dimerisation (Figure 4.6A). CD spectra analyses identified no significant differences in the secondary structures of $HMGCS1^{WT}$ and $HMGCS1^{S447P}$ (Figure 4.6C).

4.6.6 The p.(Ser447Pro) substitution may partially reduce HMGCS1 thermostability

CD melt curve analyses indicated a potential decrease in the melting temperature (Tm; temperature at which 50% of protein is denatured) of $HMGCS1^{S447P}$ (Figure 4.6D) compared with WT. The Tm's were determined by interpolating the IC50 value (point of inflection) from a sigmoidal logistic regression four parameter fit (n = 3).[37] The average Tm of $HMGCS1^{S447P}$ was ~3.62°C lower than the Tm of $HMGCS1^{WT}$ (Figure 4.6D; Table 4.2) and was statistically significant (two-tailed unpaired t-test; P = 0.0244) despite the variability within the data (Figure 4.6D). Whether such a partial difference is biologically significant warrants further investigation.

Table 4.1 Circular dichroism melt curve data

HMGCS1	WT	S447P	% Difference (S447P/WT)
Tm (IC50); °C	47.01	43.39	92.3
95% CI (IC50); °C	45.57 to 48.45	41.92 to 44.87	
R squared	0.7412	0.7066	

4.6.7 HMGCS1^{S447P} enzyme activity is moderately dysregulated

The apparent K_m and V_{max} of recombinant HMGCS1^{WT} and HMGCS1^{S447P} for Ac-CoA and AcAc-CoA were determined by fitting reactions to Michaelis-Menten curves (Ac-CoA R^2 >0.89 and AcAc-CoA R^2>0.84, respectively; Table 4.2). For wildtype HMGCS1^{WT}, the K_m for Ac-CoA ($K_m^{Ac\text{-}CoA}$) was determined as 5.4 μM (95% CI 3.4 - 8.3 μM, Table 4.2). This was lower than the $K_m^{Ac\text{-}CoA}$ reported by Rokosz *et al.*[49] for recombinant human HMGCS1 (29 ± 7 uM). The K_m for AcAc-CoA ($K_m^{AcAc\text{-}CoA}$) was undetermined by Rokosz *et al.*[49] due to complications of substrate inhibition.[48,49] While our data indicated similar substrate complications at AcAc-CoA concentrations <25 uM (Supplementary Fig 4.2), reactions performed at >12.5 uM AcAc-CoA conformed with Michaelis-Menten kinetics, thus enabling the determination of an apparent $K_m^{AcAc\text{-}CoA}$ of 43.2 μM (95% CI 29.65 - 62.67 μM; Table 4.2). Given the discrepancies between our results and the literature, we restricted the analysis to interpretation of the relative values of apparent K_m and V_{max} of HMGCS1^{S447P} and HMGCS1^{WT}.

Kinetic analyses indicated that HMGCS1^{S447P} is relatively more active than HMGCS1^{WT} when saturated with Ac-CoA (Figure 4.6E). The V_{max} for HMGCS1^{S447P} (n = 3) was ~1.4-fold (i.e., 140%) higher than the V_{max} of HMGCS1^{WT} (Figure 4.6E; Table 4.2). Alongside the increase in V_{max}, the apparent $K_m^{Ac\text{-}CoA}$ for HMGCS1^{S447P} increased by ~3-fold of HMGCS1^{WT} (Table 4.2), suggesting the S447 substitution reduces HMGCS1^{S447P} affinity for Ac-CoA. Contrastingly, when saturated with AcAc-CoA, HMGCS1^{S447P} appeared relatively less active than HMGCS1^{WT}. The V_{max} for HMGCS1^{S447P} (n>3) was ~56% that of HMGCS1^{WT} and the apparent $K_m^{AcAc\text{-}CoA}$ of HMGCS1^{S447P} was ~17-fold lower than that for HMGCS1^{WT} (Table 4.2). Given these parameters were deduced from data >12.5 μM, we were unable to calculate a complete confidence interval for the $K_m^{AcAc\text{-}CoA}$ for HMGCS1^{S447P}. Therefore, the values should be interpreted with caution. Nevertheless, a reduced $K_m^{AcAc\text{-}CoA}$ suggests that the p.S447P substitution increases HMGCS1^{S447P} affinity for AcAc-CoA. These observations suggest a shift occurs in the substrate affinity of HMGCS1^{S447P} which may cause HMGCS1^{S447P} to retain moderate activity but behave differently to the wildtype at different substrate concentrations. Knowledge of the physiological substrate concentrations is required to determine how differently HMGCS1^{S447P} may behave in the patients.

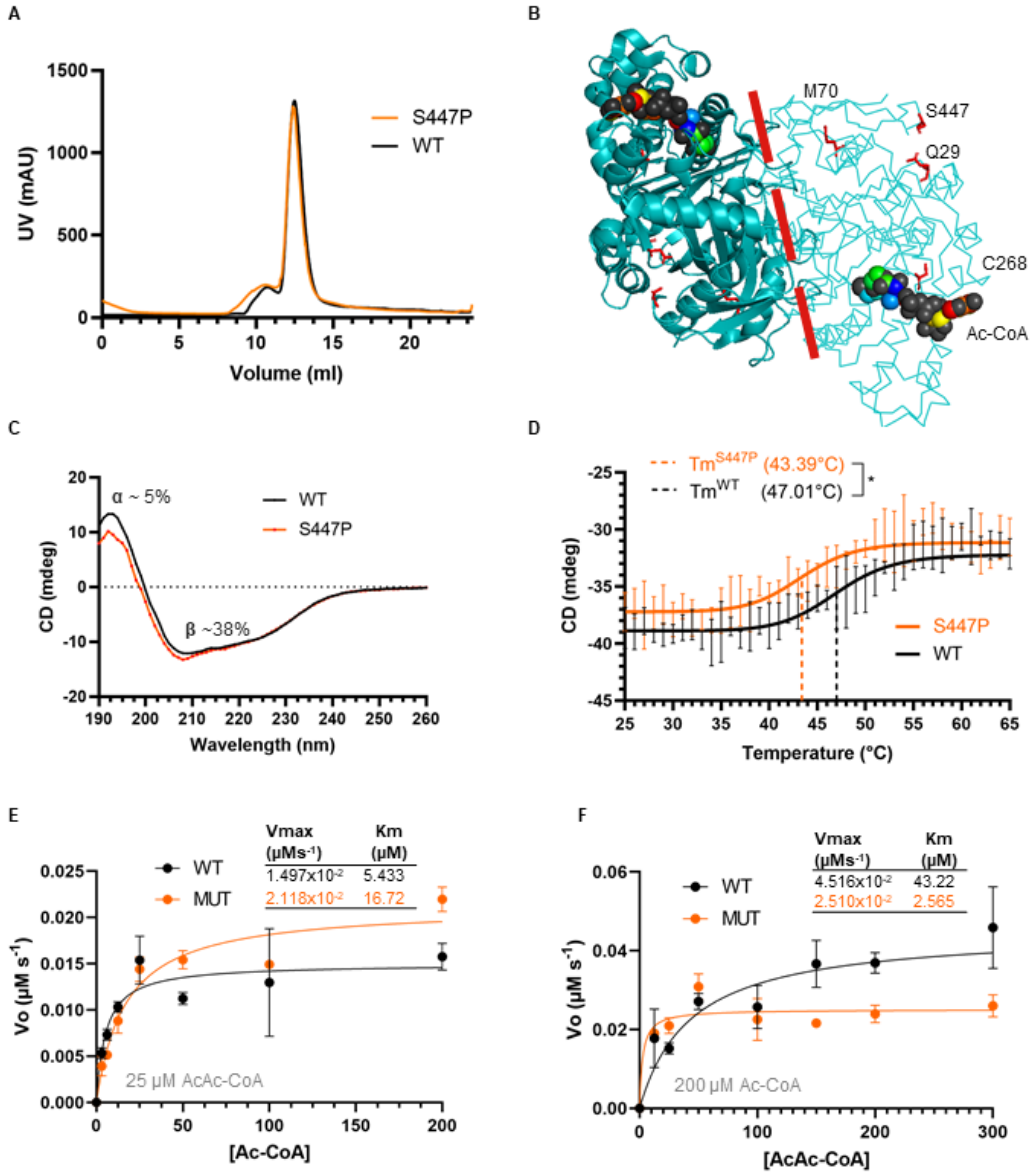


Figure 4.6 Structural and functional comparisons of wildtype and mutant HMGCS1. A) Size exclusion chromatogram of HMGCS1^{WT} and HMGCS1^{S447P} eluting at ~12 ml corresponding to the dimer state. Proteins eluted by a 10/300 GL Superdex 200 column (GE Healthcare). (B) Modelling of HMGCS1 dimer (PDB: 2P8U). Chain A (left) represented as a Cartoon. Chain B (right) represented as a Ribbon. Ac-CoA substrate represented as rainbow spheres. Dimerisation interphase indicated by red dashed line. Residues Q29, M70, C268S and S447 represented as red liquorish. (C). Average circular dichroism (CD) spectra of HMGCS1^{WT} and HMGCS1^{S447P} [0.1 mg/mL]; 25°C, pH 8.14 (n = 3). (D) Average CD melt curves of HMGCS1^{WT} and HMGCS1^{S447P} measured at 220 nm, [0.5 mg/mL] pH 8.14. Melting temperatures (Tm) were determined by interpolation of a sigmoidal logistic regression four parameter fit (n = 3). The point of inflection was taken as the Tm. Significance

was measured using a two-tailed unpaired t-test, n = 3). The asterisk indicates statistical significance (*P < 0.05).

Table 4.2 Apparent kinetic parameters for HMGCS1^{WT} and HMGCS1^{S447P}

HMGCS1	WT	S447P	Fold Difference (S447P/WT)
Ac-CoA (0-200 µM), 25 µM AcAc-CoA			
V_{max} (µMs^{-1}) [95% CI]	0.01497 [0.01359 - 0.01645]	0.02118 [0.01926 - 0.02334]	1.41
$K_m^{Ac\text{-}CoA}$ (µM) [95% CI]	5.433 [3.403 - 8.286]	16.72 [11.91 - 23.40]	3.08
R squared	0.89	0.94	
AcAc-CoA (12.5-300 µM), 200 µM Ac-CoA			
V_{max} (µMs^{-1}) [95% CI]	0.04516 [0.04070 - 0.05075]	0.02510 [0.02331 - 0.02699]	0.56
$K_m^{AcAc\text{-}CoA}$ (uM) [95% CI]	43.22 [29.65 - 62.67]	2.565 [??? - 6.212]	0.06
R squared	0.89	0.84	

4.7 Discussion

We report the identification of five biallelic variants in the *HMGCS1* gene in three unrelated families with severe rigid spine syndrome. Although the cohort thus far includes four patients, the distinct clinical descriptions shared by the patients, including scoliosis, respiratory insufficiency, disease onset, severity, and worsening of disease with infections is suggestive of a single disease entity which we define here as HMGCS1-related myopathy.

HMGCS1 encodes a cytosolic HMG-CoA synthase; a key enzyme of the mevalonate pathway which catalyses the condensation reaction between AcAc-CoA and Ac-CoA to form HMG-CoA.[26] The mevalonate pathway branches into several critical pathways including the sterol biosynthetic pathway, the isoprenoid pathway and ubiquinone pathway, which are essential for regulating cell proliferation, maturation, and maintenance.[50] Genes of these pathways, including *HMGCS1* are conserved in eukaryotes, archaea and some bacteria.[26,50-52] According to the International Mouse Phenotyping Consortium, knockdown of most genes of the mevalonate and subsequent pathways are associated with complete preweaning lethality (Figure 4.7).[53] Knockdown of *Hmgcs1* in mouse was associated with prenatal lethality (https://www.mousephenotype.org/data/genes/MGI:107592), no homozygous pups were observed at E9.5 and it is thought that this arises due to placental defects (personal communication, JAX labs staff). Placental defects are frequently observed

in mouse gene-knockout studies associated with embryonic lethality.[54] Consistent with its prenatal lethal phenotype, *HMGCS1* is highly intolerant to loss of function variants, with only four heterozygous variants reported out of 141,546 samples in gnomAD (probability of loss of function intolerance; pLI=1).[55] Additionally, part from one homozygous missense variant in one individual (p.(Ser227Thr)), all missense variants reported in gnomAD are heterozygous suggesting that biallelic variants may be unfavourable and deleterious (missense constraint score Z=3.04).[55]

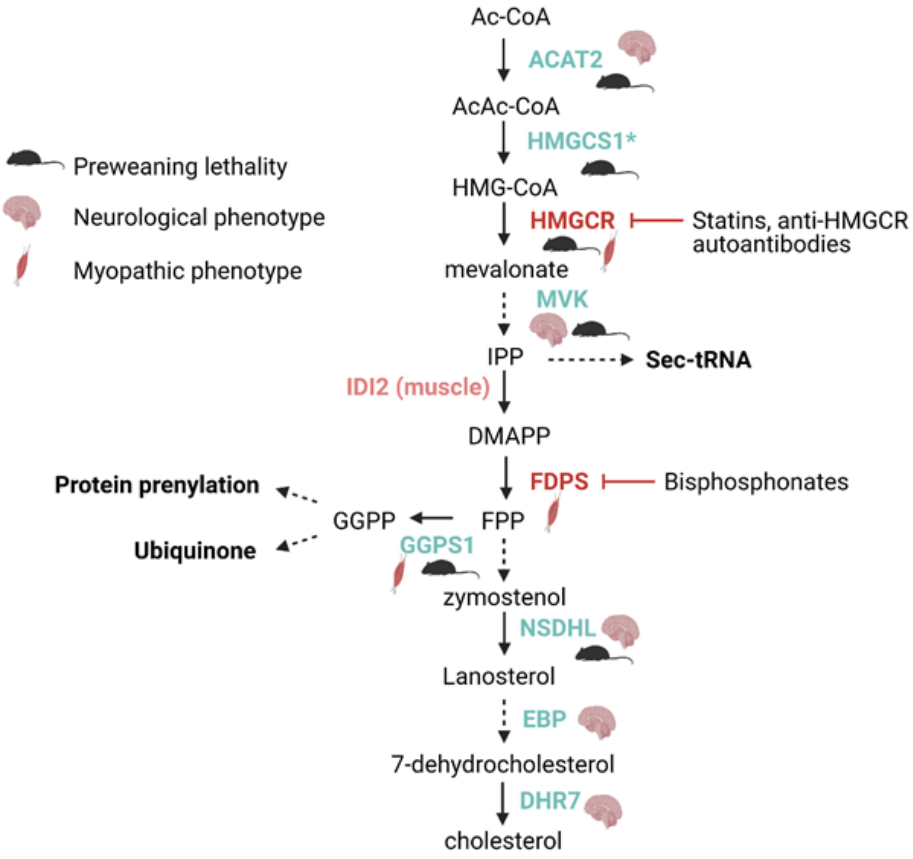


Figure 4.7 Diagrammatic summary of the mevalonate pathway. Representative enzymes of the mevalonate pathway implicated in monogenic neuromuscular disorders (blue). HMGCS1, the candidate for rigid spine syndrome is indicated by an asterisk. Enzymes associated with drug-induced and autoimmune-induced myopathy (red). Mouse symbols represent genes associated with preweaning lethality when knocked down in mice. Muscle symbol represents genes associated with myopathic manifestations. Brain symbol represents genes associated with neurological manifestations. IDI2 (pink) is a skeletal muscle specific isopentenyl diphosphate isomerase isoform not yet associated with disease. Diagram produced by BioRender.com.

While *HMGCS1* has not yet been associated with Mendelian disease, biallelic variants in the mitochondrial paralogue; *HMGCS2* (3-hydroxy-3-methylglutaryl-CoA synthase 2), are associated with childhood-onset HMG-CoA synthase deficiency (OMIM# 605911) characterised by episodic metabolic distress, febrile crisis, hepatomegaly, and respiratory distress.[56-58] At least 37 variants have been reported in 43 affected patients to date.[58,59] HMGCS1 and HMGCS2 share ~67% sequence similarity and the substituted residues in HMGCS1 (Q29, C268 and S447) are conserved in HMGCS2 (Q66, C305 and S484, respectively) suggesting they may be of structural or functional importance. Thus far, there appears to be no overlap between the reported HMGCS2 pathogenic substitutions[59] and the HMGCS1 substitutions (Figure 4.2A). Notably, according to gnomAD constraint scores, *HMGCS1* (pLI=1; Z=3.04) is substantially more intolerant to loss of function and missense variation compared to *HMGCS2* (pLI=0; Z=-0.19), thus supporting that biallelic variation in *HMGCS1* can be deleterious. While *HMGCS2* is restricted to mammals[52], studies of a homozygous missense variant in *hmgcs1* (p.H189Q) of zebrafish identified by an ENU (N-ethyl-N-nitrosourea) mutagenesis screen[60] has been associated with multiple congenital anomalies affecting myelination[60], craniofacial development[61], and erythropoiesis.[62] This reside is highly conserved across species (p.H200 in human) but not conserved in human HMGCS2. Although a muscle-related phenotype has not been associated with this particular variant, Hernandez *et al*.[62] suggest that *HMGCS1* may be associated with variable congenital phenotypes given its critical roles in cholesterol synthesis across various body systems.[61,62]

Thus far, we have characterised one of the four HMGCS1 substitutions, p.S447P, by recombinant protein investigations. As this residue is positioned in a loop stabilised by significant H-bond interactions, we initially predicted that the substitution to proline may be sufficient to increase the flexibility of the loop, destabilise secondary structures, and indirectly affect dimerisation. However, unlike initial speculations from protein modelling, our CD structural analyses and size exclusion chromatography indicated that HMGCS1^{S447P} retains its 3D structure and dimerisation tendency. This is consistent with the superficial position of the p.S447P substitution and its distance from the HMGCS1 dimerisation interphase (Figure 4.6B). Whilst HMGCS1 surface substitutions have not been previously characterised, studies of other recombinant HMGCS1 mutants (p.C129A; *H. sapiens*[49] and p.A110G; *E. faecalis*[48]) with

substitutions close to the active site have also indicated no major difference in HMGCS1 structure or conformation. However, subtle albeit significant changes were detected by X-ray crystallography.[48] We thus speculate that the structural effect of the S447P substitution (if any) may be partial and local, and therefore beyond the detection capacity of CD. This hypothesis could be evaluated by X-ray crystallography.[48]

Our CD melt curve analyses indicated that the p.S447P substitution may partially reduce the melting temperature (Tm) of $HMGCS1^{S447P}$ ($P < 0.05$; Figure 4.6D). A reduction in Tm from 47°C to 43°C could suggest that $HMGCS1^{S447P}$ is partially less stable than $HMGCS1^{WT}$, though this would need to be ascertained using a biologically relevant temperature-based assay. It is noteworthy that several inborn disorders of the mevalonate pathway have been associated with partial impairment in temperature sensitivity and/or stability when assessed between 30°C and 50°C.[63,64] This includes cases of hyper-IgD syndrome (OMIM# 260920) associated with biallelic variants in *MVK* (mevalonate kinase). Of note, *MVK* (pLI=0.17; Z-score=0.94)[55] is less constrained to variation than *HMGCS1* (pLI=1; Z=3.04)[55] and knockdown of *Mvk* in mouse also leads to embryonic lethality.[53] Thermal inactivation experiments in patient-derived fibroblasts with a common substitution in MVK (p.V377I) have been associated with reduced MVK activity by 1.3-fold at 37°C and 50°C.[63] At 30°C, the mutant p.V377I protein was as stable as the wildtype.[63] In addition, MVK activity in patient peripheral blood mononuclear cells were reduced by 2- to 8-fold during febrile attacks.[63] It has been suggested that partial changes in enzyme activity due to temperature sensitivity can have far-reaching pathogenic consequences.[63] The partial reduction in MVK activity upon minor elevations in physiological temperature can trigger a chain of events as MVK activity becomes rate-limiting. This can lead to a temporary deficiency of isoprenoid products and subsequently trigger inflammation and fever.[63]

Some variants in *NSDHL* (NAD(P) Dependent Steroid Dehydrogenase-Like) have also been associated with hypomorphic temperature sensitivity that lead to CK-syndrome (OMIM# 300831).[64,65] Expression studies in yeast indicated that at 30°C, the structure and function of mutant NSDHL were comparable with wildtype NSDHL. However, at 37°C, mutant NSDHL proteins lost the ability to correctly fold and were susceptible to protein degradation.[64] Of note, *NSDHL* also appears relatively less constrained to

loss of function variation than HMGCS1 according to gnomAD (pLI=0.96; Z-score=0.87).[55] Loss of function variants in *Nsdhl* are also associated with placental defects and embryonic lethality in hemizygous male mice.[66] It is tempting to postulate that HMGCS1^{S447P} may also be susceptible to partial temperature-sensitive effects in our patients. Indeed, it would be consistent with the worsening of phenotypes associated with episodic febrile attacks. Further, given the severe prenatal lethality in *Hmgcs1* null mouse pups, it is unlikely that variants identified in our patients would result in dramatic reduction in HMGCS1 function. One might expect that more deleterious variants would be incompatible with life in humans and would also result in prenatal lethality.

Although the p.S447P substitution is distant from the HMGCS1 catalytic sites, we predicted the substitution may have some effect on HMGCS1 activity, given that HMGCS2 substitutions distant from the catalytic sites have been associated with reduced activity.[67] Consistently, our Michaelis-Menten analyses revealed moderate differences in the substrate affinities and behaviour of HMGCS1^{S447P} and HMGCS1^{WT} under different substrate concentrations (Figure 4.6; Table 4.2). Although modest impairment in enzyme activity may be debated to be clinically insignificant, a superficial substitution in HMGCS2 p.R505Q (p.R468 in HMGCS1) has been associated with disease despite retaining 70% of normal HMGCS2 activity.[67] This suggests that partial impairment in HMGCS1 activity may similarly elicit disease.[67] A hypomorphic partial loss in HMGCS1 activity would be a plausible pathomechanism given that the patients' disorders manifested during childhood and were slowly progressive. Moreover, complete dysfunction of HMGCS1 is expected to be unviable given that homozygous deletion of *Hmgcs1* is embryonic lethal. Nonetheless, the consequence of partial impairments in HMGCS activity associated with surface variants remains unclear.[67] One consequence of hypomorphic activity may involve abnormal accumulation and processing of metabolites. This can subsequently trigger dysregulation in the activity of other enzymes of the mevalonate pathway, a mechanism that has been postulated in other hypomorphic disorders of the mevalonate pathway.[51,63,64]

Given the superficial position of the p.S447 residue (Figure 4.3), we also speculated that p.S447 may be involved in HMGCS1-protein interactions that are impaired by the

substitution to proline. Although there are limited reports of HMGCS1-protein interactions in human, HMGCS of archaea has been shown to form a thermodynamically stable multienzyme complex with ACAT2.[68] We attempted to demonstrate that HMGCS1 and ACAT2 similarly complex in human as hypothesised (Supplementary Fig 4.3A)[68], however, analyses of the recombinant proteins by analytical ultracentrifugation yielded inconclusive results. This may be due to our reaction mixture solely comprising HMGCS1 and ACAT2 without any substrates, cofactors or scaffold proteins that may be needed to keep the enzymes biologically active and allow interaction *in vitro*.[68] Alternatively, we speculate that HMGCS1 may interact with other novel binding partners[50] such as muscle-specific proteins, which would clarify the myopathic features of this disorder. A yeast two-hybrid assay involving a skeletal muscle cDNA library could be used in future to investigate HMGCS1-protein interactions in a more biologically suitable context.[69]

Collectively, the apparent intolerance of *HMGCS1* to variation judged by the gnomAD constraint data, alongside the phenotypic data from the JAX labs and our functional data suggest a hypomorphic mechanism of disease associated with the *HMGCS1* variants identified, rather than substantial or complete loss of function. This is consistent with the mechanisms of several downstream genetic defects of the mevalonate pathway.[64] Apart from hypomorphic temperature-sensitive changes in NSDHL and MVK, pathogenic substitutions in *GGPS1* have similarly been reported to have no significant impact on GGPS1 conformation and only mildly impaired enzyme activity *in vitro* and in patient-derived fibroblasts.[10] In addition, hypomorphic mechanisms of DHR7 (Δ7-sterol reductase) and EBP (mopamil-binding protein) have been reported and are associated with slower progression of disease.[70,71] Collectively, these examples support that modest changes in HMGCS1 stability or function may be pathogenic.

Phenotypically, HMGCS1 appears a suitable candidate gene for rigid spine syndrome, given the features in our patients strongly resemble those reported across various acquired and inborn errors of the mevalonate pathway (Figure 4.7).[10,18] Features include myopathy, muscular dystrophy, scoliosis, joint contractures, hypotonia, elevated serum creatine kinase, recurrent infections, febrile episodes, and respiratory insufficiency.[10,64,65,72-78] Of the acquired disorders, the most recognised cause of

myopathy includes statin inhibition of HMGCR (HMG-CoA reductase), the enzyme immediately downstream of HMGCS1 (Figure 4.7).[18] It is estimated that 5-10% of statin users experience statin-induced myopathy, which is associated with elevated CK, myalgia and rhabdomyolysis.[79] Additionally, anti-HMGCR myopathy, an autoimmune disorder that inhibits HMGCR has been associated with features resembling muscular dystrophy with some aspects of the phenotype overlapping with the phenotype observed in our HMGCS1 cohort.[73,75,80,81] Further downstream, inhibition of FDPS (farnesyl diphosphate synthase) by bisphosphonate medications has been associated with profound muscle weakness with fever and flu-like side effects in some patients.[82,83] Of the inborn errors of the mevalonate pathway, GGPS1-related muscular dystrophy appears to most closely resemble HMGCS1-myopathy both clinically and morphologically.[10] Overall, the recurrent neuromuscular and rigid-spine phenotypes implicated in the mevalonate pathway (Figure 4.7) support HMGCS1 as an underlying genetic cause for our patients disorder.[10]

We have demonstrated that *HMGCS1* is expressed and detected in myoblasts, myotubes and skeletal muscle, supporting the myopathic phenotype of the patients (Figure 4.4). Although enrichment of *HMGCS1* in the brain may suggest a neurological phenotype, other genes of the mevalonate pathway including *HMGCR* and *GGPS1* are also more abundant in brain tissue compared to skeletal muscle.[29,45-47] However, both *HMGCR* and *GGPS1* have been predominantly associated with a myopathic phenotype in patients and animal models.[10,73,74,84-88] By Western blotting, Hmgcs1 was similarly detected in mouse EDL and soleus muscle, suggesting no obvious fibre-type predominance of Hmgcs1 protein abundance. We also examined protein localisation by performing immunofluorescence on human skeletal muscle biopsies. However, the possibility of nonspecific staining observed with the antibody by immunofluorescence challenged our interpretation of these results (Figure 4.5). There was some indication that HMGCS1 may be more enriched in type II myofibres, when co-stained with antibodies against type II MHC, however the interpretation remains inconclusive. Interestingly, type II fibre type specificity has been reported in HMGCR studies in rat.[84] This may suggest an important role of the mevalonate pathway in type II myofibres. Validation of the fibre type specificity of HMGCS1 and downstream enzymes may elucidate some of the unclear myopathic mechanisms of the mevalonate pathway.

To date, it remains poorly understood why disruptions to the mevalonate pathway are associated with a myopathic phenotype.[89] One possible explanation may be attributed to dysregulation of genes and isoenzymes exclusively expressed in the mevalonate pathway of skeletal muscle.[90,91] For example, IDI2 (isopentenyl diphosphate isomerase 2; Figure 4.7) is a skeletal muscle-specific isozyme of IDI1 which appears to be regulated by HMGCR abundance and activity.[91] In one study, HMGCR statin inhibition was associated with significant reduction of *Idi2* mRNA levels in mouse skeletal muscle, whereas *Idi1* levels remained unaffected.[91] This suggests that muscle-specific regulatory changes may differentially affect downstream pathways in skeletal muscle compared to other tissue.[91]

The mevalonate pathway diverges into various pathways critical for cell proliferation, maturation, and maintenance (Figure 4.7).[50] These include the cholesterol biosynthesis pathway, the isoprenoid pathway, and the ubiquinone pathway. There are various speculations regarding which of these pathways results in the myopathic phenotype. Some research suggests that inhibition of the cholesterol biosynthetic pathway reduces cholesterol abundance in the sarcolemma and leads to a dystrophic phenotype.[18] Other research suggests that inhibition of isoprenoid synthesis may affect *SELENON* expression by reducing isopentylation of selenocysteine-tRNA, thus leading to increased oxidative stress in skeletal muscle.[18,92,93] Notably, *SELENON* defects are one of the most common causes of rigid spine syndrome.[9] The clinical resemblance between patients with SELENON-related myopathies and *HMGCS1*-associated myopathy may suggest that these disorders have a shared molecular pathway or mechanism. Another theory involves disruptions in protein prenylation or geranylgeranylation in the GGPP pathway.[18] Disruptions to this pathway by statins have been associated with impaired mitochondrial function and impairment of Ca^{2+} homeostasis in skeletal muscle, which would also be consistent with the multiminicore pathology observed in skeletal muscle from some patients.[94] It remains unclear if the myopathic phenotype results from defects in one or multiple of the downstream mevalonate pathways. Nonetheless, the plethora of myopathic cases arising due to defects within this pathway supports the association of HMGCS1 dysfunction as a cause of rigid spine syndrome.

4.8 Future Directions

Future directions to this study include investigating the effects of the remaining substitutions (p.Q29L, p.M70T and p.C268S) on the stability, structure, and catalytic activity of HMGCS1. In addition, *hmgcs1* null zebrafish will be generated for phenotypic investigations, with a particular focus on skeletal muscle. A rescue experiment will be also performed to determine whether injection of the *hmgcs1* mutants identified in this study can rescue the phenotypes of *hmgcs1* null fish.

4.9 Conclusion

We report a novel autosomal recessive rigid spine syndrome we have termed *HMGCS1*-myopathy. We speculate that *HMGCS1* variants act by a hypomorphic mechanism, however the precise cause of the myopathic phenotype in the patients remains to be elucidated. Identification of additional *HMGCS1*-myopathy cases may help clarify the pathomechanisms of myopathies reported in disorders of the mevalonate pathway. This report expands the genetic causes of rigid spine syndrome and highlights mevalonate pathway as an essential pathway for muscle function. We encourage screening of *HMGCS1* in patients with unsolved rigid spine syndrome.

4.10 Acknowledgements

We would like to thank the patients and families for their participation in this study. We thank Associate Professor Evan Ingley and Associate Professor Robert Tuckey for their assistance with the enzyme assays. We gratefully acknowledge the funding bodies for their support.

4.11 Funding

LD is supported by an Australian Government Research Training Program (RTP) Scholarship. (Investigator Grant, APP2007769) and NGL (Fellowship APP1117510) are supported by the Australian NHMRC.

4.12 Competing interests

The authors report no competing interests

4.13 Supplementary material

4.13.1 Clinical history of Family 1

The proband (P1, Family 1) was a 10-year-old Spanish boy born to suspected consanguineous parents (Figure 4.2B). Appearing healthy at birth, he developed weakness, progressive hypotonia, scapular winging, scoliosis, and spinal rigidity particularly in the cervical region during the first decade. His muscle biopsy showed multiminicore-like structures without fibrosis, atrophy, necrosis, or predominance of type 1 fibres. He experienced recurrent respiratory infections and the severity of his disorder worsened with infection. His sister (P2) was also affected. She succumbed at 15 years of age after developing an infection followed by rapid decline in muscle strength and respiratory failure. The disorder was presumed to be recessive given the appearance of the disease trait in both siblings and the possibility of the parents being consanguineous.

Supplementary Table 4.1 Primer sequences for expression analysis of *HMGCS1* normalised to *EEF2* and *TBP*.

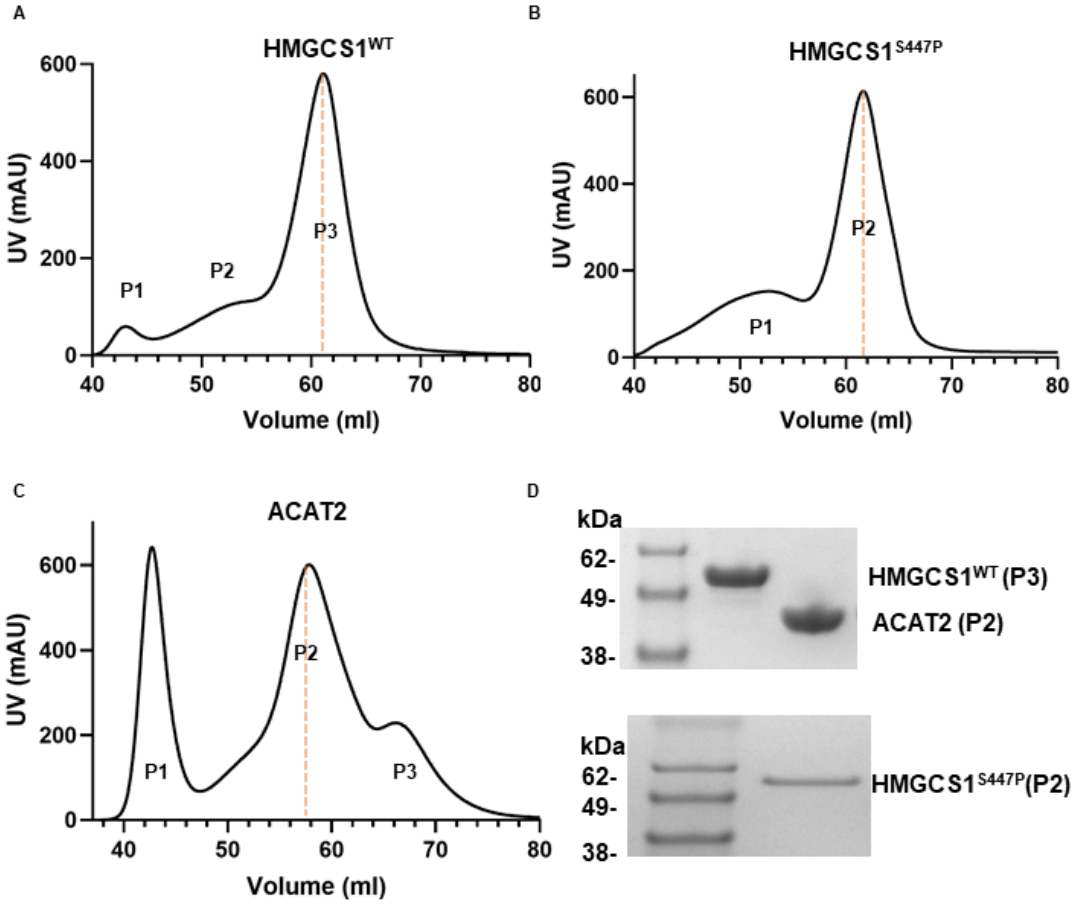
Gene	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	Amplicon size
HMGCS1	TTCCCTTGCATCTGTTCTAGC	TTTTATCAAGAGCAGACCCCG	149
EEF2	CCTTGTGGAGATCCAGTGTCC	CTCGTTGACGGGCAGATAGG	142
TBP	TTGTACCGCAGCTGCAAAAT	CGTGGTTCGTGGCTCTCTTA	149

Supplementary Table 4.2 Summary of the biallelic variants identified in the autosomal recessive HMGCS1 cohort.

Family	Nationality	C	Variant (NM_001098272.2)	gnomAD	CADD	PolyPhen-2	Provean	SIFT
1	Spanish	Y	A,B. c.1339T>C, p.(Ser447Pro)[Hom]	Absent	23.3	Benign	Neutral	Tolerated
2	Japanese	N	A. c.86A>T, p.(Gln29Leu) B. c.344_345del, p.(Ser115Trpfs*12)	A. Absent B. Absent	A. 23.1	A. Benign	A. Deleterious	A. Tolerated
3	American?	N	A. c.803G>C, p.(Cys268Ser)	A. $1.59x10^{-5}$	A. 27.7	A. probably damaging	A. Deleterious	A. Tolerated
			B. c.209T>C, p.(Met70Thr)	B. Absent	B. 23.9	B. Benign	B. Deleterious	B. Damaging
4	American?	N	A. c.209T>C, p.(Met70Thr) B. ?	A. Absent	A. 23.9	A. Benign	A. Deleterious	A. Damaging

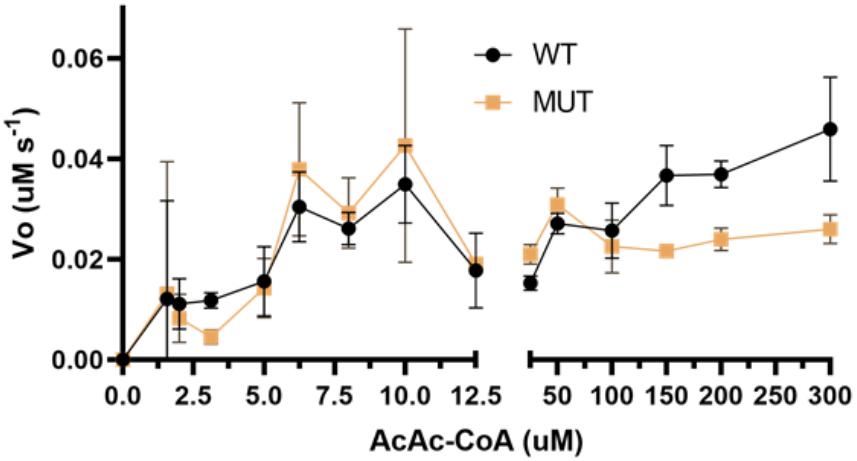
[C]Consanguineous
[A]First allele.
[B]Second allele.

4.13.2 Recombinant protein purification assessment



Supplementary Fig 4.1 Evaluation of recombinant HMGCS1 and ACAT2 purity and oligomerisation tendencies. (A-C) Size exclusion chromatograms of recombinant (A) HMGCS1^{WT} (B) HMGCS1^{S447P} and (C) ACAT2 using a XK16/600 Superdex 200 column. Peaks (P1-3) represent aggregates and oligomers of different sizes. Predicted HMGCS1^{WT} and HMGCS1^{S447P} dimers P3 (A) and P2 (B), respectively eluted ~ 61 mL. Predicted ACAT2 dimer (P2) eluted ~58 mL. (D) SDS-PAGE evaluation of fractions corresponding to dimerised protein. SDS-PAGE confirmed that HMGCS1 (~57kDa) and ACAT2 (~43 kDa) were of the expected size and of satisfactory purity.

4.13.3 Enzymatic assays



Supplementary Fig 4.2 Enzymatic reactions for HMGCS1^{WT} and HMGCS1^{S447P}. Initial velocities (Vo) plotted against different concentrations of AcAc-CoA. Substrate inhibition for both enzymes seen ~12.5-100 µM AcAc-CoA.

4.13.4 Analytical ultracentrifugation

4.13.4.1 Methodology

Given the positioning of S447 on the outer surface of HMGCS1, we hypothesised that the p.S447P substitution may impair protein interactions. HMGCS1 has been suggested to interact with ACAT2 (acetoacetyl Coenzyme A thiolase) in the dimerised states to form a ~400 kDa multienzyme complex in humans, as judged from structural comparisons of known HMGCS-ACAT2 interactions in archaea.[68] Superpositioning of human HMGCS1 (PDB 2P8U) and ACAT2 (1WL4) dimers to the *M. thermolithotrophicus* multienzyme complex (6ET9) in pyMOL indicated that the p.S447P substitution may be located near the binding interface with ACAT2 and thus possibly affect assembly of this hypothetical complex (Supplementary Fig 4.3A).

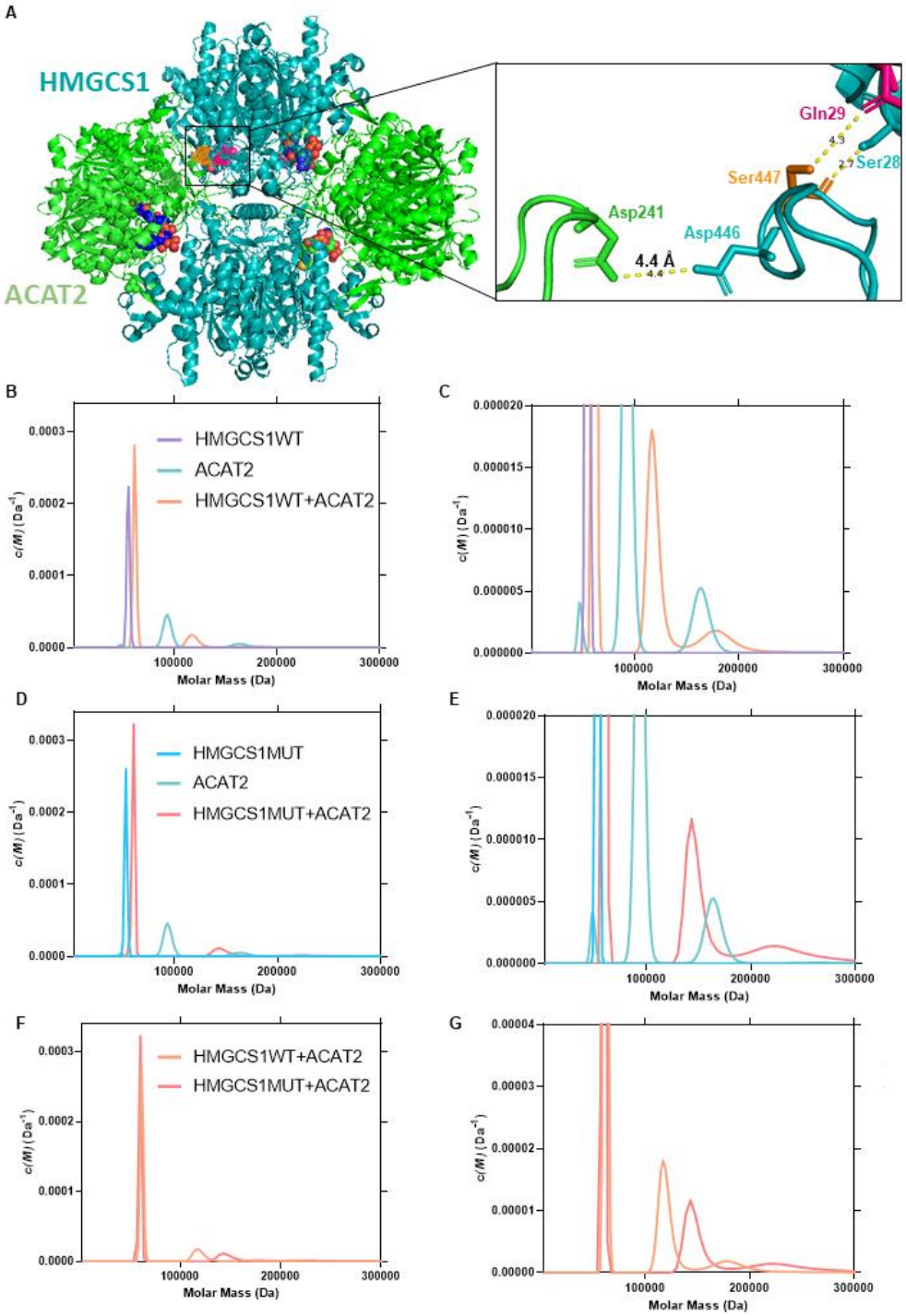
We attempted to investigate whether recombinant human HMGCS1^{WT} or HMGCS1^{S447P} complexed with ACAT2 *in vitro* using analytical ultracentrifugation (AUC). Protein production, purification and quantification were performed following previous methods.[37] Protein were eluted in a different size exclusion equilibration buffer containing 10% v/v glycerol, 20 mM Tris, pH 7.6, as described by Vögeli *et al*.[68] DTT was omitted given its interference with the AUC readings. Sedimentation velocity experiments were performed at 20°C as previously described[42], in a Beckman Coulter XL A analytical ultracentrifuge with a 4-hole and 8-hole An50-Ti rotor.[40,41]

380 µL of protein were loaded in double sector quartz cells at 1 mg/mL alongside 400 µL of the size exclusion equilibration buffer as a reference. Centrifugation was performed at 40,000 rpm. As previously described[95,96], the absorbance was fit as a function of radial position to the Lamm equation to determine the continuous sedimentation (c(*s*)) and continuous mass (c(*M*)) distribution. The c(*s*) and c(*M*) plots were determined for HMGCS1^{WT}, HMGCS1^{S447P} and ACAT2 when centrifuged individually, as well as when HMGCS1^{WT} or HMGCS1^{S447P} were complexed with ACAT2 at 1:1 ratio.

4.13.4.2 Results

Continuous mass distribution analyses at 1 mg/mL indicated that both HMGCS1^{WT} and HMGCS1^{S447P} had a molar mass of ~57 kDa, corresponding to the monomeric form. This was contrary to our size exclusion chromatography results and studies in the literature[48,49,68]. c(*M*) analyses of ACAT2 indicated two species of ~86 kDa and ~172 kDa corresponding to the ACAT2 homodimer and homotetramer respectively.

Analyses of the c(*M*) distributions for HMGCS1^{WT} and ACAT2 in combination did not indicate strong evidence of interactions between the species (Supplementary Fig 4.3B-G). Of note, there was a broadening of a peak corresponding to ~190-200 kDa (Supplementary Fig 4.3C), which may be indicative of multi-oligomeric species existing in an equilibrium. Similarly, analyses of HMGCS1^{S447P} and ACAT2 in combination revealed a broader peak ~200-300 kDa (Supplementary Fig 4.3E) suggesting the presence of a multi-oligomeric species existing within an equilibrium. However, it is difficult to conclude whether this corresponds to an interaction between the two different proteins, or solely between ACAT2 oligomers. Moreover, given the HMGCS1 proteins were analysed in the monomerised state, these results may be biologically invalid. Nevertheless, it is interesting that the broader peaks corresponding to HMGCS1^{WT}+ACAT2 (~190-200 kDa) and HMGCS1^{S447P}+ACAT2 (~200-300 KDa) appear to sediment differently (Supplementary Fig 4.3F,G).



Supplementary Fig 4.3 Analytical ultracentrifuge sedimentation velocity analyses. (A) Hypothetical model of the human HMGCS1:ACAT2 multienzyme complex containing two HMGCS1 dimers; cyan (2P8U), two ACAT2 dimers; green (1WL4) and four Ac-CoA substrates (rainbow). Closeup of S447 and local interactions shown on the right. (B-G) The molar mass distribution *c(M)* is plotted as a function of molar mass (Da) for recombinant wildtype HMGCS1 (purple), mutant HMGCS1^{S447P} (blue) and recombinant ACAT2 (green) when analysed separately and when HMGCS1-ACAT2 proteins are combined (orange).

4.14 References

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

1. Neto OA, Tassy O, Biancalana V, Zanoteli E, Pourquié O, Laporte J. Integrative Data Mining Highlights Candidate Genes for Monogenic Myopathies. *PLoS One.* 2014;9(10):e110888.

2. Tordjman M, Dabaj I, Laforet P, et al. Muscular MRI-based algorithm to differentiate inherited myopathies presenting with spinal rigidity. *Eur Radiol.* 2018;28(12):5293-5303.

3. Ravenscroft G, Davis MR, Lamont P, Forrest A, Laing NG. New era in genetics of early-onset muscle disease: Breakthroughs and challenges. *Semin Cell Dev Biol.* 2017;64:160-170.

4. Kazamel M, Milone M. Congenital myopathy with a novel SELN missense mutation and the challenge to differentiate it from congenital muscular dystrophy. *J Clin Neurosci.* 2019;62:238-239.

5. Goto I, Nagasaka S, Nagara H, Kuroiwa Y. Rigid spine syndrome. *J Neurol Neurosurg Psychiatry.* 1979;42(3):276-279.

6. Kayman-Kurekci G, Talim B, Korkusuz P, et al. Mutation in TOR1AIP1 encoding LAP1B in a form of muscular dystrophy: a novel gene related to nuclear envelopathies. *Neuromuscul Disord.* 2014;24(7):624-633.

7. Dai Y, Liang S, Huang Y, Chen L, Banerjee S. Targeted next generation sequencing identifies two novel mutations in SEPN1 in rigid spine muscular dystrophy 1. *Oncotarget.* 2016;7(51):83843-83849.

8. O'Grady GL, Best HA, Oates EC, et al. Recessive ACTA1 variant causes congenital muscular dystrophy with rigid spine. *Eur J Hum Genet.* 2015;23(6):883-886.

9. Moghadaszadeh B, Petit N, Jaillard C, et al. Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome. *Nat Genet.* 2001;29(1):17-18.

10. Foley AR, Zou Y, Dunford JE, et al. GGPS1 Mutations Cause Muscular Dystrophy/Hearing Loss/Ovarian Insufficiency Syndrome. *Ann Neurol* 2020;88(2):332-347.

11. Poewe W, Willeit H, Sluga E, Mayr U. The rigid spine syndrome--a myopathy of uncertain nosological position. *J Neurol Neurosurg Psychiatry.* 1985;48(9):887-893.

12. Arbogast S, Beuvin M, Fraysse B, Zhou H, Muntoni F, Ferreiro A. Oxidative stress in SEPN1-related myopathy: from pathophysiology to treatment. *Ann Neurol.* 2009;65(6):677-686.

13. Koul R, Sankhla D, Al-Jahdhami S, et al. Rigid Spine Syndrome among Children in Oman. *Sultan Qaboos Univ Med J.* 2015;15(3):e364-369.

14. Quijano-Roy S, Mbieleu B, Bönnemann CG, et al. De novo LMNA mutations cause a new form of congenital muscular dystrophy. *Ann Neurol.* 2008;64(2):177-186.

15. Scoto M, Cullup T, Cirak S, et al. Nebulin (NEB) mutations in a childhood onset distal myopathy with rods and cores uncovered by next generation sequencing. *Eur J Hum Genet.* 2013;21(11):1249-1252.

16. Bharucha-Goebel DX, Santi M, Medne L, et al. Severe congenital RYR1-associated myopathy: the expanding clinicopathologic and genetic spectrum. *Neurology.* 2013;80(17):1584-1589.

17. Clarke NF, Kolski H, Dye DE, et al. Mutations in TPM3 are a common cause of congenital fiber type disproportion. *Ann Neurol.* 2008;63(3):329-337.

18. Baker SK. Molecular clues into the pathogenesis of statin-mediated muscle toxicity. *Muscle Nerve.* 2005;31(5):572-580.

19. Ravenscroft G, Clayton JS, Faiz F, et al. Neurogenetic fetal akinesia and arthrogryposis: genetics, expanding genotype-phenotypes and functional genomics. *J Med Genet.* 2021;58(9):609-618.

20. Pais L, Snow H, Weisburd B, et al. seqr: a web-based analysis and collaboration tool for rare disease genomics. *medRxiv.* 2021:2021.2010.2027.21265326.

21. Cassandrini D, Trovato R, Rubegni A, et al. Congenital myopathies: clinical phenotypes and new diagnostic tools. *Ital J Pediatr.* 2017;43(1):101.

22. Gonorazky HD, Bönnemann CG, Dowling JJ. The genetics of congenital myopathies. *Handb Clin Neurol.* 2018;148:549-564.

23. North KN, Wang CH, Clarke N, et al. Approach to the diagnosis of congenital myopathies. *Neuromuscul Disord.* 2014;24(2):97-116.

24. Nykamp K, Anderson M, Powers M, et al. Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet Med.* 2017;19(10):1105-1117.

25. Lek M, MacArthur D. The Challenge of Next Generation Sequencing in the Context of Neuromuscular Diseases. *J Neuromuscul Dis.* 2014;1(2):135-149.

26. Shafqat N, Turnbull A, Zschocke J, Oppermann U, Yue WW. Crystal Structures of Human HMG-CoA Synthase Isoforms Provide Insights into Inherited Ketogenesis Disorders and Inhibitor Design. *J Mol Biol.* 2010;398(4):497-506.

27. Burley SK, Berman HM, Bhikadiya C, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* 2018;47(D1):D464-D474.

28. Wang J, Youkharibache P, Zhang D, et al. iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics.* 2020;36(1):131-135.

29. Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 2015;16(1):22.

30. Ravenscroft G, Miyatake S, Lehtokari V-L, et al. Mutations in KLHL40 Are a Frequent Cause of Severe Autosomal-Recessive Nemaline Myopathy. *Am J Hum Genet.* 2013;93(1):6-18.

31. McNamara EL, Taylor RL, Clayton JS, et al. Systemic AAV8-mediated delivery of a functional copy of muscle glycogen phosphorylase (Pygm) ameliorates disease in a murine model of McArdle disease. *Hum Mol Genet.* 2020;29(1):20-30.

32. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002;3(7):research0034.0031.

33. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9(7):671-675.

34. Stroud DA, Surgenor EE, Formosa LE, et al. Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature.* 2016;538(7623):123-126.

35. Goullée H, Taylor RL, Forrest ARR, Laing NG, Ravenscroft G, Clayton JS. Improved CRISPR/Cas9 gene editing in primary human myoblasts using low confluency cultures on Matrigel. *Skelet Muscle.* 2021;11(1).

36. Froger A, Hall JE. Transformation of plasmid DNA into E. coli using the heat shock method. *J Vis Exp.* 2007(6):253.

37. Haywood J, Schmidberger JW, James AM, et al. Structural basis of ribosomal peptide macrocyclization in plants. *eLife.* 2018;7:e32955.

38. Wilkins MR GE, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol.* 1999:531-552.

39. Lowe DM, Tubbs PK. 3-Hydroxy-3-methylglutaryl-coenzyme A synthase from ox liver. Properties of its acetyl derivative. *Biochem J.* 1985;227(2):601-607.

40. Soares da Costa TP, Yap MY, Perugini MA, et al. Dual roles of F123 in protein homodimerization and inhibitor binding to biotin protein ligase from Staphylococcus aureus. *Mol Microbiol.* 2014;91(1):110-120.

41. Peverelli MG, Soares da Costa TP, Kirby N, Perugini MA. Dimerization of Bacterial Diaminopimelate Decarboxylase Is Essential for Catalysis. *J Biol Chem.* 2016;291(18):9785-9795.

42. Christensen JB, Soares Da Costa TP, Faou P, Pearce FG, Panjikar S, Perugini MA. Structure and Function of Cyanobacterial DHDPS and DHDPR. *Sci Rep.* 2016;6(1):37111.

43. Yu J, Deng J, Guo X, et al. The GGC repeat expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy type 3. *Brain.* 2021;144(6):1819-1832.

44. Benarroch L, Bonne G, Rivier F, Hamroun D. The 2021 version of the gene table of neuromuscular disorders (nuclear genome). *Neuromuscul Disord.* 2020;30(12):1008-1048.

45. Fagerberg L, Hallström BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics.* 2014;13(2):397-406.

46. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585.

47. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.

48. Steussy CN, Robison AD, Tetrick AM, et al. A structural limitation on enzyme activity: the case of HMG-CoA synthase. *Biochemistry.* 2006;45(48):14407-14414.

49. Rokosz LL, Boulton DA, Butkiewicz EA, et al. Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase: expression, purification, and characterization of recombinant wild-type and Cys129 mutant enzymes. *Arch Biochem Biophys.* 1994;312(1):1-13.

50. Sapir A, Tsur A, Koorman T, et al. Controlled sumoylation of the mevalonate pathway enzyme HMGS-1 regulates metabolism during aging. *Proc Natl Acad Sci U S A.* 2014;111(37):E3880-3889.

51. Woollett LA. Where does fetal and embryonic cholesterol originate and what does it do? *Annu Rev Nutr.* 2008;28:97-114.

52. Bagheri-Fam S, Chen H, Wilson S, et al. The gene encoding the ketogenic enzyme HMGCS2 displays a unique expression during gonad development in mice. *PLoS One.* 2020;15(1):e0227411.

53. Dickinson ME, Flenniken AM, Ji X, et al. High-throughput discovery of novel developmental phenotypes. *Nature.* 2016;537(7621):508-514.

54. Perez-Garcia V, Fineberg E, Wilson R, et al. Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature.* 2018;555(7697):463-468.

55. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.

56. Lee T, Takami Y, Yamada K, et al. A Japanese case of mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase deficiency who presented with severe metabolic acidosis and fatty liver without hypoglycemia. *JIMD Rep.* 2019;48(1):19-25.

57. Thompson GN, Hsu BY, Pitt JJ, Treacy E, Stanley CA. Fasting hypoketotic coma in a child with deficiency of mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase. *N Engl J Med.* 1997;337(17):1203-1207.

58. Rojnueangnit K, Maneechai P, Thaweekul P, et al. Expanding phenotypic and mutational spectra of mitochondrial HMG-CoA synthase deficiency. *Eur J Med Genet.* 2020;63(12):104086.

59. Kılıç M, Dorum S, Topak A, Yazıcı MU, Ezgu FS, Coskun T. Expanding the clinical spectrum of mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase deficiency with Turkish cases harboring novel HMGCS2 gene mutations and literature review. *Am J Med Genet A.* 2020;182(7):1608-1614.

60. Mathews ES, Mawdsley DJ, Walker M, Hines JH, Pozzoli M, Appel B. Mutation of 3-hydroxy-3-methylglutaryl CoA synthase I reveals requirements for isoprenoid and cholesterol synthesis in oligodendrocyte migration arrest, axon wrapping, and myelin gene expression. *J Neurosci.* 2014;34(9):3402-3412.

61. Quintana AM, Hernandez JA, Gonzalez CG. Functional analysis of the zebrafish ortholog of HMGCS1 reveals independent functions for cholesterol

and isoprenoids in craniofacial development. *PLoS One.* 2017;12(7):e0180856.

62. Hernandez JA, Castro VL, Reyes-Nava N, Montes LP, Quintana AM. Mutations in the zebrafish hmgcs1 gene reveal a novel function for isoprenoids during red blood cell development. *Blood Adv.* 2019;3(8):1244-1254.

63. Houten SM, Frenkel J, Rijkers GT, Wanders RJ, Kuis W, Waterham HR. Temperature dependence of mutant mevalonate kinase activity as a pathogenic factor in hyper-IgD and periodic fever syndrome. *Hum Mol Genet.* 2002;11(25):3115-3124.

64. McLarren KW, Severson TM, du Souich C, et al. Hypomorphic temperature-sensitive alleles of NSDHL cause CK syndrome. *Am J Hum Genet.* 2010;87(6):905-914.

65. du Souich C, Chou A, Yin J, et al. Characterization of a new X-linked mental retardation syndrome with microcephaly, cortical malformation, and thin habitus. *Am J Med Genet A.* 2009;149A(11):2469-2478.

66. Cunningham D, Swartzlander D, Liyanarachchi S, Davuluri RV, Herman GE. Changes in gene expression associated with loss of function of the NSDHL sterol dehydrogenase in mouse embryonic fibroblasts. *J Lipid Res.* 2005;46(6):1150-1162.

67. Puisac B, Marcos-Alcalde I, Hernández-Marcos M, et al. Human Mitochondrial HMG-CoA Synthase Deficiency: Role of Enzyme Dimerization Surface and Characterization of Three New Patients. *Int J Mol Sci.* 2018;19(4):1010.

68. Vögeli B, Engilberge S, Girard E, et al. Archaeal acetoacetyl-CoA thiolase/HMG-CoA synthase complex channels the intermediate via a fused CoA-binding site. *Proc Natl Acad Sci U S A.* 2018;115(13):3380-3385.

69. Garg A, O'Rourke J, Long C, et al. KLHL40 deficiency destabilizes thin filament proteins and promotes nemaline myopathy. *J Clin Invest.* 2014;124(8):3529-3539.

70. Barboza-Cerda MC, Barboza-Quintana O, Martínez-Aldape G, Garza-Guajardo R, Déctor MA. Phenotypic severity in a family with MEND syndrome is directly associated with the accumulation of potentially functional variants of cholesterol homeostasis genes. *Mol Genet Genomic Med.* 2019;7(9):e931.

71. Witsch-Baumgartner M, Fitzky BU, Ogorelkova M, et al. Mutational spectrum in the Delta7-sterol reductase gene and genotype-phenotype correlation in 84 patients with Smith-Lemli-Opitz syndrome. *Am J Hum Genet.* 2000;66(2):402-412.

72. Grange DK, Kratz LE, Braverman NE, Kelley RI. CHILD syndrome caused by deficiency of 3beta-hydroxysteroid-delta8, delta7-isomerase. *Am J Med Genet.* 2000;90(4):328-335.

73. Mohassel P, Mammen AL. Anti-HMGCR Myopathy. *J Neuromuscul Dis.* 2018;5(1):11-20.

74. Mammen AL, Chung T, Christopher-Stine L, et al. Autoantibodies against 3-hydroxy-3-methylglutaryl-coenzyme A reductase in patients with statin-associated autoimmune myopathy. *Arthritis Rheum.* 2011;63(3):713-721.

75. Kishi T, Rider LG, Pak K, et al. Association of Anti-3-Hydroxy-3-Methylglutaryl-Coenzyme A Reductase Autoantibodies With DRB1*07:01 and Severe Myositis in Juvenile Myositis Patients. *Arthritis Care Res (Hoboken).* 2017;69(7):1088-1094.

76. Porter FD, Herman GE. Malformation syndromes caused by disorders of cholesterol synthesis. *J Lipid Res.* 2011;52(1):6-34.

77. Peciuliene S, Burnyte B, Gudaitiene R, et al. Perinatal manifestation of mevalonate kinase deficiency and efficacy of anakinra. *Pediatr Rheumatol Online J.* 2016;14(1):19.

78. Gibson KM, Hoffmann G, Nyhan WL, et al. Mevalonate kinase deficiency in a child with cerebellar ataxia, hypotonia and mevalonic aciduria. *European Journal of Pediatrics.* 1988;148(3):250-252.

79. Thompson PD, Panza G, Zaleski A, Taylor B. Statin-Associated Side Effects. *J Am Coll Cardiol.* 2016;67(20):2395-2410.

80. Tiniakou E, Pinal-Fernandez I, Lloyd TE, et al. More severe disease and slower recovery in younger patients with anti-3-hydroxy-3-methylglutaryl-coenzyme A reductase-associated autoimmune myopathy. *Rheumatology (Oxford).* 2017;56(5):787-794.

81. Ghaoui R, Cooper ST, Lek M, et al. Use of Whole-Exome Sequencing for Diagnosis of Limb-Girdle Muscular Dystrophy: Outcomes and Lessons Learned. *JAMA Neurol.* 2015;72(12):1424-1432.

82. Badayan I, Cudkowicz ME. Profound Muscle Weakness and Pain after One Dose of Actonel. *Case Reports in Medicine.* 2009;2009:693014.

83. DeMonaco HJ. Patient- and Physician-Oriented Web Sites and Drug Surveillance: Bisphosphonates and Severe Bone, Joint, and Muscle Pain. *Arch Intern Med.* 2009;169(12):1164-1166.

84. Westwood FR, Scott RC, Marsden AM, Bigley A, Randall K. Rosuvastatin: characterization of induced myopathy in the rat. *Toxicol Pathol.* 2008;36(2):345-352.

85. Yablonka-Reuveni Z. The skeletal muscle satellite cell: still young and fascinating at 50. *J Histochem Cytochem.* 2011;59(12):1041-1059.

86. Feng X, Naz F, Juan AH, Dell'Orso S, Sartorelli V. Identification of Skeletal Muscle Satellite Cells by Immunofluorescence with Pax7 and Laminin Antibodies. *J Vis Exp.* 2018(134).

87. Grunwald SA, Popp O, Haafke S, et al. Statin-induced myopathic changes in primary human muscle cells and reversal by a prostaglandin F2 alpha analogue. *Sci Rep.* 2020;10(1):2158.

88. Trapani L, Segatto M, La Rosa P, et al. 3-hydroxy 3-methylglutaryl coenzyme a reductase inhibition impairs muscle regeneration. *J Cell Biochem.* 2012;113(6):2057-2063.

89. Ahmadi Y, Ghorbanihaghjo A, Naghi-Zadeh M, Yagin NL. Oxidative stress as a possible mechanism of statin-induced myopathy. *Inflammopharmacology.* 2018;26(3):667-674.

90. Huang J, Du J, Lin W, et al. Regulation of lactate production through p53/β-enolase axis contributes to statin-associated muscle symptoms. *EBioMedicine.* 2019;45:251-260.

91. Clizbe DB, Owens ML, Masuda KR, Shackelford JE, Krisans SK. IDI2, a second isopentenyl diphosphate isomerase in mammals. *J Biol Chem.* 2007;282(9):6668-6676.

92. Moosmann B, Behl C. Selenoproteins, Cholesterol-Lowering Drugs, and the Consequences Revisiting of the Mevalonate Pathway. *Trends Cardiovasc Med.* 2004;14(7):273-281.

93. Kromer A, Moosmann B. Statin-Induced Liver Injury Involves Cross-Talk between Cholesterol and Selenoprotein Biosynthetic Pathways. *Mol Pharmacol.* 2009;75(6):1421-1429.

94. Sirvent P, Mercier J, Vassort G, Lacampagne A. Simvastatin triggers mitochondria-induced Ca2= signaling alteration in skeletal muscle. *Biochem Biophys Res Commun.* 2005;329:1067-1075.

95. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J.* 2000;78(3):1606-1619.

96. Schuck P, Perugini MA, Gonzales NR, Howlett GJ, Schubert D. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys J.* 2002;82(2):1096-1111.

Chapter 5

General discussion

5.1 Summary of findings

This thesis aimed to investigate genetic causes of Mendelian congenital myopathies in patients with unresolved genetic aetiology. The approach involved two overall steps. Initially, whole exome sequencing (WES) analysis was performed on a cohort of eight families clinically diagnosed with congenital myopathy. Subsequently, functional genomics was performed to further explore and validate selected candidate pathogenic variants in two of these families.

Two key candidate variants were prioritised by WES analysis and were the focus of this thesis: a likely pathogenic variant (c.*152G>T) in a known congenital myopathy gene (*KLHL40*) that was inherited *in trans* with a *KLHL40* multi-exon deletion, and a homozygous missense variant (c.1339T>C, p.(Ser447Pro)) in *HMGCS1*, a novel gene candidate for rigid spine syndrome.

5.2 Factors facilitating novel disease gene and variant discovery in this thesis

5.2.1 Comprehensive gene panel pre-screening

Prior to conducting WES, the majority of families recruited in this study underwent comprehensive gene panel testing. This strategy aimed to rapidly identify a diagnosis for patients with disorders associated with known disease genes, and prioritise panel-negative families for further investigation[1]. The families who remained panel-negative were assumed to have disorders associated with novel disease variants or genes and were subsequently recruited for WES analysis. This helped enrich the resulting cohort for the discovery of novel disease variants and genes[1].

5.2.2 Availability of familial WES

The availability of familial genetic data can be crucial for the identification and prioritisation of novel variants and disease gene candidates[2]. In Chapter 4, I showed how WES analysis of four family members (i.e., quad analysis) facilitated the identification and support of *HMGCS1* as a candidate disease gene. Given the suspected recessive nature of disease in this family, the availability of WES data from both affected siblings and the unaffected parents enabled filtration for rare coding and essential splice variants (<0.01 allele frequency) that co-segregated with disease[3]. This

facilitated the capturing of the private *HMGCS1* c.1339T>C variant, the only candidate remaining in the filtered WES data. Of note, while evidence of variant co-segregation is inadequate to confirm a diagnosis alone[2], it provided an encouraging starting point for further characterisation of this candidate. Overall, this study supports that analysis of familial WES can expedite the identification and prioritisation of novel disease gene candidates and help reduce the diagnostic odyssey for patients and families[4].

5.2.3 Availability of patient clinical data and muscle biopsy for deep phenotyping

Another factor facilitating the prioritisation of *KLHL40* and *HMGCS1* as causative disease genes was detailed clinical information and deep phenotyping via muscle pathology and muscle MRI results[5]. As demonstrated in Chapter 3, having *a priori* knowledge of the clinical diagnosis of nemaline myopathy enabled a targeted analysis of non-coding variants in known nemaline myopathy genes (Chapter 1, Table 1.1). This facilitated the identification of the 3' UTR variant in *KLHL40*[6]. Deep phenotyping was also informative for the analysis of the rigid spine syndrome cohort in Chapter 4 and helped attribute *HMGCS1* as a novel genetic cause for their disorder[2]. Particularly, the distinct phenotype shared by all probands with biallelic *HMGCS1* variants, including spinal rigidity, scoliosis, respiratory insufficiency, and age of onset increased support for the *HMGCS1* variants as the cause for their disease. Overall, both studies in this thesis reflect the importance of comprehensive phenotyping during genetic investigations of congenital myopathies[7].

5.2.4 Data sharing and collaboration

The identification of additional individuals with matching phenotypes and variants in the same gene provides strong support for disease causality[2,5,8,9]. In Chapter 4, sharing of the *HMGCS1* data and clinical phenotypes of the siblings with global clinical networks and matchmaker exchange[8] enabled the identification of two additional unrelated patients with genetically unresolved rigid spine myopathy and biallelic variants in *HMGCS1*. A potential fifth patient was also identified by data sharing, with a monoallelic variant and the same rigid spine phenotype. Research is ongoing to identify a potential second variant in *HMGCS1* in this individual. Overall, this chapter highlights the potential of data sharing in accelerating matchmaking and contributing to improved variant interpretation and validation of novel disease genes[4].

Additionally, in Chapter 3, sharing of the *KLHL40* c.*152G>T variant with 3' UTR experts and bioinformaticians was crucial for elucidating the mechanism of disease in the patient with nemaline myopathy. The implications of the c.*152G>T variant on splicing were identified by bioinformatic collaborators using recently-developed tools Introme[10] and MINTIE[11]. In addition, discovery of the splice-activating nonsense mediated decay mechanism was facilitated by discussions with a 3' UTR expert (Traude Beilharz, personal communication). This highlights the benefits of collaborations and data sharing during genetic investigations of rare diseases[12].

5.2.5 Insights from animal models

A valuable approach to evaluating the pathogenicity of candidate genes is the analysis of available animal model data[6]. In Chapter 4, access to phenotypic data from previous *HMGCS1* animal studies enabled further characterisation and confirmation of *HMGCS1* intolerance to variation. According to the International Mouse Phenotyping Consortium[13], homozygous *Hmgcs1* knockout mice showed complete preweaning lethality, suggesting *HMGCS1* is critical for life[13,14]. Further, data from the Zebrafish Information Network indicated mutant zebrafish with a homozygous point mutation in *hmgcs1* had a plethora of congenital disorders[15-18]. Although a muscle phenotype was not reported from these studies, it is known that animal models can show phenotypic variability and fail to recapitulate the specific phenotype observed in affected humans[15,19,20]. Nonetheless, the availability of animal experimental data contributed to the evaluation of the pathogenesis of *HMGCS1* as a novel disease gene[21].

5.2.6 Accessibility of large-scale multi-omic data for novel disease gene evaluation

Other factors facilitating evaluation of disease gene candidates in this study included analysis of available online resources including large-scale normal population genetic databases, transcriptomic databases and proteomic databases[6].

The gnomAD population genetic database was particularly useful in Chapter 4 for the evaluation of *HMGCS1* intolerance to missense and loss of function variation[22]. The lack of homozygous *HMGCS1* missense variants and relatively high constraint scores supported that biallelic missense variants in this gene can be deleterious. Of note, current gene constraint parameters in gnomAD are solely suited to assessing gene intolerance to synonymous, missense and loss of function variants[22]. However, gene

constraint predictions for variants in non-coding regions such as promoters, enhancers and UTRs have not yet been reported in gnomAD. It has been demonstrated that analysis of non-coding variation from large-scale population data can reveal genes susceptible to pathogenic non-coding variation and help define particular non-coding regions of high constraint[23]. Therefore, it may be expected that inclusion of constraint scores for non-coding variants in gnomAD may aid in the identification of pathogenic variants such as the *KLHL40* 3' UTR variant described in Chapter 3.

Transcriptomic and proteomic data from freely accessible datasets such as GTEx[24], FANTOM5[25], NCBI and the human protein atlas[26] were also useful resources in exploring candidate gene dysfunction in the context of muscle disease. In my analyses of *HMGCS1*, I did not set a threshold for gene expression in skeletal muscle. The rationale was that an increasing numbers of novel disease genes, including *GGPS1* (another enzyme involved in the same pathway as *HMGCS1*) are detected, but are not necessarily enriched in skeletal muscle despite causing a muscle phenotype[27]. Therefore, the detection (but not enrichment) of *HMGCS1* in skeletal muscle nevertheless supported the hypothesis that HMGCS1 may be required for muscle function[28,29]. Overall, as a number of congenital myopathy genes are identified with ubiquitous expression, this study supports that gene candidates should not be solely prioritised based on gene enrichment in the tissue of interest.

A combined analysis of the available resources, including those described above can enhance NGS variant interpretation and understanding of genotype-phenotype associations[23,26,30,31]. Using data mining, gene clustering, and network analyses, Neto *et al.*[31] devised a list of proposed novel genes that were predicted, and some now demonstrated to be associated with myopathy[31-33]. Remarkably, from their predictions, Neto *et al*.[31] listed *HMGCS1* as a potential novel candidate disease gene that may be implicated in congenital muscular dystrophy and/or congenital myopathy. Together these data highlight the power of integrated and multi-scale analyses in accelerating the discovery of novel disease genes.

5.3 Complementary diagnostic tools and functional genomics

In this thesis, a significant component of the work involved functionally characterising the key candidates identified from WES. As will be discussed below, Chapters 3 and

4 particularly highlight RNA-seq, western blotting, quantitative mass spectrometry and *in vitro* assays as powerful approaches that can clarify the pathogenicity of genes and variants of unknown significance[2].

5.3.1 RNA-seq

In Chapter 3, RNA-seq is highlighted as a powerful complementary tool for the identification of novel pathogenic variants and mechanisms in regulatory regions of known disease genes[34]. During analyses of the 3' UTR variant in *KLHL40*, the detection of RNA-seq reads supporting splicing at the variant site suggested the c.*152G>T variant may be pathogenic and reduce transcript stability. Moreover, the pronounced reduction of *KLHL40* RNA-seq coverage in the patient sample further supported the diagnosis of a KLHL40-related myopathy. Of note, while RNA-seq data can identify aberrant splicing, it is suggested that this evidence alone may be insufficient in informing clinical significance[35,36]. Thus, additional data, for example at the protein level or *in vitro* assays may be required to provide further evidence of variant pathogenecity[36,37].

5.3.2 Western blotting

In both chapters, western blotting is highlighted as a useful approach in confirming the effect of candidate variants on protein abundance (where validated antibodies are available). Western blotting for KLHL40 in patient muscle biopsy confirmed that KLHL40 deficiency was the likely cause of the patient's nemaline myopathy. In contrast, the detection of HMGCS1 by western blotting of patient skeletal muscle indicated that the studied genetic substitution likely does not affect protein stability or abundance, as was supported by my structural studies on recombinant protein. Given that *Hmgcs1* knockdown in mice is lethal, this suggested that variants in the patients would need to retain sufficient function to prevent embryonic lethality. Thus, in this instance, western blotting was useful for investigating the molecular aetiology of HMGCS1, suggesting that HMGCS1 substitutions likely have more subtle functional effects. Despite the utility of western blotting, several disadvantages accompany this approach[38], as will be discussed in Section 5.5.3.

5.3.3 Quantitative mass spectrometry

In Chapter 4, quantitative mass spectrometry was a useful approach for validating the HMGCS1 antibody and supporting the western blotting data relating to HMGCS1

abundance in myoblasts and myotubes. The consistent detection of HMGCS1 in myoblasts and myotubes by both approaches suggested that HMGCS1 may be required for skeletal muscle function and myogenesis[29]. Aside from validation of antibodies, quantitative mass spectrometry can offer other advantages[39] including A) unbiased analysis of protein abundance in e.g. patient tissue compared to controls[40], B) simultaneous quantitation of multiple proteins enabling proteomic investigations[39] and C) distinguished detection of protein isoforms and paralogues[41,42]. Additionally, this technique may be useful for investigating novel candidate proteins that do not have a suitable or validated antibody[43]. Although there are limited reports of quantitative mass spectrometry in the context of congenital myopathy research, it is expected that there will be a shift to this method as it becomes cheaper and more accessible[39]. Overall, Chapter 4 highlights quantitative mass spectrometry as a useful complementary approach for investigating the muscle proteome and elucidating pathomechanisms of skeletal muscle diseases[42,44].

5.3.4 *In vitro* assays

In vitro expression assays can also aid in clarifying the roles of genes and variants of unknown significance in disease[2,45]. For example, in Chapter 3, *in vitro* assays (adapted from Dusl *et al.*[45]) proved useful to clarify the impacts of the c.*152G>T variant on splicing and *KLHL40* expression. These assays showed that it is specifically 3' UTR splicing that leads to *KLHL40* deficiency via induction of nonsense mediated decay[46]. While in most cases patient-derived cells or tissue are sufficient to confirm variant pathogenicity, this study highlights how *in vitro* analyses provide a controlled system that can unravel pathomechanisms specific to a variant of interest that may otherwise be more complicated to deduce from patient tissue or cells alone[36].

5.4 Limitations and challenges of NGS analysis

5.4.1 Uncapturable variants

In this thesis, WES analysis identified candidate disease variants in three of eight families (data not shown for remaining candidates). This is a typical yield for NGS investigations in congenital myopathy cohorts (<60%)[47], which are generally higher than other neuromuscular disorders[1,47]. There are several reasons that may have limited the identification of candidate variants in the remaining families. One bottleneck includes the identification too many preliminary candidates to follow up for functional

characterisation under the available resources and duration of the study[48]. A second limitation may relate to the limitations of WES in capturing pathogenic variants in regions that are "dark" by sequencing[49]. The selective capturing of variation in coding regions may have precluded identification of some non-coding variants such as deep intronic variants[50]. In such cases, WGS may be an informative future approach for analysis[50]. In addition, given that NGS methods rely on short-read sequencing technologies, this may have precluded detection of pathogenic variants that are difficult to sequence and map with short-read NGS, such as regions that are not unique (e.g. pseudogenes), repetitive regions (including STRs), and complex structural variation[34,51,52]. It is expected that long-read sequencing may help uncover such variation[51], as discussed in Chapter 1.

5.4.2 Variants with conflicting *in silico* pathogenicity predictions

While *in silico* analyses are generally valuable tools for evaluating variant pathogenicity[6], in some cases, these tools can provide conflicting and inaccurate pathogenicity predictions[53,54]. In both Chapters 3 and 4, most tools suggested the *KLHL40* and *HMGCS1* variants would be tolerated, despite the variants showing apparent functional differences in the *in vitro* assays. The false negative predictions in these two cases may be because *in silico* tools are generally tailored for evaluating loss-of-function variants and missense variants in well-annotated regions[55]. Consequently, pathogenic variants in unannotated regions may be misjudged[55-57]. This may explain why for example, the *KLHL40* c.*152G>T donor splice gain variant in the 3' UTR was associated with low SpliceAI predictions. It is expected that as more functional splice variants are identified in such poorly understood regions, the algorithms for *in silico* tools can be better tailored for more accurate variant predictions. Interestingly, the algorithms for Introme[10], a relatively new splice prediction tool enabled strong detection of the *KLHL40* 3' UTR variant as a donor splice site. Therefore, Introme may be a useful tool for predicting the effect of other single nucleotide 3' UTR variants on splicing[10]. Overall, these analyses suggest that *in silico* outputs should be interpreted with context of the variant's location and assessed alongside multiple suitable *in silico* pathogenicity prediction tools[6].

The molecular aetiology of variants should also be considered during the interpretation of *in silico* variant pathogenicity predictions[2]. In Chapter 4, both structure-based *in*

silico and circular dichroism analyses suggested the p.S447P substitution in HMGCS1 does not affect protein structure. Of note, such results may not necessarily refute the pathogenicity of the p.S447P substitution. Rather, this could suggest that the substitution acts by an alternative pathomechanism, for example involving protein-protein interactions[58] or post translational modifications, e.g. sumoylation[14], which are functions that have been described in HMGCS of other species[14,58].

5.5 Limitations and challenges of functional genomics

There are vital factors to consider in the implementation and interpretation of functional genomics assays. As described in Chapter 1, the ACMG-AMP and VCEP guidelines emphasise the importance of assessing assay validity, reproducibility, and biological relevance[2,36]. In addition, the clinical (diagnostic) significance of the assays must be evaluated to confirm variant pathogenicity[59]. While current guidelines provide attributes for "well-established" functional assays, the methods for how to assess these attributes are not well described[36]. Moreover, such guidelines are not suitable for investigations of novel disease genes and proteins with poorly understood aetiologies[2]. The limited guidelines for functionally evaluating the novel candidates in this thesis resulted in some limitations and challenges as will be elaborated below.

5.5.1 Evaluating the biological relevance of a HEK293FT expression system for *in vitro KLHL40* analyses.

In Chapter 3, the *in vitro* assays for *KLHL40* involved use of a HEK293FT expression system. Given the patient had a muscle disease, a major consideration was whether the use of HEK293FT as a system was biologically reflective and capable of capturing the molecular pathomechanism occurring in muscle. For example, defective muscle-specific trans-acting binding proteins and microRNA interactions, if they were involved in the pathogenesis of the c.*152G>T variant[45,60]. The choice for a HEK293FT expression system was based on two factors: 1) the c.*152G>T variant was predicted to affect splicing and HEK293FT cells have been previously successfully used in minigene assays to investigate splice variants in myopathy-associated genes[61,62], and 2) skeletal muscle cell lines are notorious for their poor transfection efficiencies[63]. Thus, HEK293FT cells were used for the study given they were more likely to provide robust results.

5.5.2 Evaluating the clinical significance of the *in vitro* HMGCS1 assays

One major challenge in the functional characterisation of recombinant HMGCS1 was evaluating the clinical significance of the data. Partial differences were seen in the thermostability and activity of $HMGCS1^{WT}$ and $HMGCS1^{S447P}$. While statistically significant, it was difficult to determine whether such changes were biologically and clinically relevant and thus pathogenic. In particular, the assays were adapted from nonclinical experiments which involved expression of recombinant His-tagged HMGCS1 in a bacterial system[64,65]. Given that eukaryotic HMGCS1 is reported to undergo multifaceted post-translational modifications[14], consideration should be given as to whether the structure, activity and interactions of $HMGCS1^{WT}$ and $HMGCS1^{S447P}$ are biologically representative[64,66]. It is noteworthy however, that clinical studies of HMGCS2 mutants were also performed using bacterial expression systems[67]. Nevertheless, further studies for example using a mammalian expression system may be used to confirm the findings observed in the HMGCS1 assays using bacterially-expressed proteins[36,59].

According to VCEP guidelines and most enzymatic studies in the literature, the majority of enzymatic defects described in rare diseases are typically associated with markedly reduced or diminished activity[36]. Using the VCEP guidelines for interpretation, the differences between $HMGCS1^{WT}$ and $HMGCS1^{S447P}$ activity were below the typical pathogenic threshold (activity retained <50%)[36] suggesting that the effects of the S447P substitution on HMGCS1 activity are insignificant. However, it is apparent that several enzymes including HMGCS2 and GGPS1, which sit in the same enzymatic pathway, have had pathogenic substitutions identified that only partially disrupt enzymatic activity but yet still lead to disease[27,67]. For example, as discussed in Chapter 4, a p.R505Q substitution in HMGCS2 was shown to retain 70% of activity in a patient presenting with HMGCS2 deficiency[67]. This substitution was previously described in a compound heterozygous state in another patient with HMGCS2 deficiency, thus supporting it is pathogenic despite causing partial reduction on enzymatic function. Such examples suggest that current guidelines may not fit for all enzymes and may thus lead to inaccurate interpretation of functional data[36]. As previously discussed, prenatal lethality is associated with mice that are null for most genes of the mevalonate pathway[13]. Given how critical such genes appear to be for life, it would therefore be expected that the effects of pathogenic variants identified in

living patients would be partial and not complete loss of enzymatic in order to sustain life. Therefore, improvements to the guidelines and interpretation of enzymatic assays are needed[36].

5.5.3 Technical difficulties associated with assays

Several technical difficulties were encountered during the functional analyses of HMGCS1. In particular, these relate to the reproducibility and validation of the enzymatic assays. Variability in HMGCS1 activity was apparent at various substrate concentrations, most notably at 100 µM of either substrate (Chapter 4, Figure 4.6). Variability was also apparent between runs performed on different days. This reduced the overall fit of the Michaelis-Menten plots and thus affected the accuracy of the K_m and V_{max} derived for wildtype and mutant HMGCS1. In addition, the apparent K_m determined for $HMGCS1^{WT}$ in the analyses (5.433 µM) was markedly different to the previous report by Rokosz *et al.*[64] (29 ± 7 µM). Such variation may be due to differences in the experimental design and approach, as has been suggested in studies of other recombinant enzymes such as PAH (phenylalanine hydroxylase), implicated in PKU (Phenylketonuria)[36]. In particular, enzymatic analyses of milder PAH variants were associated with variable results. In some instances, activity of several variants were ≥50% of the wild type, which exceeds the VCEP threshold and incorrectly "disproves" variant pathogenicity[36]. As such, further guidance is required for standardising *in vitro* enzyme assay approaches and determining how to resolve variable and inconsistent results[36].

Another technical challenge was obtaining a valid HMGCS1 antibody for western blotting and immunofluorescence. Three different HMGCS1 antibodies had to be tested to establish a suitable and specific antibody for western blotting (based on detection of recombinant HMGCS1 protein and mass spectrometry analysis). However, even the validated antibody showed potentially nonspecific binding in skeletal muscle sections by immunofluorescence. While some signal was lost when the antibody was preincubated with HMGCS1 protein, some signals remained present which raises the possibility of non-specific staining and complicates conclusions made about HMGCS1 localisation. The inconsistencies and recognised issues of antibody non-specificity highlights the importance of antibody validation despite the expense, time and efforts required[38].

5.6 The remaining families

Though beyond the scope of this thesis, the future research pathway for the patients in the congenital myopathy cohort that remained without a definitive candidate following NGS analysis needs to be considered. The major issues preventing identification and further analysis of candidates in these families are briefly addressed below.

All the remaining families were analysed by trio WES which enabled filtering for variants that segregated with disease. However, numerous candidate variants were still identified in each family following stringent filtering. Many of these were VUS in known muscle disease genes such as *RYR1* and *TTN*. Such genes naturally harbour many polymorphisms given their large size and thus can make variant interpretation challenging[68]. The VUS identified in this study thus had insufficient evidence to warrant further pursuit[69].

In addition, patients from highly consanguineous families had multiple candidate homozygous variants. This made it difficult to ascertain a single likely causative genetic candidate without additional information about variant segregation in the extended family. While trio WES analysis in consanguineous populations has enabled discovery of numerous novel disease genes and variants such as *de novo* and recessive disorders[4,7], some researchers argue that the diagnostic efficiency in families with very high degrees of consanguinity may be hindered due to the higher degree of homozygousity[70]. Nevertheless, it appears that WES findings in inbred populations are generally higher than outbred populations[4,7].

It is suspected that novel disease genes may be an underlying cause of disease for the remaining families. However, these genes may remain bottlenecked in NGS pipelines under the current variant and gene classification guidelines. Moreover, the abundance of GUS, as well as limitations in time and resources may have precluded further analysis of these candidates[48]. From the lessons learned in this thesis and in the literature, improving WES analysis for the remaining families may require accessibility to DNA from extended family members where relevant, additional patient clinical history and patient-derived samples where possible[34]. Techniques including RNA-seq and proteomics may be performed to complement WES data, followed by WGS if no variants are uncovered. However, such complementary approaches may

not be possible due to issues with consent, invasive biopsy procedures and in some cases, unavailability of samples due to death of family members or patients[34,71].

5.7 Significance of thesis

This thesis contributes two novel pieces of knowledge to the genetics of congenital myopathies and provides insight into molecular pathways that are uncommonly explored[27]. Reporting of these candidates should facilitate a genetic diagnosis for other affected families worldwide.

I describe the identification of a novel 3' UTR splice variant with a nonsense mediated decay mechanism that appears to be underrepresented in the literature[72]. Discovery of the splice activating nonsense mediated decay mechanism may provide future insights into potential gene therapies such as antisense oligonucleotide (ASO) therapy, which has been used to restore defective splicing in disease contexts[73]. ASO therapy is also promising given some reports demonstrate that ASO binding to 3' UTRs can stabilise mRNA[74,75].

Additionally, I describe the discovery of *HMGCS1* as a novel disease gene candidate for rigid spine syndrome. These findings expand the number of genes implicated in this disorder and highlight the mevalonate pathway as being crucial for muscle function[27]. Reporting of *HMGCS1* as a disease gene may lead to further studies that clarify links between the mevalonate pathway and neuromuscular disease, which to date are poorly understood[76,77]. Moreover, future studies of this pathway may also provide additional insight into the muscle-damaging side effects of statins[77].

5.8 Future directions

I speculate that hidden pathogenic 3' UTR variants may be implicated in patients with rare genetically unsolved diseases. Thus, analysis of 3' UTR variants in cohorts of patients with unsolved genetic disorders may identify the cause of disease for many patients. Disease gene databases such as ClinVar may also provide a reservoir of 3' UTR VUS that may cause 3' UTR-related nonsense mediated decay. The use of *in vitro* expression systems such as the one utilised in this study may aid in characterisation and validation of additional pathogenic 3' UTR splice variants

identified. Finally, research involving targeted gene therapy for 3' UTR splice variants is warranted.

As the molecular aetiology for *HMGCS1*-myopthy remains undetermined, future research includes phenotyping zebrafish *hmgcs1* knockout models and performing rescue assays with a focus on muscular phenotypes[78-80]. Ongoing research also includes *in vitro* characterisation of the other HMGCS1 substitutions reported in this study, which will add further confidence to the observations for the HMGCS1 p.Ser447Pro substitution. Additionally, testing *HMGCS1* in unsolved rigid spine syndrome cohorts may provide a diagnosis for more families and further elucidate the mechanisms of *HMGCS1* disease.

Given the crucial roles of the mevalonate pathway, it might be expected that by nature, additional inborn disorders will be associated with this pathway[13,55]. Therefore, genes of the mevalonate pathway should be considered during NGS analysis of patients with unresolved myopathic phenotypes.

The families remaining without a genetic diagnosis warrant further investigations. Approaches including whole genome sequencing, long read sequencing and RNA-seq analyses can be performed to search for the missing genetics of these families[47]. Identification of novel disease genes or variants will provide further insights into pathological mechanisms of myopathies and can define novel therapeutic targets[31].

5.9 Thesis conclusion

A precise genetic diagnosis is the crux in rare diseases such as congenital myopathies[69]. However, identifying and confirming a genetic diagnosis can be challenging[48]. In this thesis, I show that combined NGS and functional genomics methods can facilitate the investigation and evaluation of novel genetic causes of congenital myopathies in both known disease genes and novel disease candidates. The shift from a pure genetic approach to the implementation of well-established functional genomic and multi-omics approaches are expected to increase the diagnostic yield for NGS[47,81]. Moreover, data sharing, and increased collaborations will be fundamental to reducing the diagnostic odyssey, clarifying disease mechanisms, and developing treatments for congenital myopathies[5,47,55,82].

5.10 References

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

1. Beecroft SJ, Yau KS, Allcock RJN, et al. Targeted gene panel use in 2249 neuromuscular patients: the Australasian referral center experience. *Ann Clin Transl Neurol.* 2020;7(3):353-362.

2. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-423.

3. Ghaoui R, Cooper ST, Lek M, et al. Use of Whole-Exome Sequencing for Diagnosis of Limb-Girdle Muscular Dystrophy: Outcomes and Lessons Learned. *JAMA Neurol.* 2015;72(12):1424-1432.

4. Monies D, Abouelhoda M, AlSayed M, et al. The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. *Hum Genet.* 2017;136(8):921-939.

5. Sobreira NL, Valle D. Lessons learned from the search for genes responsible for rare Mendelian disorders. *Mol Genet Genomic Med.* 2016;4(4):371-375.

6. Dashti MJS, Gamieldien J. A practical guide to filtering and prioritizing genetic variants. *BioTechniques.* 2017;62(1):18-30.

7. Beaulieu CL, Majewski J, Schwartzentruber J, et al. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *Am J Hum Genet.* 2014;94(6):809-817.

8. Sobreira NLM, Arachchi H, Buske OJ, et al. Matchmaker Exchange. *Curr Protoc Hum Genet.* 2017;95:9.31.31-39.31.15.

9. Strande NT, Riggs ER, Buchanan AH, et al. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am J Hum Genet.* 2017;100(6):895-906.

10. Sullivan P, Mayoh C, Wong-Erasmus M, et al. NEW GENES AND DISEASES / NGS & RELATED TECHNIQUES: P.334 Introme identifies non-canonical splice-altering variants in neuromuscular patients resulting in multiple new genetic diagnoses. *Neuromuscul Disord.* 2020;30:S144.

11. Cmero M, Schmidt B, Majewski IJ, Ekert PG, Oshlack A, Davidson NM. MINTIE: identifying novel structural and splice variants in transcriptomes using RNA-seq data. *Genome Biol.* 2021;22(1):296.

12. Boycott KM, Campeau PM, Howley HE, et al. The Canadian Rare Diseases Models and Mechanisms (RDMM) Network: Connecting Understudied Genes to Model Organisms. *Am J Hum Genet.* 2020;106(2):143-152.

13. Dickinson ME, Flenniken AM, Ji X, et al. High-throughput discovery of novel developmental phenotypes. *Nature.* 2016;537(7621):508-514.

14. Sapir A, Tsur A, Koorman T, et al. Controlled sumoylation of the mevalonate pathway enzyme HMGS-1 regulates metabolism during aging. *Proc Natl Acad Sci U S A.* 2014;111(37):E3880-3889.

15. Varga M, Ralbovszki D, Balogh E, Hamar R, Keszthelyi M, Tory K. Zebrafish Models of Rare Hereditary Pediatric Diseases. *Diseases.* 2018;6(2).

16. Mathews ES, Appel B. Cholesterol Biosynthesis Supports Myelin Gene Expression and Axon Ensheathment through Modulation of P13K/Akt/mTor Signaling. *J Neurosci.* 2016;36(29):7628-7639.

17. Quintana AM, Hernandez JA, Gonzalez CG. Functional analysis of the zebrafish ortholog of HMGCS1 reveals independent functions for cholesterol and isoprenoids in craniofacial development. *PLoS One.* 2017;12(7):e0180856.

18. Hernandez JA, Castro VL, Reyes-Nava N, Montes LP, Quintana AM. Mutations in the zebrafish hmgcs1 gene reveal a novel function for isoprenoids during red blood cell development. *Blood Adv.* 2019;3(8):1244-1254.

19. Lek M, MacArthur D. The Challenge of Next Generation Sequencing in the Context of Neuromuscular Diseases. *J Neuromuscul Dis.* 2014;1(2):135-149.

20. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature.* 2014;508(7497):469-476.

21. Vainzof M, Ayub-Guerrieri D, Onofre PCG, et al. Animal Models for Genetic Neuromuscular Diseases. *J Mol Neurosci.* 2008;34(3):241-248.

22. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.

23. Petrovski S, Gussow AB, Wang Q, et al. The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. *PLoS Genet.* 2015;11(9):e1005492.

24. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585.

25. Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 2015;16(1):22.

26. Fagerberg L, Hallström BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics.* 2014;13(2):397-406.

27. Foley AR, Zou Y, Dunford JE, et al. GGPS1 Mutations Cause Muscular Dystrophy/Hearing Loss/Ovarian Insufficiency Syndrome. *Ann Neurol* 2020;88(2):332-347.

28. Kitsak M, Sharma A, Menche J, et al. Tissue Specificity of Human Disease Module. *Sci Rep.* 2016;6:35241.

29. Trapani L, Segatto M, La Rosa P, et al. 3-hydroxy 3-methylglutaryl coenzyme a reductase inhibition impairs muscle regeneration. *J Cell Biochem.* 2012;113(6):2057-2063.

30. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci U S A.* 2007;104(21):8685-8690.

31. Neto OA, Tassy O, Biancalana V, Zanoteli E, Pourquié O, Laporte J. Integrative Data Mining Highlights Candidate Genes for Monogenic Myopathies. *PLoS One.* 2014;9(10):e110888.

32. Ravenscroft G, Zaharieva IT, Bortolotti CA, et al. Bi-allelic mutations in MYL1 cause a severe congenital myopathy. *Hum Mol Genet.* 2018;27(24):4263-4272.

33. Nilipour Y, Nafissi S, Tjust AE, et al. Ryanodine receptor type 3 (RYR3) as a novel gene associated with a myopathy with nemaline bodies. *Eur J Neurol.* 2018;25(6):841-847.

34. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med.* 2017;9(386):1-11.

35. Marco-Puche G, Lois S, Benítez J, Trivino JC. RNA-Seq Perspectives to Improve Clinical Diagnosis. *Front Genet.* 2019;10(1152).

36. Kanavy DM, McNulty SM, Jairath MK, et al. Comparative analysis of functional assay evidence use by ClinGen Variant Curation Expert Panels. *Genome Med.* 2019;11(1):77.

37. Brnich SE, Rivera-Muñoz EA, Berg JS. Quantifying the potential of functional evidence to reclassify variants of uncertain significance in the categorical and Bayesian interpretation frameworks. *Hum Mutat.* 2018;39(11):1531-1541.

38. Mahmood T, Yang P-C. Western blot: technique, theory, and trouble shooting. *N Am J Med Sci.* 2012;4(9):429-434.

39. Aebersold R, Burlingame AL, Bradshaw RA. Western blots versus selected reaction monitoring assays: time to turn the tables? *Mol Cell Proteomics.* 2013;12(9):2381-2382.

40. Schessl J, Bach E, Rost S, et al. Novel recessive myotilin mutation causes severe myofibrillar myopathy. *Neurogenetics.* 2014;15(3):151-156.

41. Stastna M, Van Eyk JE. Analysis of protein isoforms: can we do it better? *Proteomics.* 2012;12(19-20):2937-2948.

42. Jayasena T, Poljak A, Braidy N, et al. Application of Targeted Mass Spectrometry for the Quantification of Sirtuins in the Central Nervous System. *Sci Rep.* 2016;6(1):35391.

43. Cozzolino F, Landolfi A, Iacobucci I, et al. New label-free methods for protein relative quantification applied to the investigation of an animal model of Huntington Disease. *PLoS One.* 2020;15(9):e0238037.

44. De Palma S, Capitanio D, Vasso M, et al. Muscle Proteomics Reveals Novel Insights into the Pathophysiological Mechanisms of Collagen VI Myopathies. *J Proteome Res.* 2014;13(11):5022-5030.

45. Dusl M, Senderek J, Müller JS, et al. A 3'-UTR mutation creates a microRNA target site in the GFPT1 gene of patients with congenital myasthenic syndrome. *Hum Mol Genet.* 2015;24(12):3418-3426.

46. Thermann R, Neu-Yilik G, Deters A, et al. Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.* 1998;17(12):3484-3494.

47. Yubero D, Natera-de Benito D, Pijuan J, et al. The Increasing Impact of Translational Research in the Molecular Diagnostics of Neuromuscular Diseases. *Int J Mol Sci.* 2021;22(8):4274.

48. Bamshad MJ, Nickerson DA, Chong JX. Mendelian Gene Discovery: Fast and Furious with No End in Sight. *Am J Hum Genet.* 2019;105(3):448-455.

49. Ebbert MTW, Jensen TD, Jansen-West K, et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol.* 2019;20(1):97.

50. Efthymiou S, Manole A, Houlden H. Next-generation sequencing in neuromuscular diseases. *Curr Opin Neurol.* 2016;29(5):527-536.

51. Mitsuhashi S, Matsumoto N. Long-read sequencing for rare human genetic diseases. *J Hum Genet.* 2020;65(1):11-19.

52. Scriba CK, Beecroft SJ, Clayton JS, et al. A novel RFC1 repeat motif (ACAGG) in two Asia-Pacific CANVAS families. *Brain.* 2020;143(10):2904-2910.

53. Sun H, Yu G. New insights into the pathogenicity of non-synonymous variants through multi-level analysis. *Sci Rep.* 2019;9(1):1667.

54. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat.* 2011;32(4):358-368.

55. Lappalainen T, MacArthur DG. From variant to function in human disease genetics. *Science.* 2021;373(6562):1464-1468.

56. Frisso G, Detta N, Coppola P, et al. Functional Studies and In Silico Analyses to Evaluate Non-Coding Variants in Inherited Cardiomyopathies. *Int J Mol Sci.* 2016;17(11):1883.

57. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-315.

58. Vögeli B, Engilberge S, Girard E, et al. Archaeal acetoacetyl-CoA thiolase/HMG-CoA synthase complex channels the intermediate via a fused CoA-binding site. *Proc Natl Acad Sci U S A.* 2018;115(13):3380-3385.

59. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* 2019;12(1):3.

60. Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell.* 2009;101(5):251-262.

61. Surikova Y, Filatova A, Polyak M, Skoblov M, Zaklyazminskaya E. Common pathogenic mechanism in patients with dropped head syndrome caused by different mutations in the MYH7 gene. *Gene.* 2019;697:159-164.

62. Osborn DPS, Emrahi L, Clayton J, et al. Autosomal recessive cardiomyopathy and sudden cardiac death associated with variants in MYL3. *Genet Med.* 2021;23(4):787-792.

63. Goullée H, Taylor RL, Forrest ARR, Laing NG, Ravenscroft G, Clayton JS. Improved CRISPR/Cas9 gene editing in primary human myoblasts using low confluency cultures on Matrigel. *Skelet Muscle.* 2021;11(1).

64. Rokosz LL, Boulton DA, Butkiewicz EA, et al. Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase: expression, purification, and characterization of recombinant wild-type and Cys129 mutant enzymes. *Arch Biochem Biophys.* 1994;312(1):1-13.

65. Steussy CN, Robison AD, Tetrick AM, et al. A structural limitation on enzyme activity: the case of HMG-CoA synthase. *Biochemistry.* 2006;45(48):14407-14414.

66. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol.* 2014;5:172.

67. Puisac B, Marcos-Alcalde I, Hernández-Marcos M, et al. Human Mitochondrial HMG-CoA Synthase Deficiency: Role of Enzyme Dimerization Surface and Characterization of Three New Patients. *Int J Mol Sci.* 2018;19(4):1010.

68. Chae JH, Vasta V, Cho A, et al. Utility of next generation sequencing in genetic diagnosis of early onset neuromuscular disorders. *J Med Genet.* 2015;52(3):208-216.

69. Ravenscroft G, Davis MR, Lamont P, Forrest A, Laing NG. New era in genetics of early-onset muscle disease: Breakthroughs and challenges. *Semin Cell Dev Biol.* 2017;64:160-170.

70. Kurul SH, Oktay Y, Töpf A, et al. High diagnostic rate of trio exome sequencing in consanguineous families with neurogenetic diseases. *Brain.* 2021.

71. Gaildrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol.* 2010;653:249-257.

72. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci.* 1998;23(6):198-199.

73. Tsoumpra MK, Fukumoto S, Matsumoto T, Takeda Si, Wood MJA, Aoki Y. Peptide-conjugate antisense based splice-correction for Duchenne muscular dystrophy and other neuromuscular diseases. *EBioMedicine.* 2019;45:630-645.

74. Li Y, Li J, Wang J, et al. Targeting 3′ and 5′ untranslated regions with antisense oligonucleotides to stabilize frataxin mRNA and increase protein expression. *Nucleic Acids Res.* 2021;49(20):11560-11574.

75. Nakamori M, Gourdon G, Thornton CA. Stabilization of expanded (CTG)•(CAG) repeats by antisense oligonucleotides. *Mol Ther.* 2011;19(12):2222-2227.

76. Mohassel P, Mammen AL. Anti-HMGCR Myopathy. *J Neuromuscul Dis.* 2018;5(1):11-20.

77. Baker SK. Molecular clues into the pathogenesis of statin-mediated muscle toxicity. *Muscle Nerve.* 2005;31(5):572-580.

78. Lin YY. Muscle diseases in the zebrafish. *Neuromuscul Disord.* 2012;22(8):673-684.

79. Ravenscroft G, Miyatake S, Lehtokari V-L, et al. Mutations in KLHL40 Are a Frequent Cause of Severe Autosomal-Recessive Nemaline Myopathy. *Am J Hum Genet.* 2013;93(1):6-18.

80. Mathews ES, Mawdsley DJ, Walker M, Hines JH, Pozzoli M, Appel B. Mutation of 3-hydroxy-3-methylglutaryl CoA synthase I reveals requirements for isoprenoid and cholesterol synthesis in oligodendrocyte migration arrest, axon wrapping, and myelin gene expression. *J Neurosci.* 2014;34(9):3402-3412.

81. Dhamija R, Chambers C. Diagnostic NGS for Severe Neuromuscular Disorders. *Pediatr Neurol Briefs.* 2015;29(11):82-82.

82. Philippakis AA, Azzariti DR, Beltran S, et al. The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery. *Hum Mutat.* 2015;36(10):915-921.

Appendix

I. Authorship attributions

Chapter 3: in preparation for publication

To whom it may concern, I, Lein Dofash, hereby testify my contribution to the study. This includes the analysis of patient whole exome sequences, discovery of the 3' UTR variant in *KLHL40,* Sanger sequencing validation of the variant, design of the *KLHL40* constructs, design and implementation of the *in vitro* assays, data collection and analysis, as well as drafting and editing the manuscript. The patient's clinical assessments, muscle histopathology and western blotting were performed by clinical colleagues in Spain.

Lein Dofash, Fathimath Faiz, Emilia Servián-Morilla, Eloy Rivas, Patricia Sullivan, Joshua Clayton, Rhonda Taylor, Nigel G. Laing, Macarena Cabrera, Gianina Ravenscroft. A *KLHL40* 3'UTR splice-altering variant causes milder NEM8. *[in preparation]*

I, as a co-author, certify that the majority of the work in this chapter is attributed to the candidate. The candidate's statements regarding their contribution to the work indicated above is appropriate.

Fathimath Faiz

Emilia Servián-Morilla

Eloy Rivas

Patricia Sullivan

Joshua S. Clayton

Rhonda L. Taylor

Nigel G. Laing

Macarena Cabrera

Gianina Ravenscroft

Chapter 4: in preparation for publication

To whom it may concern, I, Lein Dofash, hereby testify my contribution to this study. This includes the analysis of patient whole exome sequences, discovery of the pathogenic *HMGCS1* variant in the Spanish family, Sanger sequencing validation of the variant, protein modelling, design of the *HMGCS1* constructs, production of the recombinant proteins and all *in vitro* experiments except for analytical ultracentrifugation. The Spanish patients' clinical assessments and muscle histopathology were performed by colleagues in Spain. Mass spectrometry was performed in collaboration with colleagues at Bio 21, Melbourne. In addition, I affirm my contribution to the analysis of data generated, as well as drafting and editing the manuscript.

Lein Dofash, Joel Haywood, Jessica Wyllie, Tatiana P Soares da Costa, Eloy Rivas, Joshua S Clayton, David Stroud, Rhonda L Taylor, David Groth. Nigel G. Laing, Macarena Cabrera, Gianina Ravenscroft. Biallelic variants in *HMGCS1* are a novel cause of rare rigid spine syndrome. *[in preparation]*

I, as a co-author, certify that the majority of the work in this chapter is attributed to the candidate. The candidate's statements regarding their contribution to the work indicated above are appropriate.

Joel Haywood

Jessica Wyllie

Tatiana P Soares da Costa

Eloy Rivas

Joshua S. Clayton

David Stroud

Rhonda L. Taylor

David Groth

Nigel G. Laing

Macarena Cabrera

Gianina Ravenscroft

II. Conference Participation

Abstract for World Muscle Society. Published in *Neuromuscular Disorders.*

https://doi.org/10.1016/j.nmd.2021.07.059

EP.34

3'UTR variant in *KLHL40* causes nemaline myopathy

L. Dofash [1], F. Faiz [2], E. Servián-Morilla [3], E. Rivas [4], P. Sullivan [5], E. Oates [6], J. Clayton [1], R. Taylor [1], M. Davis [2], N. Laing [1], M. Cabrera-Serrano [7], G. Ravenscroft [1]

1 *Harry Perkins Institute of Medical Research, Perth, Australia;* [2] *PathWest, Perth, Australia;* [3] *Hospital Universitario Virgen del Rocío, IBiS, Sevilla, Spain;* [4] *Hospital Universitario Virgen del Rocío, Sevilla, Spain;* [5] *Children's Cancer Institute, Lowy Cancer Centre, Sydney, Australia;* [6] *The University of New South Wales, Sydney, Australia;* [7] *Instituto de Biomedicina de Sevilla/CSIC, Sevilla, Spain*

Nemaline myopathy 8 (NEM8) is typically a severe autosomal recessive disorder associated with variants in the kelch-like family member 40 gene (*KLHL40*). To date, patients with NEM8 have only been identified with coding pathogenic variants and almost all have presented with severe disease. Common features include fetal akinesia/hypokinesia, fractures, contractures, dysphagia, respiratory failure, and neonatal death (average age at death 5 months). Here, we describe a 26-year-old man with milder NEM8. He presented with hypotonia and bilateral femur fractures at birth, later developing bilateral Achilles' contractures and scoliosis, as well as elbow and knee contractures. He had walking difficulties throughout his childhood and never ran or jumped. He was wheelchair bound from age 15 after prolonged immobilization. Muscle MRI at age 13 indicated prominent fat replacement in his pelvic girdle, posterior compartments of thighs, vastus intermedius and legs. Muscle biopsy showed nemaline bodies and intranuclear rods. RNA sequencing and western blotting revealed striking reduction in KLHL40 expression and abundance in patient skeletal muscle. Using a combination of NGS panel testing and exome sequencing, we identified a likely pathogenic variant (c.∗152G>T) in the 3' untranslated region (UTR) of KLHL40 in trans with a truncating 10.9 kb deletion. In silico tools SpliceAI and Introme predicted that the c.∗152G>T variant creates a donor splice site. Confirming this, MINTIE alignment of patient muscle RNA-seq data indicated a cryptic intron is

spliced from the variant site. Our data demonstrates that c.∗152G>T splicing induces a 78 bp deletion (c.∗151_∗228del) in the 3'UTR, likely reducing mRNA stability. Investigations are ongoing to determine whether regulatory element sites are excised or if illegitimate miRNA sites are introduced. We describe for the first time a 3'UTR splice-mediated mechanism of NEM8 and encourage consideration of 3'UTRs during screening of patients.