

Diese Dissertation wurde angefertigt
unter der Leitung von Prof. Benedikt Schoser
im Bereich vom Friedrich-Baur-Institut
an der Ludwig-Maximilians-Universität München
Direktor: Univ. Prof. Dr. med. Günter U. Höglinger



***Comparison of splicing and nuclear envelope alterations in
three muscular dystrophies***

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Fakultät für Biologie der
Ludwig-Maximilians-Universität München

vorgelegt von
Vanessa Jimayma Todorow
aus
München
2023

Ersgutachter/in:	Professor Dr. Benedikt Grothe
Zweitgutachter/in:	Professor Dr. Benedikt Schoser
Tag der Abgabe:	29.11.2023
Tag der mündlichen Prüfung:	19.03.2024

EIDESSTATTLICHE ERKLÄRUNG

Ich versichere hiermit an Eides statt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist.

Die vorliegende Dissertation wurde weder ganz noch teilweise bei einer anderen Prüfungskommission vorgelegt.

Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 29.11.2023

Vanessa Todorow

Table of Contents

ABBREVIATIONS	5
LIST OF PUBLICATIONS	6
ABSTRACT	7
1 INTRODUCTION	8
1.1 Splicing: a source of diversity	8
1.2 Splicing complexes and process	10
1.3 Alternative splicing in muscle	12
1.4 Alternative splicing in disease	15
1.5 Neuromuscular diseases	16
1.5.1 Myotonic dystrophy (DM)	16
1.5.1.1 RNA toxicity.....	18
1.5.1.2 DMPK haploinsufficiency	19
1.5.1.3 Chromatin changes	20
1.5.2 Emery-Dreifuss-Muscular-Dystrophy (EDMD).....	22
1.5.2.1 Historical remarks and clinical presentation	22
1.5.2.2 The nuclear envelope.....	24
1.5.2.3 Mutation of NE proteins in EDMD	25
1.5.3 Facioscapulohumeral muscular dystrophy (FSHD).....	26
1.5.3.1 Contraction and hypomethylation of the 4q35 subtelomeric region	27
1.5.3.2 Aberrant expression of DUX4	28
2 RESULTS	30
2.1 Splicing is affected in DM1, EDMD and FSHD	30
2.2 The nuclear envelope is affected in DM1, EDMD and FSHD	35
3 DISCUSSION	38
3.1 Advances and limitations in splicing analysis	38
3.1.1 DEXSeq.....	39
3.1.2 MAJIQ.....	40
3.1.3 ISA	40
3.2 Comparison of DM1, EDMD and FSHD symptomology	41
3.3 Splicing in DM1, EDMD and FSHD	42
3.4 The nuclear envelope is affected in DM1, EDMD and FSHD	45
3.5 Final remarks and future directions	48
4 REFERENCES	50
5 APPENDICES	64
5.1 Publication I	64
5.2 Publication II	80
5.3 Publication III	90
5.4 Manuscript I	113
5.5 Manuscript II	143
ACKNOWLEDGEMENTS	200

CURRICULUM VITAE 201

ABBREVIATIONS

A3SS	alternative 3' splice site
A5SS	alternative 5' splice site
AE	alternative exon usage
AS	alternative splicing
ASO	antisense oligonucleotide
BPA	branch point adenosine
CDM	congenital DM
DM1	myotonic dystrophy type 1
DM2	myotonic dystrophy type 2
EDMD	Emery-Dreifuss-muscular-dystrophy
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
ES	exon skipping
ER	endoplasmic reticulum
FSHD	Facioscapulohumeral muscular dystrophy
GO	gene ontology
GSEA	gene set enrichment analysis
INM	inner nuclear membrane
IR	intron retention
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
LAD	lamin associated domains
LINC	linker of nucleoskeleton and cytoskeleton
LSV	local splicing variation
logFC	logarithmic fold change
mRNA	messenger RNA
MRI	magnetic resonance imaging
ncRNA	non-coding RNA
NE	nuclear envelope
NET	nuclear envelope transmembrane protein
ONM	outer nuclear membrane
PCR	polymerase chain reaction
PNS	perinuclear space
PSI	percent spliced in
RBP	RNA binding protein
RNAseq	RNA-sequencing
rRNA	ribosomal RNA
snRNP	small nuclear ribonucleoprotein
snRNA	small nuclear RNA
SS	splice site
TAD	topologically associating domains
UTR	untranslated region

LIST OF PUBLICATIONS

Main publications and manuscripts included in this dissertation:

- Todorow, V.**, S. Hintze, A. R. W. Kerr, A. Hehr, B. Schoser, and P. Meinke. 2021. 'Transcriptome Analysis in a Primary Human Muscle Cell Differentiation Model for Myotonic Dystrophy Type 1', *International Journal of Molecular Sciences*, 22.
- Todorow, V.**, S. Hintze, B. Schoser, and P. Meinke. 2022. 'Nuclear envelope transmembrane proteins involved in genome organization are misregulated in myotonic dystrophy type 1 muscle', *Frontiers in Cellular and Developmental Biology*, 10: 1007331.
- de Las Heras, J. I., **Todorow, V.**, L. Krečinić-Balić, S. Hintze, R. Czapiewski, S. Webb, B. Schoser, P. Meinke, and E. C. Schirmer. 2022. 'Metabolic, fibrotic, and splicing pathways are all altered in Emery-Dreifuss muscular dystrophy Spectrum patients to differing degrees', *Human Molecular Genetics*.
- Schätzl, T., **Todorow, V.**, L. Kaiser, H. Weinschrott, B. Schoser, H. Deigner, P. Meinke, M. Kohl. 2023. 'New insights in FSHD pathology: meta-analysis identifies involvement of the neuromuscular junction, nuclear envelop and spliceosome', in submission
- Kao, S-Y., **Todorow, V.**, Monteagudo Mesas, P., Hoelze, L., Walper, P., Ravichandran, K., Weinert, O., Graupner, A., Alqassem, A., Borozan, L., Barone, M, Re Cecconi, A., Forne, I., Menden, M., Straub, T., Piccirillo, R., Meinke P., Schoser, B., Hayashi,R., Canzar, S., Spletter ML 'Splicing mediated by U2-associated Scaf6/CHERP is necessary for myogenesis in Drosophila and vertebrates', in preparation

Additional Publications:

- Donandt, T., **V. Todorow**, S. Hintze, A. Graupner, B. Schoser, M. C. Walter, and P. Meinke. 2023. 'Nuclear Small Dystrophin Isoforms during Muscle Differentiation', *Life (Basel)*, 13.
- Hintze, S., L. Baber, F. Hofmeister, S. Jarmusch, **V. Todorow**, S. Mehaffey, F. Tanganelli, U. Ferrari, C. Neuerburg, D. Teupser, M. Bidlingmaier, J. G. Marques, B. Koletzko, B. Schoser, M. Drey, and P. Meinke. 2022. 'Exploration of mitochondrial defects in sarcopenic hip fracture patients', *Heliyon*, 8: e11143.

ABSTRACT

Pre-mRNA splicing is a vital step in global and tissue-specific gene regulation and the source of proteomic diversity in eukaryotes. When misregulated, splicing can lead to detrimental disease states, and muscle and brain belong to the tissues that are most sensitive to such alterations. Consequently, many diseases caused by splicing misregulation are known, often affecting the neuromuscular system. Myotonic dystrophy type I (DM1) is described as a trans-splicing disease caused by an expansion of a CTG repeat in the *DMPK* gene. The transcription of this repeat-containing gene leads to hairpin structures in the *DMPK*-mRNA, which are recognized by various RNA-binding proteins, whose usual function is then inhibited. This results in multisystemic mis-splicing events in muscle and brain causing muscle weakness and wasting, myotonia, insulin resistance, sleepiness, cognitive decline, and a whole range of additional clinical symptoms. Among others, it has been shown that DM1 shows signs of segmental progeria, leading to premature aging symptoms in the muscular system. Yet, splicing alterations cannot explain all the symptoms observed in patients and more contributors have been identified and are still under investigation. Here, I explore new contributors to the pathogenesis of DM1 and further compare it with two other hereditary muscular dystrophies, Emery-Dreifuss-Muscular-Dystrophy (EDMD) and Facioscapulohumeral Muscular Dystrophy type 1 (FSHD1). I can thereby show that these muscle diseases not only share similar symptoms but also mechanisms linked to the phenotype. Notably, splicing is not only broadly altered in DM1 but also in EDMD and FSHD1 with partially overlapping mis-splicing events. Still, there are differences in splicing factor expression patterns between the three muscle diseases, which are also reflected in the type of events detected via bioinformatic tools. I thus question which splicing events in DM1 are disease-specific due to splicing factor sequestration and which are secondary to general muscular dystrophy alterations. Similarly, genes of the nuclear envelope are affected in all three diseases on the expression level, with sometimes similar, sometimes opposite trends. Importantly, the expression of these genes is correlated with disease severity which proposes them as potential biomarkers and candidates for differential diagnosis. In summary, this work demonstrates the benefits of comparative bioinformatic analyses of big data sets for causally unrelated but similar diseases, whose current view is thereby challenged.

1 INTRODUCTION

1.1 Splicing: a source of diversity

At the present day, a midsummer day in 2022, it is thoroughly evident to any life scientist that splicing is the removal of introns from the pre-mRNA to produce mature mRNA. Any comprehensive textbook will point to splicing as the main source of proteomic diversity in higher eukaryotes with hundreds of thousands of proteins arising from “only” 25,000 genes. Take a color-by-number blueprint and, by changing the order of the numbers, create not one but many pictures, each telling a different story, finely tuned to time, space and purpose. Before we delve into the intricate details of this complex process, let’s look back upon the events leading to its’ discovery – for seldom can we learn more than by unravelling the mysteries of the past.

60 years ago, scientists had elucidated how proteins are synthesized in bacteria: manufactured in massive macromolecular machineries, the ribosomes, with mRNAs determining the amino acid sequence (Palade 1955). Although this process seemed to apply in principle to animal cells, the nature and synthesis of the informational RNA remained unclear and extensive studies were undertaken to shed light on the matter (Scherrer et al. 1963). A very useful tool at the time was pulse-chase radioactive labelling, in which tritiated uridine is added to the culture medium of cells which then incorporate the labelled uridine into their RNA. After a period of time, the labelled culture medium is replaced with normal medium, allowing the labelled RNA to be tracked over time. This revealed that the nuclear RNA was many kilobases in length – far longer than the cytoplasmic RNA - and was rapidly degraded, with only a small fraction being exported to the cytoplasm (Harris et al. 1962). Was the nuclear RNA a precursor of the actual mRNA? A preliminary version, that was modified before being exported to the cytoplasm? It certainly seemed so. The nuclear RNA had the same poly-A tail and 7-methylguanine 5’-5’ phosphotriester cap as the cytoplasmic mRNA used for protein synthesis (Darnell et al. 1971; Rottman et al. 1974). But whether it was really the *same* molecule, reduced to about a quarter of its length and then transported out of the nucleus – it was impossible to prove at the time (mid-1970s).

A crucial hint finally came from observations of pre-rRNA derived from a single 13.7 kb transcript resulting in 18S, 5.8S and 28S rRNA. Different products from a single transcript – generated by cleavage. It was therefore not too far a stretch to suggest that the same cleavage process occurred in mRNA. The proof finally came in 1977 with hybridisation experiments of purified adenovirus 2 (Ad2) mRNA with its' DNA, described in the now famous paper by Berget *et al.* In 70% formamide, RNA-DNA hybrids have a melting temperature just above that of DNA-DNA hybrids, allowing the hybridisation of mRNA to DNA to be observed by electron microscopy (**Figure 1**, (Berget *et al.* 1977; Berk 2016)).

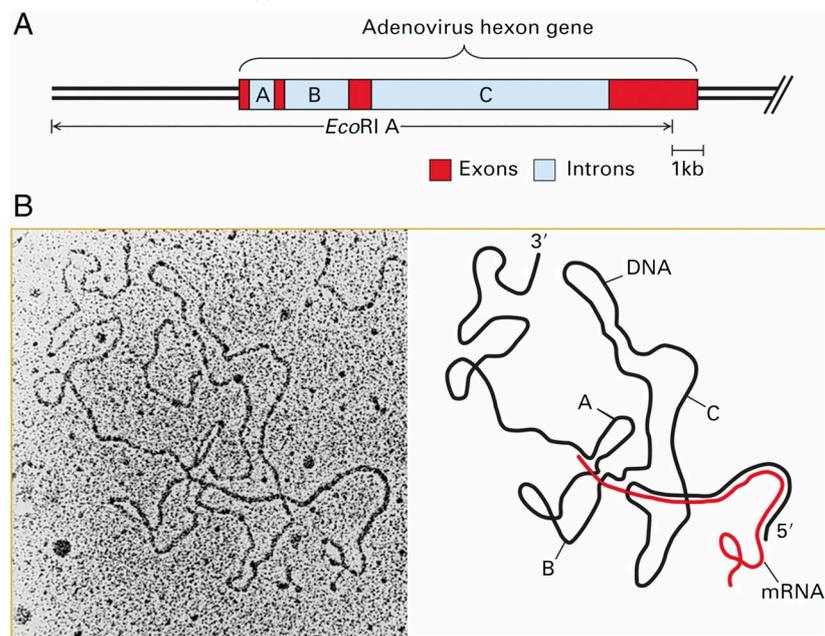


Figure 1: DNA-RNA hybridization of the Adenovirus 2 hexon gene. Hybridization can be observed via electron microscopy in 70% formamide. Original electron microscopy image from (Berget *et al.* 1977) and simplified scheme from (Berk 2016). Permission is granted by PNAS under <https://www.pnas.org/pb-assets/authors/authorlicense-1633461587717.pdf>.

This clearly showed that the processed mRNA transcript is complementary to certain - shorter - segments of the genome (now known as exons, in red) with larger segments spanning the in-betweens, which are non-complementary, forming DNA-loops (now called introns, in blue, regions A, B and C). The authors of the manuscript correctly connected the dots by pointing out that the mature mRNA is likely to be produced by splicing out the longer segments and joining of the shorter ones, while retaining both the poly-A tail and 5' cap. This explained the greater length of nuclear transcripts and the rapid degradation of most of them. Shortly afterwards, the same hybridisation technique was used to show that several different Ad2 mRNA transcripts could be generated from the same original sequence by joining different segments - now known as alternative

splicing. Advances in cloning allowed the same observations to be made the following year for non-viral genes such as chicken ovalbumin (Breathnach et al. 1977) or mouse β -globin (Kinniburgh et al. 1978).

Scientists quickly understood the implications of this discovery for the diversity and complexity of multicellular organisms, most of which have the exon-intron structure of genes. Different gene products can be made from a single locus, depending on the needs of the cell – but that is not all. Exons correspond to structural domains in the final protein with specific functions and foldings, as shown in 2004 (Liu et al. 2004) but suggested already in 1978 by Walter Gilbert (Gilbert 1978). Intriguingly, large intronic regions, especially if they contain repeats, allow DNA recombination and thus the shuffling of exons between genes, providing entire functional units for the generation of new proteins – and, if advantageous, evolutionary selection and fixation. A source of diversity without the need to duplicate a whole gene to vary the function. It was also Walter Gilbert who coined the terms intron (for **in**tragenic region) and exon (for **exp**ressed region).

1.2 Splicing complexes and process

Biochemically, splicing is a combination of two successive transesterifications in which a hydroxyl (OH) group of one nucleotide "attacks" the phosphate group of another nucleotide. This reaction is called a "nucleophilic attack". Rather than being "cleaved", a particular nucleotide forms a new bond, releasing the previous one (Filipowicz et al. 1983; Konarska et al. 1982; Konarska et al. 1985). The nucleotides at which transesterifications occur are located in introns, which are evolutionarily conserved and are called splice sites (SS). There are three conserved sites: the 5' SS at the beginning of the intron with an AG, the 3' SS at the end of an intron with a GU and finally the branch point adenosine (BPA) 15-50 base pairs upstream of the 3' SS (Ruskin et al. 1985; Ruskin et al. 1984). Splicing is strictly dependent on the presence of these sequences, as evidenced by the complete inhibition of the entire process when mutated. Less conserved but necessary for correct splicing are enhancer and silencer regions, which are mostly bound by SR (serine and proline rich) proteins and either promote or repress splice site usage. These regions occur in exons (exonic splicing enhancer (ESE), exonic splicing silencer (ESS)) and introns (intronic splicing enhancer (ISE), intronic splicing silencer (ISS)) (Blencowe 2000; Wu et al. 1993; Lavigne et al. 1993).

In the first transesterification, the 2'OH group of the BPA attacks the phosphate of the 5' SS guanine phosphate group, which then form a new bond leaving a free OH group at the 3' end of the 5' splice site. This structure looks like a lasso and is thus called lariat (**Figure 2**). The free hydroxy group of the 5' exon then attacks the phosphodiester of the 3' exon, finally releasing the intron and joining the two exons together (Will et al. 2011).

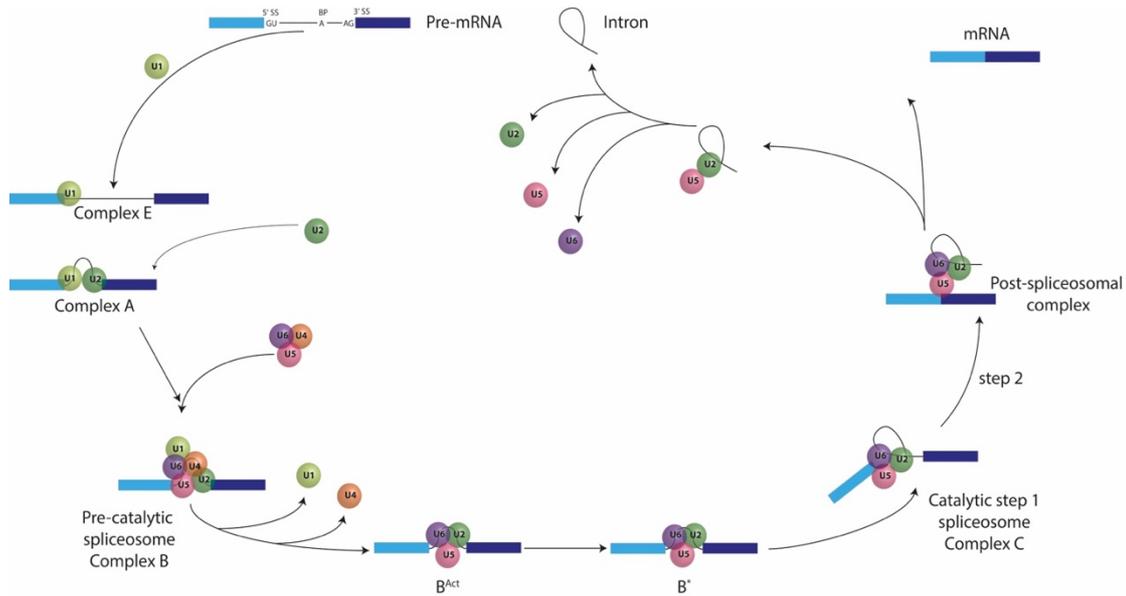


Figure 2: Schematic overview of the splicing reaction.

The transesterifications are mediated by the spliceosome, a huge machinery that contains five RNAs and about 150 RNA-binding proteins (RBPs). The sequential steps and re-formations of the spliceosome assembly, binding and catalytic conformations have been studied in detail (Hicks et al. 2005) and reviewed in (De Conti et al. 2013). Prior to the actual splicing reaction, the three conserved sequences are recognized by specific subcomplexes. In the case of introns between 200-250 nucleotides in length, the U1 small nuclear ribonucleoprotein (snRNP) complex binds the 5'SS through base pairing of the U1 snRNA to the pre-mRNA. Shortly thereafter, the BPA is recognized by non-snRNPs such as SF1, which now form the early complex (complex E). U2 snRNA then binds to the BPA region and conformational changes bring the 5'SS, 3'SS and the BPA into close proximity. This process is ATP-dependent, unlike the formation of complex E, and the resulting complex is called the pre-spliceosome or complex A. The pre-assembled U4/U6.U5 tri-snRNP is then recruited (pre-catalytic complex B) and several conformational rearrangements lead to U1 and U4 snRNPs leaving the spliceosome after the catalysis has been prepared. The sites of the first nucleophilic attack, the BPA and the 5'SS, are brought into the active center (B_{act} complex). DEAH-box RNA helicase Prp2 is

then activated, which is referred to as the B* complex, and catalyzes the first transesterification, yielding the C complex which then mediates the second transesterification. The lariat intermediate is released and quickly degenerated, although it has been reported that released introns might act as regulatory non-coding RNA (ncRNA)(Rearick et al. 2011).

If the intron length exceeds 250 nucleotides – which is often the case in mammals – the intron-defined spliceosome assembly described above switches to an exon-defined assembly in which the 3'SS and 5'SS of an exon are recognized first (Fox-Walsh et al. 2005). The BPA downstream of the 5'SS is then bound and the exon definition complex switches to an intron definition complex. However, the mechanistic details of this switch remain unknown. Correct splicing is a challenge for several reasons. First, exons are much shorter than introns and are therefore masked by sheer mass. Second, splicing usually occurs in parallel with transcription elongation by RNA polymerase II, requiring excessive interplay between two tightly regulated mechanisms involving hundreds of proteins (Tennyson et al. 1995; Howe 2002). Third, splice site selection must be precise (a single nucleotide can cause a frame shift and thus nonsense-mediated decay) and specific to the respective tissue and developmental state, often with several options per exon, while some are skipped altogether – this is known as alternative splicing.

1.3 Alternative splicing in muscle

The process described above is called constitutive splicing – the removal of a single intron and the joining of two adjacent exons. Alternative splicing was first observed in 1977 (Chow et al. 1977) and since then, several types of events have been proposed, which can also be combined into complex events. The most studied alternative splicing events are exon skipping (ES), intron retention (IR), alternative 3' and 5' SS (A3SS, A5SS) and mutually exclusive exons. Pan et al. estimated that more than 95% of all human genes undergo alternative splicing resulting in gene isoforms with different functions (Pan et al. 2008). This allows adaptation to the specific needs of different tissues, as well as to developmental and environmental changes. While some genes have only a few annotated isoforms, others have several dozen, such as the giant protein dystrophin (DMD) with more than 40 annotated protein-coding isoforms. **Figure 3** shows the most common types of AS.

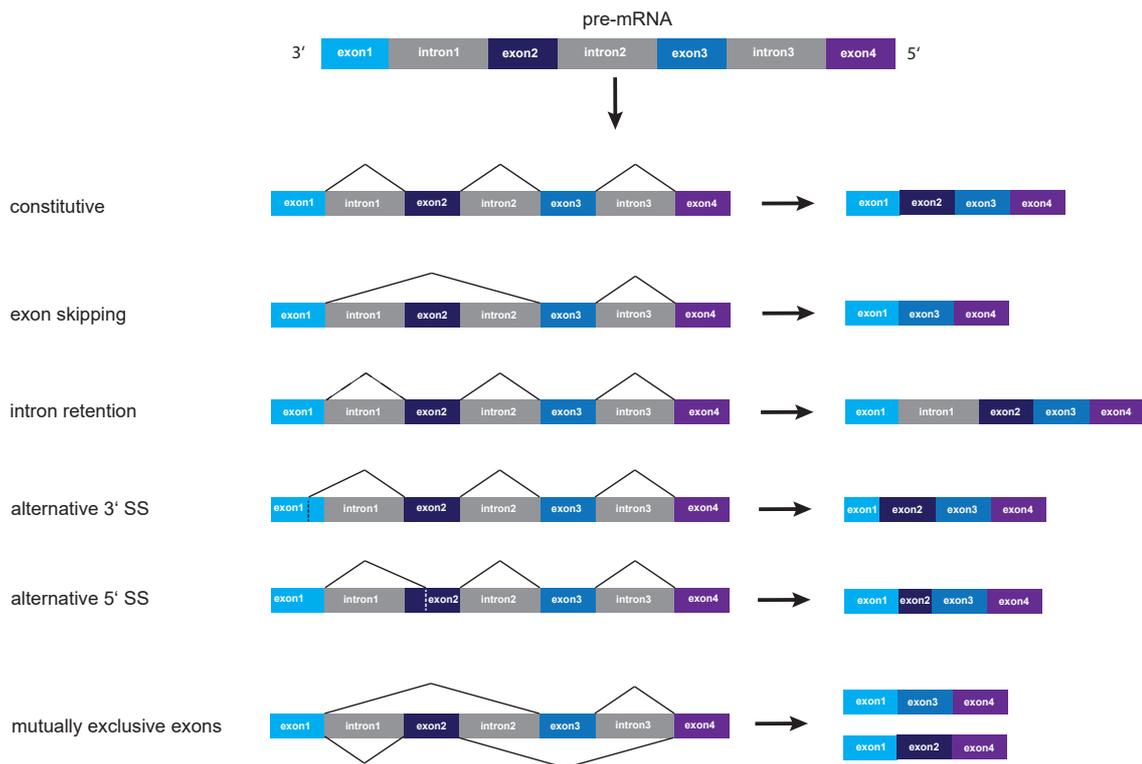


Figure 3: Most frequent splicing modes and resulting mRNA transcripts

Muscle is a highly ordered tissue consisting of post-mitotic, multinucleated myocytes which are able to contract and are thus the basis of all voluntary (and some involuntary) motion in animals. In these huge cells, also called myofibrils, we find the smallest contractile unit of muscles, the sarcomeres. Sarcomeric structure and components are highly conserved among species, between invertebrates and vertebrates, from fly to man (Fukushige et al. 2006; Jung et al. 2008; Buckingham 2017). Thus, studies in model organisms like *Drosophila melanogaster* (fruit fly) or *Danio rerio* (zebrafish) advanced our knowledge about muscle structure and development enormously. Sarcomeres consist of thin filaments of actin and tropomyosin-troponin complexes which are anchored in the Z-disc, and of thick myosinII-filaments anchored in the M-line (Squire 1997). Through interaction, these long filaments can slide past each other and thus shorten the sarcomere in length, which is an ATP-dependent process (also sliding-filament theory) (Lorand 1953; Huxley et al. 1954). If this happens in all consecutive sarcomeres in a myofibril – and all myofibrils in a myofiber – and all myofibers in a muscle bundle – the entire muscle shortens, i.e. contracts. For this to work, a multitude of additional interactors, enzymes and anchor proteins like troponins (TNNT), tropomodulins (TMOD), tropomyosins (TPM) and titin (TTN), are necessary (Brenner et al. 1987). Their main

function is to anchor actin and myosin to the Z-disc and M-line, as well as regulating their interaction which is Calcium dependent: tropomyosin blocks the interaction of myosin with actin but uncovers the binding site upon calcium release and interaction (Lehman et al. 1994).

Muscle was one of the first tissues in which alternative splicing was demonstrated to be massively used and essential for proper function, development and maintenance. However, even after extensive screenings of different tissues, muscle is still among the ones with the highest proportion of differentially spliced genes, along with leukocytes, mammary glands, testes, brain and several cancers (Castle et al. 2008; Pan et al. 2008). In contrast, adrenal gland, fetal lung and adipose tissue display the fewest events. Further, the number of AS events is reflected in the amount of differential expression in each tissue. Much research has been focused on embryonic and postnatal heart development, revealing that most changes are driven transcriptionally and post-transcriptionally (Baralle et al. 2017). Many AS events have been identified that occur in highly regulated temporal patterns and are conserved across species (Staudt et al. 2012). These splicing events are regulated by several RNA-binding proteins (RBPs), including CUGBP and ETR3-like factors (CELFs) and the muscleblind-like (MBNLs) family proteins. CELF1 and MBNL1 were shown to act antagonistically with CELF1 being highly expressed in the early stages and more than 10-fold downregulated during heart development, while MBNL1 is more than 4-fold upregulated. Impressively, restoring the embryonic protein levels of both proteins in adult mouse hearts results in embryonic splicing patterns of more than 50% of the developmentally regulated AS events (Kalsotra et al. 2008).

One of the most prominent examples of the importance of splicing in muscle is the giant protein titin (TTN) whose gene contains a remarkable 364 exons and with this, the highest number of exons in all human genes. TTN spans from the M-line to the Z-disc, which means both the I-band as well as the A-band region (Labeit et al. 1990). *TTN* exons corresponding to the I-band region undergo differential splicing resulting in a longer isoform (N2BA isoform) and a shorter one (N2B) (Labeit et al. 1995). The ratio of N2BA and N2B determines the length of the sarcomere and thus the passive tension of the muscle. Accordingly, N2BA can be found in higher proportions in neonates, while adults express primarily N2B, a result of increased muscle activity and physical forces exerted on adult muscle. Similarly, the ratio of N2B to N2BA in cardiomyocytes is higher than in skeletal muscle, since the passive tension needs to be manifolds higher to ensure efficient blood pressure (Loescher et al. 2021).

1.4 Alternative splicing in disease

Given the fundamental role of splicing in gene expression and the enormous complexity of its regulatory system, it is not surprising that it has been implicated in a wide range of diseases, sometimes as a major player, sometimes as one of many dysregulated levels, as in cancer. An early study in 1992 estimated that approximately 15% of all genetic diseases are caused by mutations that result in splicing defects, but at that time only splice sites and BPA sequences were considered, while enhancer and silencer regions were neglected (Krawczak et al. 1992). Even today, 30 years later, the number of hereditary diseases caused by misregulated splicing can only be estimated, but it may be as high as 50% (Cartegni et al. 2002). Scientists distinguish between *cis*-acting and *trans*-acting splicing alterations. The former involve mutations in the DNA sequence required for splicing: 5'SS, 3'SS, BPA as well as enhancer and silencer regions. The second refers to the dysregulation of splicing factors. Table 1 shows some of the best-studied splicing disorders, their genetic cause (the gene in which the mutation occurs) and the consequence for splicing. Here, I will focus on neurodegenerative and muscular diseases as these are the most relevant to my field of research, but many others are known (Scotti et al. 2016; Douglas et al. 2011; Li et al. 2021).

Disease	Gene	Mechanism	Consequence	Ref.
<i>Cis</i>				
Early-onset Parkinson Disease (PD)	<i>PINK1</i>	5'SS mutation	Cryptic SS usage → exon7 skipping	(Samaranch et al. 2010)
Frontotemporal dementia with parkinsonism chromosome 17 (FTDP-17)	<i>MAPT</i>	ESS mutation	Increased exon 10 inclusion	(Iovino et al. 2014)
Duchenne muscular dystrophy (DMD)	<i>DMD</i>	Exon deletions and skipping	Frameshift → NMD	(Muntoni et al. 2003)
Becker muscular dystrophy (BMD)	<i>DMD</i>	ESS introduction	Exon 31 skipping	(Disset et al. 2006)
Dilated cardiomyopathy (DCM)	<i>LMNA</i>	A3SS	Exon 4 extension by 9 nt	(Otomo et al. 2005)
Limb girdle muscular dystrophy type 1B (LGMD1B)	<i>LMNA</i>	5'SS mutation	Intron 9 retention → NMD	(Muchir et al. 2000)
Hutchinson–Gilford progeria syndrome (HGPS)	<i>LMNA</i>	A5SS	Exon 11 reduction by 150 nt → progerin	(Eriksson et al. 2003)

Trans

Myotonic Dystrophy (DM)	<i>DMPK</i> <i>CNBP</i>	CTG/CCTG repeat expansion	Splicing factor sequestration, imbalance between MBNL and CELF protein families	(Miller et al. 2000)
Spinal muscular atrophy (SMA)	<i>SMN1</i>	deletion	Loss-of-function	(Lorson et al. 1999)
Amyotrophic lateral sclerosis (ALS)	<i>FUS</i>	Point mutation	Altered interaction with splicing factors → mis-splicing of targets	(Sun et al. 2015)
Dilated cardiomyopathy (DCM)	<i>RBM20</i>	Point mutation	Altered interaction with splicing factors → TTN mis-splicing	(Guo et al. 2012)

Table 1: *Cis- and trans-splicing diseases, causing genes and effect on splicing*

One of the best studied *trans*-splicing diseases is myotonic dystrophy (DM), caused by expansion of CTG/CCTG repeats and subsequent splicing factor sequestration to the mutated transcripts. In the following paragraph I will describe this muscle disease and its proposed pathomechanisms as well as two other, splicing-unrelated muscular dystrophies.

1.5 Neuromuscular diseases

The Friedrich-Baur-Institute, in which this doctoral thesis was conducted, is specialized in the clinical treatment and scientific investigation of all types of neuromuscular diseases. Hereditary neuromuscular disorders are generally defined by the malfunction of peripheral nerves, the neuromuscular endplate, or skeletal muscles with different sites of injury: neuronal cell bodies, Schwann cells, axons, neuromuscular junction or skeletal muscle. In the following I will outline three muscular dystrophies, DM1, Emery-Dreifuss-muscular-dystrophy (EDMD) and facioscapulohumeral muscular dystrophy (FSHD); and summarize the current point of view on their pathomechanisms. In the discussion, I will compare these diseases by challenging some of the established opinions.

1.5.1 Myotonic dystrophy (DM)

Myotonic dystrophies (dystrophia myotonica; DM) are hereditary, muscular disorders which affect the muscles, central nervous system (CNS) and endocrine system marking them as multisystemic diseases (Minnerop et al. 2011; Dahlqvist et al. 2015). Two types of DM are described, and both are caused by a repeat expansion in the untranslated regions of two distinct genes, *DMPK* (dystrophia myotonica protein kinase, DM type 1) and *CNBP* (CCHC-type zinc finger nucleic acid binding protein, DM 2). With a

prevalence of ~1:8000 -1:10000, DM1 is the most common late-onset muscular dystrophy (Liao et al. 2022), and a recent study identified an even higher prevalence 0 1:2100 of DM in newborns in the state New York (Johnson et al. 2021). Its' congenital form (CDM1) is rarer and the most severe one with a high mortality rate in the neonatal state (Echenne et al. 2013). This type is only well established for DM1. The symptoms are wide-ranging and variably pronounced between patients (Wenninger et al. 2018). As the name suggests, one of the characteristic symptoms is myotonia, which is defined by prolonged relaxation after muscle contractions and mostly affects the skeletal muscles of the extremities, i.e. hands (grip myotonia) and lower legs. Muscle weakness and wasting, however, is the most prevalent symptom with more than 45% of patients suffering from it (De Antonio et al. 2016).

In DM1, distal muscles are more affected than proximal muscles, while DM2 patients experience muscle weakness more often in the proximal muscles (Wenninger et al. 2018). Further, insulin resistance, cardiac arrhythmia, cataracts, sleep-wake-cycle disturbances, and decreased intelligence and cognitive decline occur. Pathologically, dystrophic findings encompassing central nuclei, pyknotic nuclear clumps, angulated fibers, and slow fiber type atrophy with fast fiber type hypertrophy, are commonly observed (**Figure 4**).

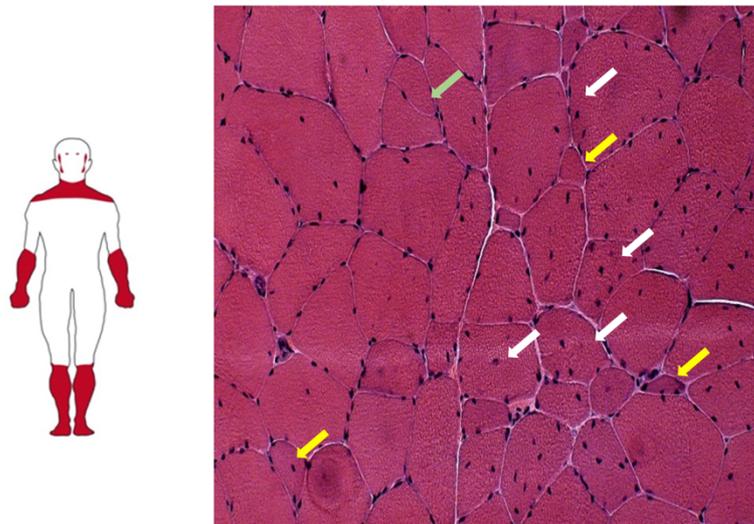


Figure 4: DM1 pathology. Primarily, distal muscle groups are affected as shown on the left. H&E stain shows typical pathological changes, including internal nuclei (white arrows), small and angular fibers (yellow arrows) and splitting fibers (green arrow).

So how does a nucleotide repeat expansion cause a diverse symptomology like that on the mechanistic level? In fact, many repeat expansion diseases are described, among others Huntington's disease and Fragile X syndrome, and they all have a highly

variable phenotype (Paulson 2018). The first possibility that comes to mind is a loss of function of the gene harboring the repeat expansion, due to reading frame shifts (non-triplet repeat expansions) or altered transcription factor binding (repeat in regulatory regions). In the case of Fragile X syndrome for instance, a CGG-repeat expansion in the 5' UTR of the *FMRI* gene leads to hypermethylation and subsequent silencing of the gene (McConkie-Rosell et al. 1993). This is, however, the minority of cases. More common is a toxic gain of function: in protein-coding sequences, a triplet repeat codes for a specific amino acid which can then be found in the protein ultimately altering its behavior. For example, CAG codes for glutamine and several polyglutamine diseases are known, among others Huntington's disease or spinal and bulbar muscular atrophy. In other disease entities, however, as in DM1 and DM2, it is not the protein that displays the toxic gain of function, but the RNA.

1.5.1.1 RNA toxicity

As an important regulatory feature, RNA can form secondary structures which are then recognized by RNA-binding proteins and regulate RNA stability and decay, splicing, post-transcriptional modifications and transport (Sanchez de Groot et al. 2019). Unsurprisingly, RNA repeats can self-assemble and thus form highly ordered hairpin structures. In DM1, the CUG stem-loop structures in the *DMPK*-mRNA is then bound by muscleblind-like protein family members (*MBNL1-3*), which act as splicing regulators and are consequently depleted from the nucleoplasm, unable to fulfill their usual function (Miller et al. 2000; Mankodi et al. 2000). MBNL1 sequestration and resulting foci can be visually observed in immunofluorescence experiments in DM patient-derived muscle cell cultures (**Figure 5**). Interestingly, MBNL proteins were found to be expressed at low levels in developing embryonic muscle, but are highly elevated in adult muscle, suggesting that it regulates the expression of adult muscle-specific isoforms (Kalsotra et al. 2008). Consistently, its depletion in DM1 yields an increase of embryonic splice isoforms which are insufficient to meet the physiological requirements of adult muscle tissue (Lin et al. 2006). This is supported by the increased activity of CELF-family members in DM due to hyperphosphorylation, which are usually active in embryonic muscle and display reduced activity in adult muscle (Kalsotra et al. 2008). As mentioned under chapter 1.3, MBNL and CELF are thus thought to be antagonistic players in muscle development, a mechanism that is reversed in DM.

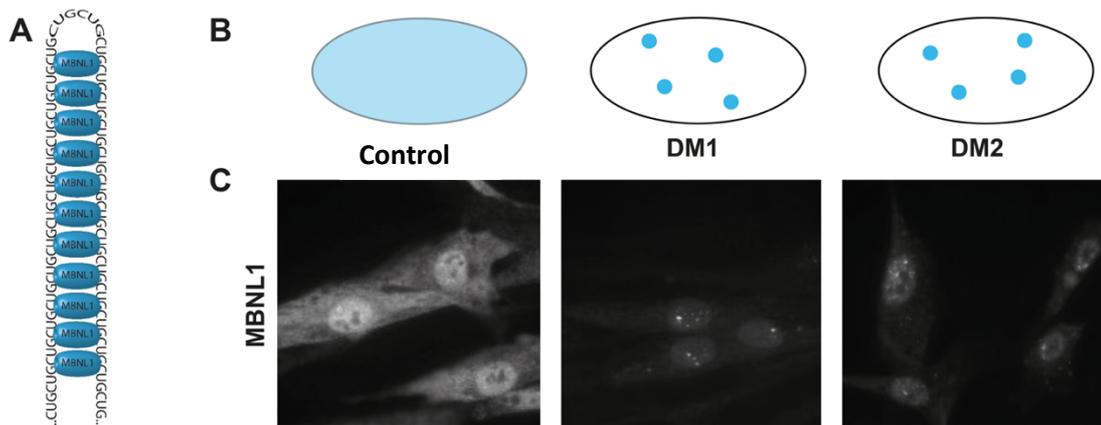


Figure 5: RNA-toxicity in DM1. A) Simplified depiction of the CUG-repeat containing DMPK transcript, which forms hairpin structures and thereby sequesters MBNL1. B) Nuclei of control, DM1 and DM2 cells, in blue MBNL1 distribution. In controls, MBNL1 is distributed evenly throughout the nucleus, but forms bright foci in DM1 and DM2. C) Immunofluorescence images of MBNL1 staining in Controls, DM1 and DM2 myoblasts. Foci can be observed in DM1 and DM2 but not in Controls. Further, MBNL1 is highly expressed in the cytoplasm in controls, but not in DM1 and DM2.

Many resulting mis-splicing events have been described in DM, among others the inclusion of exon 7a in the chloride channel *CLCN1* leading to a frame shift and premature stop codon (Charlet et al. 2002). *CLCN1* is necessary for reaching the resting membrane potential quickly after muscle excitation, and its mutation or mis-splicing yields myotonia, one of the signature symptoms of DM as described above. The inclusion of exon 17a in myomesin (*MYOM1*) results in the encoded protein being more elastic and occurs usually in embryonic muscle, but also in DM1 (Koebis et al. 2011). Other well studied mis-spliced genes in DM1 include *BINI*, *ATP2A1*, *CACNA1S*, *RYR1*, *LDB3* and *INSR*, all of which can be coupled to pathways contributing to DM1 pathology (Nakamori et al. 2013). The importance of splicing misregulation in the DM1 pathology is undoubted, however, no DM1 model, neither MBNL knockout, CELF overexpression nor repeat expansion models, are sufficient to explain the whole range of symptoms and the variability between patients. Further, many of the described mis-splicing events are also observed in other neuromuscular diseases or during muscle regeneration (Bachinski et al. 2014; Orengo et al. 2011), potentially suggesting splicing as a general secondary effect of muscle disease. Thus, the search for alternative or additional explanations revealed more mechanisms contributing to the DM1 etiology, which I want to outline briefly.

1.5.1.2 DMPK haploinsufficiency

It was shown that *DMPK*-mRNA accumulates in the nuclei of affected individuals, probably because the binding of MBNL proteins and foci formation prevents further

processing and the transport to the cytoplasm (Novelli et al. 1993; Michel et al. 2015). Naturally, this leads to reduced DMPK protein levels which might contribute to the DM1 phenotype. However, *Dmpk*^{-/-} mouse models yield controversial results: some do not display a muscle phenotype while others have mild DM-like symptoms (Reddy et al. 1996; Carrell et al. 2016). Intriguingly, a mouse model in which human *DMPK* with normal repeat length is overexpressed develops several DM1 key features like myotonia, myofiber type I atrophy and central nuclei (O'Coilain et al. 2004).

DMPK is a serine- threonine-kinase whose cellular localization is yet unclear – it was described to localize at the plasma membrane, nuclear envelope as well as the cytoplasm (Harmon et al. 2008; Harmon et al. 2011; van der Ven et al. 1993). As a kinase, DMPK phosphorylates its substrates and regulates their activity. Suggested substrates include phospholamban, which functions as an inhibitor of a calcium pump in the sarcoplasmic reticulum of the heart SERCA2a (Kaliman et al. 2005), and phospholemman, a regulator of the Na⁺/K⁺-ATPase important for excitation-contraction (EC) coupling in cardiac and skeletal muscle (Mounsey et al. 2000). Phospholemman is mainly phosphorylated by the protein kinases A and C (PKA, PKC), it is thus conceivable that these kinases compensate for the reduction of DMPK protein (Palmer et al. 1991; Lindemann 1986). Additionally, the activity of PKC was shown to be increased in DM1 and responsible for the hyperphosphorylation of CELF1 (Kuyumcu-Martinez et al. 2007). Whether DMPK has exclusive functions that cannot be compensated by other kinases and consequently, whether protein loss of DMPK contributes to DM1 is currently unknown.

1.5.1.3 Chromatin changes

Another hypothesis that was tested is based on the observation that CTG-repeat expansions can lead to chromatin changes: nucleosomes are packed tighter and thus reduce the expression of e.g. neighboring genes (Poeta et al. 2020; Wang et al. 1994). In case of *DMPK*, the transcription factor (TF) *SIX5* is in close proximity and its expression is actually downregulated in DM1 patients (Thornton et al. 1997). Consistently, a potential contribution of *SIX5* reduction is supported by a mouse knockout model which develops an increased risk of cataracts, a frequent symptom in DM1 (Klesert et al. 2000; Voermans et al. 2015).

Further, DNA methylation patterns have been investigated. Hypermethylation, especially at CpG islands (cytosine and guanosine rich DNA regions), is generally

associated with gene silencing due to impaired binding of transcription factors and enhancers (Newell-Price et al. 2000). Moreover, methylation as posttranslational modification of histones has a similar effect: hypermethylation results in a denser chromatin structure and thus gene silencing, while hypomethylation leads to loose chromatin packaging and genes can be easier accessed and transcribed (Strahl et al. 2000). In DM1, hypermethylation of CpG islands nearby the CTG (up- and downstream) expansion have been described by several labs (López Castel et al. 2011; Steinbach et al. 1998) and in CDM this hypermethylation concurred with binding sites for the transcription regulatory protein CTCF (Barbé et al. 2017). Although a methylation of the CTCF binding sites (and its consequences) is controversial, their existence at various sites in *DMPK* is still highly interesting (two of them shown in **Figure 6**). Notably, the CTCF binding site in the 3' UTR which also harbors the CTG repeat expansion was only found in myoblasts and myotubes but not in non-muscle cell lines (Buckley et al. 2016).

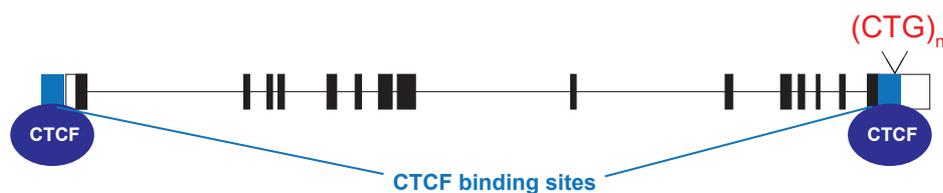


Figure 6: *DMPK* displays CTCF binding sites in the CTG repeat region in the 3' UTR, as well as in the 5' UTR. The latter was shown to be muscle specific.

CTCF is involved in the generation of topologically associating domains (TADs), self-interacting chromatin regions shaping the regulatory landscape, i.e. genes depending on the same TFs and enhancers form domains to ensure robust gene expression (Dixon et al. 2012; Phillips-Cremins et al. 2013; Sofueva et al. 2013; Zuin et al. 2014). The consequential 3D organization of the genome is specific to cell type, developmental state and environmental cues. Interestingly, the repeats associated to diseases but not repeats in general can be found at the boundaries of TADs, making it likely that their genetic instability might be a cause of the 3D localization (Sun et al. 2018). Sun et al. showed that the CTG repeat in the 3' UTR of *DMPK* belongs to the TAD marking repeats and in *FMRI*, which causes Fragile X Syndrome, TAD boundaries are disrupted in patients with expanded repeats, probably due to hypermethylation of CTCF binding sites. Since 3D genome organization is highly cell type specific, this would also explain why some tissues

are more affected than others. If these boundaries can be affected by repeat expansion without hypermethylated CTCF binding sites is currently unknown.

Therapeutic approaches in DM target either specific symptoms like Mexiletine improving myotonia by blocking sodium channels, or the effects of toxic-RNA using antisense oligonucleotides (ASOs). An interesting approach is the blocking of GSK3 β by Tideglusib, which should reduce CELF1 activity. Tideglusib was shown to mainly help with neurological symptoms like sleepiness and social distancing, but also with myotonia in some cases (Horrigan et al. 2020). ASOs were demonstrated to improve mis-splicing events, however the overall effect is often small and the delivery to site of action is still problematic (Pascual-Gilabert et al. 2021).

1.5.2 Emery-Dreifuss-Muscular-Dystrophy (EDMD)

1.5.2.1 Historical remarks and clinical presentation

EDMD is a rare disorder with a prevalence of 1: 100.000 (Nelson 2000), and its systematic investigation is relatively young. It was first described in 1966 by Emery and Dreifuss as a neuromuscular disorder that causes early joint contractures, progressive muscle weakness and wasting and cardiac conduction defects (Meinke ; Emery et al. 1966). But it was not until 30 years later that the genetic basis was identified – or at least, one of many: *EMD*, or emerin, was found to be mutated in EDMD patients (Bione et al. 1994). In 1996, *EMD* was identified as a nuclear envelope (NE) gene, which came as a surprise at the time because most muscular dystrophies were thought to be caused by sarcolemmal proteins (Manilal et al. 1996). However, not all patients had mutations in *EMD* which led to the search of EDMD-related genetic changes. Not all of these have yet been identified, but all known are associated with the NE. Lamin A/C (*LMNA*) was the second to be discovered, but together, *EMD* and *LMNA* mutations constituted less than half of all mutations found in patients (Bonne et al. 1999). By now, *FHL1* (Gueneau et al. 2009), the nesprins *SYNE1-2* (Zhang et al. 2007) and *SUN1* (Meinke et al. 2014) have been linked to EDMD. Intriguingly, these EDMD-linked genes are ubiquitously expressed, and their products are required for proper gene expression in all cell types. In particular, lamin A is necessary for chromatin organization in general and thus regulates thousands of genes (Broers et al. 2006; Verstraeten et al. 2007; Meinke et al. 2015). So how do mutations in these genes cause a tissue-restricted phenotype? Before describing

the hypotheses scientists came up with for this conundrum, I will summarize the clinical presentation of EDMD.

As is typical of muscular dystrophies, EDMD is very variable and heterogeneous among patients and even within family members in terms of symptomology and genetics – so much so that Emery himself pleaded for the term “Emery-Dreifuss-syndrome” in 1989 (Emery 1989). In line with this variability, EDMD can be classified as X-linked (*EMD*, *FHL1*), autosomal dominant (*LMNA*, *SYNE1*) and autosomal recessive (*LMNA*, *SUNI*) and mutations in additional genes can act as modifiers, either improving or worsening the phenotypic outcome (Meinke et al. 2020). For example, one family member may show signs of all three EDMD-like symptoms described (contractures, muscle weakness, and heart defects), while another has only heart defects (Bonne et al. 2000). Contractures occur usually in the elbows, Achilles tendons and postcervical muscles and are among the first symptoms to appear in the first two decades of life. As a result, the neck and spine become rigid and the patient’s mobility is impaired. Progressive muscle weakness and wasting can be seen in the proximal upper and distal lower extremities, although this can vary between X-linked and autosomal dominant forms of EDMD (Bonne G 2004 ; Meinke). Muscle pathology can also vary between *EMD* and *LMNA* mutation patients. Both display varied fiber size, increased central nuclei and immature fibers, while little necrosis is observed (**Figure 7**). However, patients with mutations in *EMD* may have pronounced atrophy and predominance of the slow-fiber type and fast-fiber hypertrophy, while *LMNA* patients rather display atrophic fibers in both types, but other studies found no difference between emerinothic and laminopathic EDMD regarding myopathic changes (Astejada et al. 2007) (<https://neuromuscular.wustl.edu/pathol/edmdpath2.htm>, <https://neuromuscular.wustl.edu/pathol/cmdlamac.htm>). Finally, cardiomyopathy is very common and usually occurs after muscle weakness but may be isolated from other symptoms and is then called dilated cardiomyopathy type 1A. Arrhythmias and conduction defects may require a pacemaker or heart transplant.

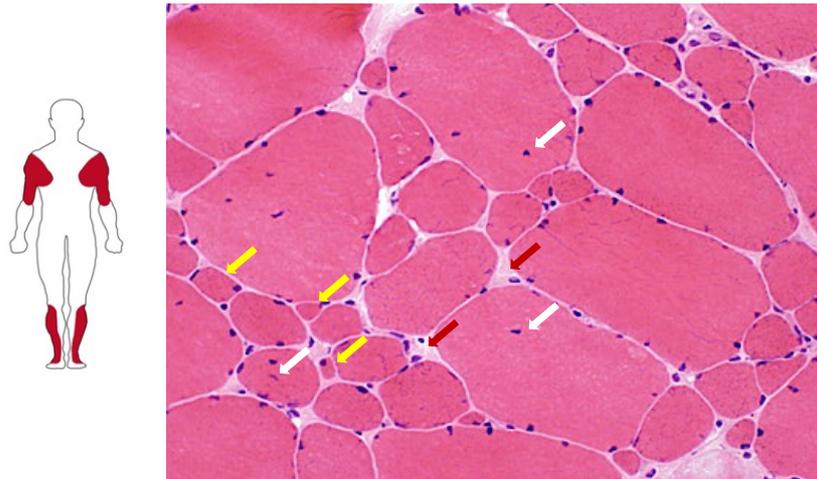


Figure 7: EDMD pathology. Proximal upper and distal lower muscle groups are affected as shown on the left. H&E stain shows typical pathological changes, including internal nuclei (white arrows) and small and angular fibers (yellow arrows). Red arrows show increases endomysial fibrosis. Muscle biopsy staining is provided by <https://neuromuscular.wustl.edu>.

1.5.2.2 The nuclear envelope

Since EDMD is clearly understood as a disease of the nuclear envelope, it is necessary to explain NE structure and functions. The nucleus is surrounded by two lipid bilayer membranes and contains the genome of eukaryotic cells (Kite 1913), as well as a myriad of regulatory proteins and RNAs. Structurally, the NE consists of the inner lamina, the double membrane and many nuclear pore complexes, which are huge mega-dalton complexes spanning both membranes and allow transport of large molecules like proteins (**Figure 8**)(Watson 1955). The outer nuclear membrane (ONM) is contiguous with the endoplasmic reticulum (ER), while transmembrane proteins can interact with nuclear lamina, chromatin or cytoskeleton (Hetzer et al. 2005). The connection between the cytoskeleton and the nucleus has several important functions like nuclear positioning and migration, and especially mechanotransduction: the conversion of mechanical into chemical signals. The main player in mechanotransduction is the LINC complex (Linker of nucleoskeleton and cytoskeleton) consisting of an ONM KASH domain protein (SYNE1-4, KASH5 and LRMP) and an INM SUN domain protein (SUN1/2) which interact in the perinuclear space (PNS) (Starr et al. 2003; Wilhelmsen et al. 2006; Sosa et al. 2012). KASH domain proteins can interact with the cytoskeleton, e.g. nesprin 1 (SYNE1) interacts with dynein and kinesin through the N-terminus of the KASH domain and with F-actin through a calponin homology domain (Wilhelmsen et al. 2005; Zhang et al. 2002). SUN proteins on the other hand interact with the nuclear lamina which has

vital functions in chromatin organization (Haque et al. 2006). Different LINC complexes can be formed with differential KASH domain lengths supporting various degrees of mechanical force (Alam et al. 2015). The lamina consists of different lamins type A and B. A type lamins derive from the same gene (*LMNA*) which gives rise to lamin A and lamin C through alternative splicing (Broers et al. 2006). In contrast, B type lamins are products of two distinct genes *LMNB1* and *LMNB2* (Höger et al. 1990). Lamins were found to anchor the chromatin to the nuclear envelope and while it was believed for a long time that they bind only the heterochromatin and are thus involved in gene silencing, this view has long been corrected as many genes are transcribed although located at the NE (Fawcett 1966; Guelen et al. 2008; Wu et al. 2017). However, the general trend of the heterochromatin being at the NE while the euchromatin is in the nuclear interior is true (van Steensel et al. 2017). Importantly, lamins interact with many INM transmembrane proteins (NETs) to achieve the anchorage of huge molecules as chromosomes, among others emerin and SUN1/2, but also LAP1 and a FHL1 isoform (Wilson et al. 2010).

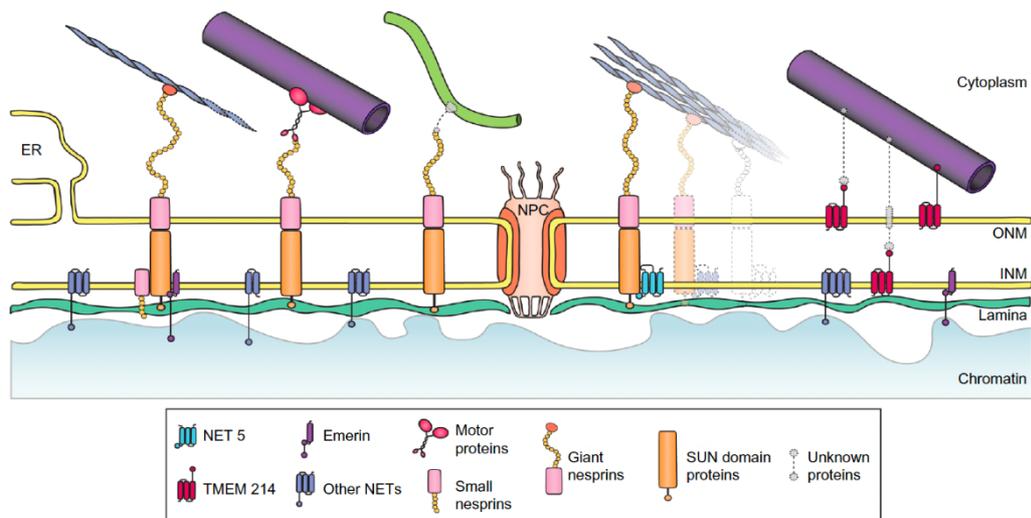


Figure 8: Schematic of the nuclear envelope. From (Meinke et al. 2015). Permission granted according to <https://creativecommons.org/licenses/by-nc/4.0/>.

1.5.2.3 Mutation of NE proteins in EDMD

As mentioned above, all proteins of the NE described in this section can cause EDMD if mutated. The question of how they receive tissue specificity has been addressed in the Schirmer lab by the identification of tissue specific NETs that might interact with ubiquitous NETs and lamins in very specific regions (Korfali et al. 2012). Notably, hundreds of disease-causing *LMNA* mutations have been identified and many of them are unexpectedly restricted to a certain tissue which gave rise to a whole field of research

called laminopathies (Vigouroux et al. 2013). A mutation in *LMNA* at the amino acid 377 or 453 causes a muscle phenotype, while a mutation in 482 has a phenotype in fat tissue. Proposedly, these mutations reduce the binding affinity of lamin A for muscle-specific or fat-specific NETs, respectively, but this has not been proven yet. Among the muscle-specific NETs, *PLPP7* and *TMEM38A* have been studied in detail and indeed both play an important role in muscle differentiation, and their overexpression in heterologous systems causes muscle specific genes being pulled towards or released from the NE (Robson et al. 2016; de Las Heras et al. 2017). Importantly, mutations in *PLPP7* and *TMEM38A* have been found in EDMD patients by now (Meinke et al. 2020). As for the pathomechanism behind the observed phenotype, it is expected that muscle-specific gene expression and signaling through mechanotransduction at the NE is impaired, eventually affecting muscle strength, sarcomeric structure, contraction and metabolism.

To date, there is no specific treatment for EDMD and similarly to DM1, alleviation of symptoms is the main focus. Stretching and muscle strengthening exercises help against joint contractures and slow down progressive muscle wasting. Since cardiac involvement is frequent, pacemakers are often necessary (Heller et al. 2020).

1.5.3 Facioscapulohumeral muscular dystrophy (FSHD)

The last muscle disease I want to summarize is FSHD. As depicted above, all muscular dystrophies are highly diverse among patients, making it painfully difficult to distinguish between primary and secondary effects. FSHD is again a highly variable muscular dystrophy. There is not only a high degree of inter-individual and intra-familial variability, but even between muscle groups with sometimes only the left or right ones being affected (left-right asymmetry) or even in the same muscle group with a certain muscle bundle being affected while one close to it being undistinguishable from healthy muscle (Dixit et al. 2007; Tawil et al. 2006; Schätzl et al. 2021). Consistently, the age of onset varies from early childhood to middle age with most patients showing first signs of the disease between 20 and 40 (Attarian et al. 2012). The first skeletal muscles that are involved are often face and shoulder muscles while the limb-girdle and leg muscles follow later (Tawil et al. 2006). While FSHD is seen as a relatively benign muscular dystrophy by some, as much as 20% of patients become wheelchair-bound above the age of 50 (Statland et al. 2013; Tawil et al. 2006). On the other hand, up to 20% of patients are paucisymptomatic or asymptomatic. As for the prevalence, FSHD is the third most common muscular

dystrophy with 1:20.000 people affected, which is likely underestimated due to the high number of asymptomatic carriers (Sposito et al. 2005; Lunt et al. 1991). Compared to other muscular dystrophies, tissues other than muscle are usually not affected and the life expectancy is not reduced (Hamel et al. 2018). As for the muscle histopathology, it has been found difficult to establish a specific picture. Nonetheless, in later disease stages a dystrophic pattern with muscle fibrosis, limited necrosis, variable fiber size (atrophy and hypertrophy) as well as central nucleation can be observed (**Figure 9**).

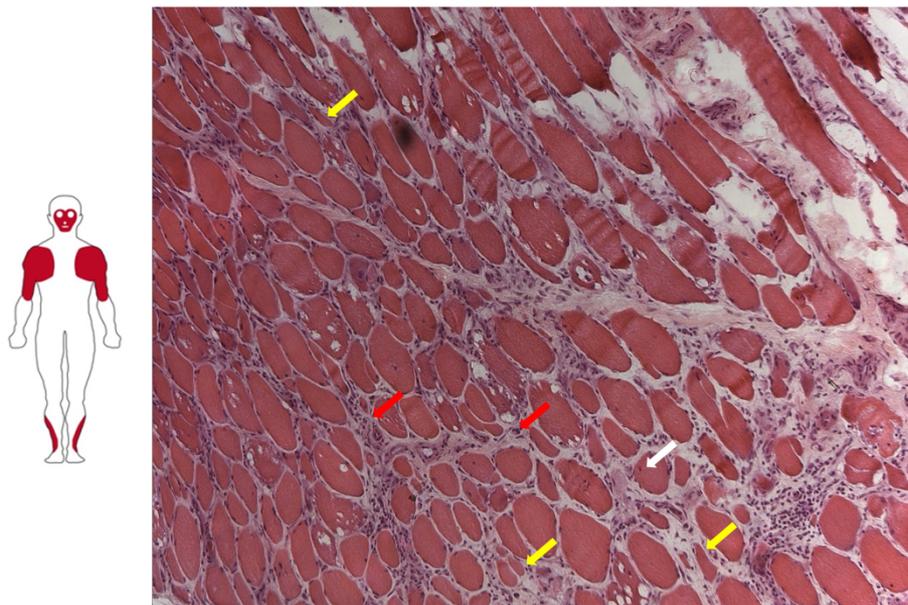


Figure 9: FSHD pathology. Facial, proximal upper and distal lower muscle groups are affected as shown on the left. H&E stain shows typical pathological changes: small and angular fibers (yellow arrows), fibrosis (red arrows) and central nuclei (white arrow).

1.5.3.1 Contraction and hypomethylation of the 4q35 subtelomeric region

The genetic cause of FSHD has been found only 100 years after the first description of the disease in 1885 (Landouzy et al. 1885; Wijmenga et al. 1990). In the subtelomeric region of chromosome 4 exists a macro-satellite repeat consisting of D4Z4 units (3.2 kb long, location 4q35) which are in head-to-tail orientation (van Deutekom et al. 1993). These repeats are scattered throughout the genome but only in the telomeric region of chromosome 10 they are similarly ordered in tandems. Interestingly, the repeat array on chromosome 10 is insensitive to repeat number as there can be between 1 and 100 units, while on chromosome 4, there are 11 to 100. A reduction below 11 repeats is associated with type 1 FSHD (FSHD1), which is the predominant form with 95% of all FSHD patients suffering from it (Preston 1999). The DNA content of these regions is complex

and usually associated with heterochromatin, but there is also a small retrogene encoded, a homeobox transcription factor called DUX4 (Wijmenga et al. 1990; Wijmenga et al. 1992). The remaining 5% of the FSHD cases (FSHD2) is caused by a hypomethylation of the 4q35 region, mediated through SMCHD1 or DNMT3B, which regulate chromatin architecture (Lemmers et al. 2015; van den Boogaard et al. 2016).

Since the repeats on chromosomes 4 and 10 are strikingly similar but only contractions on the first cause FSHD, a still ongoing hunt after the mechanistic explanation began. Many theories have been proposed, among others that the repeat functions as a barrier between heterochromatin and euchromatin and its reduction leads to a spreading of heterochromatin to neighboring regions on chromosome 4 and subsequent silencing of genes encoded there (Tawil et al. 2014). Another hypothesis suggested that the function of transcriptional repressors that bind in the D4Z4 region is inhibited by the contraction leading to an upregulation of adjacent genes like FRG1 and FRG2 (Gabellini et al. 2002). Needless to say, these findings are contradictory and could not be verified by other research groups (Masny et al. 2010; Klooster et al. 2009), or as in the case of FRG2 are indeed frequently upregulated in most FSHD patients, while in others, the whole locus is deleted alongside the repeat array (Deak et al. 2007). Thus, an FRG2 upregulation is most likely a secondary effect that might contribute but is not necessary for FSHD pathology.

1.5.3.2 Aberrant expression of DUX4

A change in FSHD research came with the observation that the retrogene *DUX4*, that is present in every D4Z4 unit, is aberrantly expressed in FSHD patient muscle (Dixit et al. 2007) although usually restricted to early developmental stages as well as testes and thymus (Snider et al. 2010). As mentioned above, DUX4 is a homeobox-containing transcription factor and its targets include further transcription factors like ZSCAN4 and LEUTX. The levels of DUX4 in patient cell lines and muscle specimens were always found to be very low, raising the question whether a gene with less than one copy per nucleus can cause FSHD. It is thus suggested that *DUX4* expression comes in wave-like patterns and its downstream targets inhibit skeletal muscle regeneration, induce stress and immune responses and thereby apoptosis (Geng et al. 2012; Knopp et al. 2016; Bosnakovski et al. 2008). The overexpression of DUX4 in various cell lines revealed increased states of apoptosis, but also yielded a list of DUX4 target genes that were also

found to be elevated in FSHD patients, now referred to as DUX4 biomarkers – which are more reliably detected than DUX4 expression itself (Geng et al. 2012; Rickard et al. 2015; Yao et al. 2014). Notably, Gaillard et al. proposed long-distance interactions between the D4Z4 region on chromosome 4 and the nuclear envelope which could lead to vast expression changes – which are indeed observed in FSHD (Gaillard et al. 2019). Since the discovery of its involvement in FSHD, DUX4 is the prime target of therapeutic approaches. However, researchers are confronted with major difficulties: for high-throughput drug screening, cell culture is the starting point – however, DUX4 is very lowly and variably expressed in control as well as FSHD myoblast or myotube cell cultures. Trying to repress a gene that is only low abundant in the first place is surely challenging. Still, transcriptional activators (i.e. SMARCA5 and BRD2) and repressors (SMCHD1 and PARP1) have been identified which are now subject of therapeutic approaches, namely small-molecule inhibitors of activators and overexpression of repressors (Himeda et al. 2018; Sharma et al. 2016; Goossens et al. 2019). Risks and benefits of such a therapy have to be carefully weighed though, as these transcription factors have more targets than only DUX4. On top of that, the actual contribution of DUX4 to the FSHD phenotype is yet to be clarified – the detection of this protein is so rare and moreover, animal models overexpressing DUX4 fail to reproduce the FSHD symptomology, that some scientists question DUX4 having the central role assigned to it, but rather acts as a mediator (Salsia et al. 2023; Banerji et al. 2021).

2 RESULTS

For detailed results found in each muscular dystrophy, please refer to the publications in the Appendix section. In the following, I use the RNAseq data from these publications for a deeper comparison of DM1, EDMD and FSHD which goes beyond what is described in the publications. I will focus on splicing regulation and alterations of the nuclear envelope and discuss the results in the next chapter.

2.1 Splicing is affected in DM1, EDMD and FSHD

Using a list of splicing factors divided in alternative splicing, core complex, A complex, B complex, Bact complex, C complex and P complex, I analyzed the expression of splicing factors in severely affected DM1 and FSHD patient muscle biopsies (Todorow et al. 2021; Todorow et al. 2022)(Schätzl et al. in submission), as well as differentiated EDMD myotubes, as quantified by RNAseq (de Las Heras et al. 2022)(**Figure 10**). Surprisingly, there are more splicing factors misregulated in EDMD (~170) and FSHD (~130) compared to DM1 (~80), although splicing changes are (to the best of our knowledge) a secondary effect there. Moreover, most splicing factors are downregulated in EDMD and FSHD, while there are similarly many factors up- as downregulated in DM1. Noteworthy, the downregulation is more severe in DM1 with $\log_2FCs < -1$, while the foldchanges in FSHD and EDMD range mostly between 0 and -1. When comparing all splicing factors misregulated in each dystrophy, they share 24, among others NOVA1/2, SF3A3, CELF1 and RBM20 (Venn Diagram **Figure 10**). Intriguingly, MBNL1 is upregulated in EDMD, and downregulated in FSHD, and although not significantly downregulated on the RNA-level in DM1, its protein displays a loss-of-function in DM1 as described in the introduction (chapter 1.5.11).

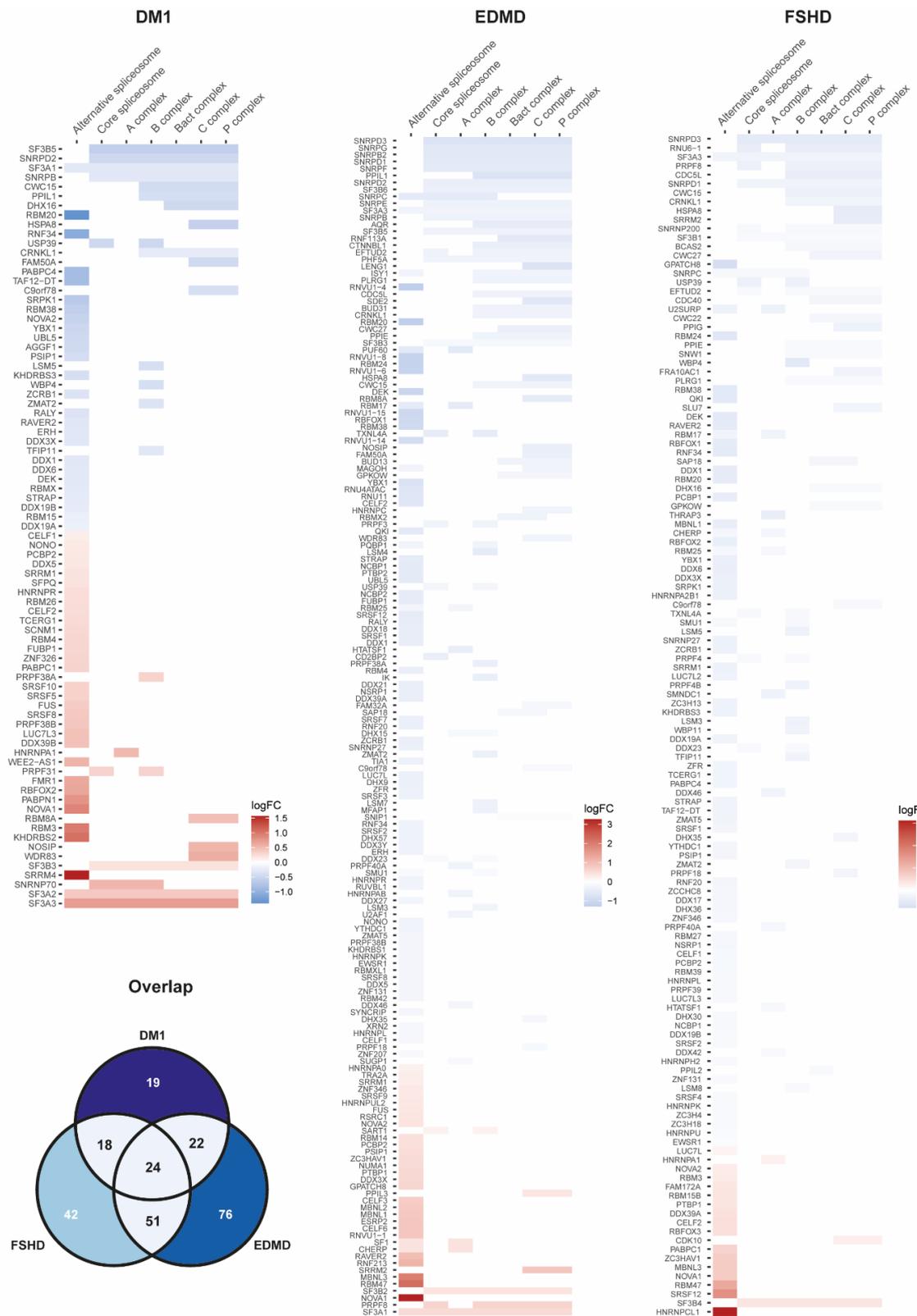


Figure 10: Heatmaps of splicing factor gene expression (transcriptional level) in DM1, EDMD and FSHD. On the bottom left, the overlap between them is shown.

This raises the questions of whether 1) the transcriptomic changes are reflected at the proteomic level and 2) the general trends at the expression level match the types of mis-splicing events detected. To this end, the differential usage of local splicing variations (LSV) was quantified for each dystrophy compared to healthy controls (delta-PSI: differential percent spliced in) and finally compared to each other. In DM1, MAJIQ detected 342 genes being mis-spliced, in EDMD and FSHD 552 and 462, respectively. The bar chart in **Figure 11** shows the distribution of the most common alternative splicing events (exon skipping (ES), alternative exon (AE), intron retention (IR), alternative 3' SS (A3SS) and alternative 5' SS (A5SS)) in each disease. Strikingly, there are by far more alternative exons used in DM1 compared to EDMD and especially FSHD. In contrast, EDMD and FSHD display high amounts of exon skipping and even double or multiple exon skipping events (Schätzl et al.).

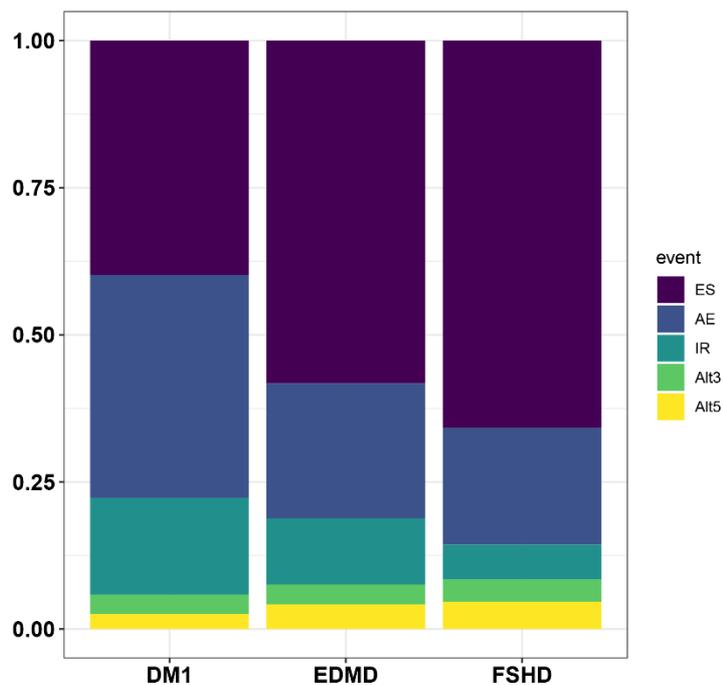


Figure 11: Alternative splicing events distribution in DM1, EDMD and FSHD. Event type indicated on the right; ES: exon skipping, AE: alternative exon usage, IR: intron retention, Alt3: alternative 3' SS, Alt5: alternative 5' SS.

I then compared the genes mis-spliced in each muscular dystrophy. 35 genes are commonly mis-spliced among them and ~180 are shared by at least two of them (**Figure 12 A**), many of which encode for sarcomeric structure proteins like *NEB*, *TNNTs*, *TTN* and *MYHs*. Importantly, the genes that are unique to each dystrophy contribute to very

similar pathways that seem to play a role in muscle weakness and wasting, namely muscle development, contraction, structure, NMJ formation and protein synthesis (**Figure 12 C**).

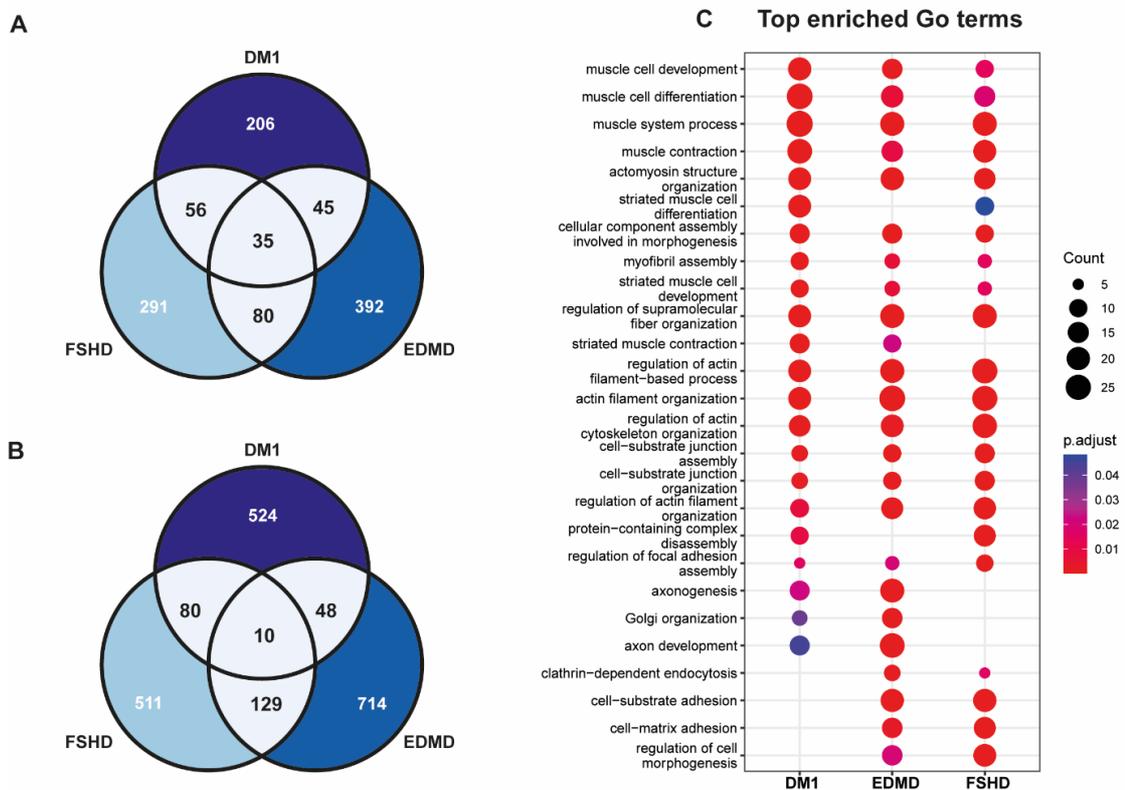


Figure 12: Comparison of mis-spliced genes and splicing events in DM1, EDMD and FSHD. A) Overlap of genes mis-spliced in the respective dystrophy. B) Overlap of specific splicing events (LSVs). C) Comparison of GO term enrichment analysis between dystrophies. Size of dots corresponds to number of genes, color corresponds to adjusted p-values.

The pathways that are not shared, are mostly redundant terms, for instance “striated muscle cell differentiation” is not present in EDMD, but “muscle cell differentiation” is. Notably, there is a higher overlap of mis-spliced genes between FSHD and EDMD, in which similar splicing factors are affected and importantly, they also share the same pattern of downregulation. On the LSV level, 10 loci are alternatively spliced between all three dystrophies and a remarkable 129 between EDMD and FSHD – even though myotubes and mature muscle have different splicing profiles (**Figure 12 B**). Among the 10 shared LSVs is *GFPT1*, an NMJ architectural protein encoding gene, which is one of the best described mis-spliced genes in DM1 (Figure 13)(Nakamori et al. 2013). *GFPT1* exon 10 skipping is increased in all DM1, EDMD and FSHD (PSI-values indicated in numbers under bars). In *BINI*, there is a preferential intron 14 retention in

DM1 but a double exon skipping in FSHD (**Figure 13**), while in *MBNL1* there is a triple exon skipping (exons 4,5 and 6) in DM1, but a double exon skipping (exons 5 and 6) in FSHD.

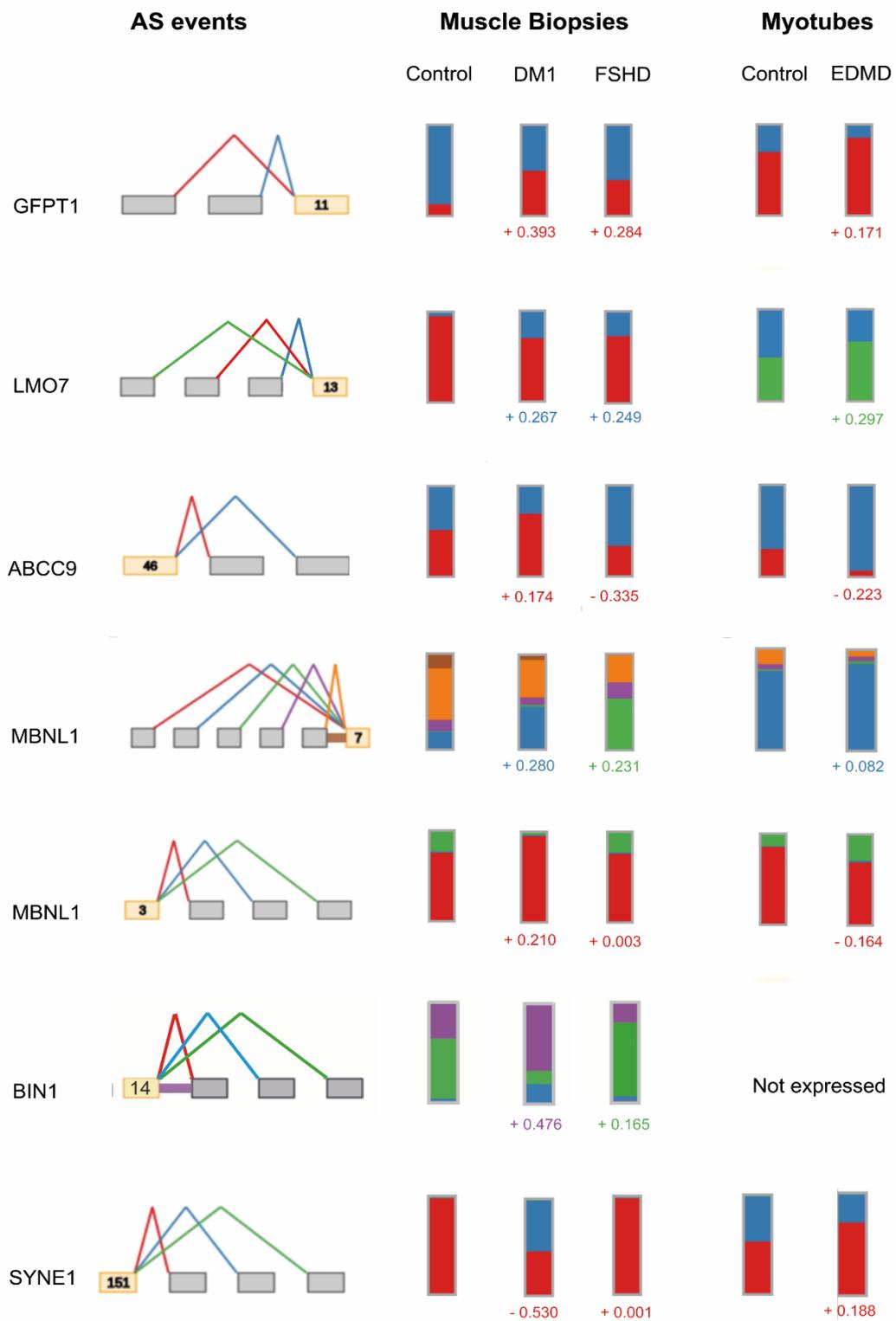


Figure 13: Splicing events in DM1, EDMD and FSHD. Exons are indicated as grey boxes, reference exon in yellow. Events are displayed as colored lines. Percentage of splice event usage are shown in the bars for control muscle biopsy and control myotubes, as well as DM1 and FSHD biopsies, and EDMD myotubes. PSI-values are below the bars, color corresponds to the event in the same color.

Regarding *MBNL1*, MAJIQ also identifies an exon 4 inclusion in DM1 (PSI-value +0.21), while controls show a double exon skipping in about 20% of the cases. DM1 thus resembles control myotubes more than mature muscle (**Figure 13**, event 5), while EDMD myotubes are more similar to mature muscle. FSHD has no significant change at this LSV. Similarly, skipping of the muscle-specific exon DV23 (exon 152) in *SYNE1* in DM1 resembles control myotubes, while EDMD myotubes have a higher inclusion of DV23 and thus shift towards adult muscle splicing (see also (Todorow et al. 2022)).

MAJIQ further identifies LSVs shared with EDMD or FSHD that have not been investigated before, for instance the transcription factor *LMO7* exon 12 inclusion. *LMO7* has important muscle-specific functions and interestingly, interacts with emerin at the NE (Holaska et al. 2006). Wang et al. found that exon 20 skipping in *LMO7* was among the most dysregulated splice events in DM1 heart (Wang et al. 2019), highlighting an important function of this gene, but also that the exact splice event differs between tissues even in the same disease. *ABCC9* is an ATP-dependent potassium channel in skeletal and cardiac tissue and thus plays important roles in innervation. In *ABCC9*, MAJIQ detects increased levels of exon 47 inclusion in DM1 (PSI-value +0.17), and exclusion in FSHD and EDMD (-0.34 and -0.22, respectively).

2.2 The nuclear envelope is affected in DM1, EDMD and FSHD

As shown in the publications (see Appendix), several nuclear envelope proteins are misregulated in DM1, EDMD and FSHD. Notably, the muscle-specific NETs *TMEM38A* and *PLPP7* are down-regulated in all of them: in EDMD myotubes ($\log_2FC = -0.9$ and -0.5) in DM1 (both -0.7) and FSHD muscle (-1.6 and -1.3 , respectively, **Figure 14**). Further, varying numbers of target genes (Robson et al. 2016) of both *TMEM38A* and *PLPP7* were found to be misregulated in all dystrophies. Similarly, *KLHL31*, which is also a muscle-specific NET and causes EDMD when mutated (Korfali et al. 2012), is downregulated in all three diseases. There are, however, also ubiquitously expressed NETs dysregulated and while some are similarly up- or downregulated in all three dystrophies, others seem to be disease specific or shared between only two dystrophies. This leads to an expression pattern of NETs as shown in **Figure 14**. The LINC complex (*SYNE* and *SUN* proteins) is distinctly misregulated in all three dystrophies: *SYNE1* is only upregulated in DM1, *SYNE2* is upregulated in EDMD but downregulated in DM1

and FSHD, while *SYNE3* is upregulated in all three diseases with the highest upregulation in FSHD. *SUN1* is only downregulated in FSHD and *SUN2* only in DM1 and EDMD.

POPDC2 is expressed mainly in cardiac and skeletal muscle, but its function in heart has been described in more detail: it is an important modulator of cardiac pacemaking (Froese et al. 2012). Here, *POPDC2* is transcriptionally upregulated only in DM1, but downregulated in FSHD and unchanged in EDMD. *TOR1AIP1* is a transmembrane protein of the inner nuclear membrane and interacts with various proteins, among others with lamins, emerin and torsinA (Mackels et al. 2023). It thus takes part in multiple functions related to genome organization and gene expression. Here, it is significantly downregulated in DM1 and FSHD muscle, but upregulated in EDMD myotubes. Further, the lamin *LMNA*, which has been described in the introduction, is upregulated in FSHD patients, but not in DM1 and EDMD.

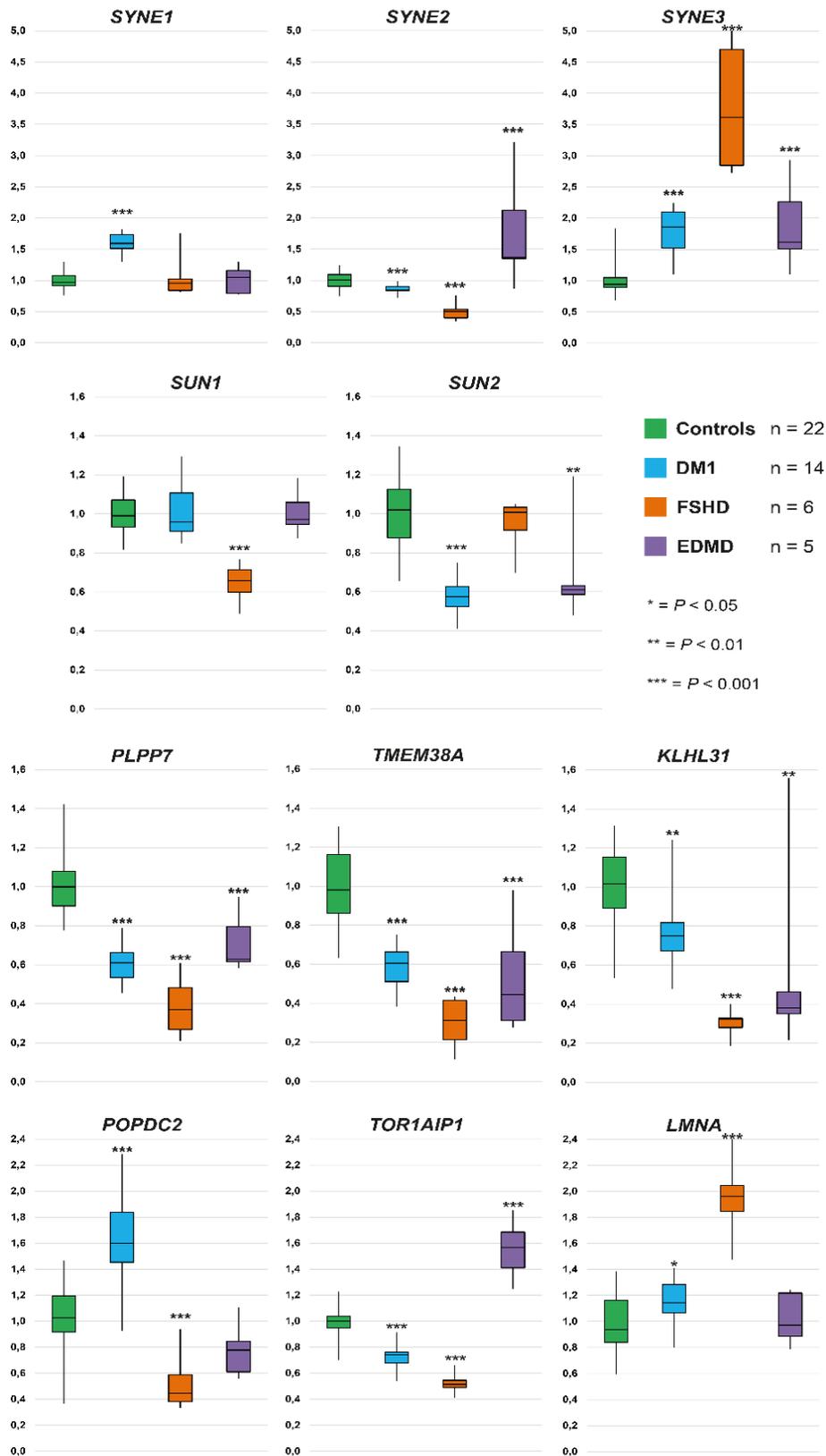


Figure 14: Expression pattern of nuclear envelope transmembrane proteins (NETs) in DM1, EDMD and FSHD. Number of replicates (n) and color code as indicated in the legend. SYNE and SUN proteins form the LINC complex which is involved in mechanotransduction (refer to Introduction 1.5.2.2). PLPP7, TMEM38A and KLHL31 are muscle specific, genome organizing NETs. POPDC2, TOR1AIP1 and LMNA are frequently mutated in EDMD and are genome organizing. Asterisks indicate significance.

3 DISCUSSION

3.1 Advances and limitations in splicing analysis

Big data and data science are the dictums of modern research. Advances in both methodology and computing power have unleashed a stream of massive data production that has changed science forever. Where we used to run specific experiments to answer specific questions, we now run general experiments that could answer a hundred questions – if only we had the expertise and time to ask them. For the publications and results in this thesis, I used a set of bioinformatic tools, that I want to evaluate regarding their power and easiness to use, before I will discuss the similarities and differences of the three neuromuscular disorders DM1, EDMD and FSHD.

Over the last decade, the price of RNA-sequencing (RNAseq) has dropped every year, and has now completely overtaken its predecessor, the microarray. This is clearly due to the following advantages: 1) RNAseq is not limited to transcript-specific probes and is therefore able to detect novel transcripts, 2) its detection rate of differentially expressed genes is significantly higher and 3) it allows for the analysis of isoform expression next to expression changes, whereas microarrays are designed specifically for one or the other (Zhao et al. 2014). Furthermore, reliability and sensitivity have increased significantly with next-generation sequencing methods, such as those deployed by Illumina, which use amplification of transcripts similar to PCR. For the analysis of expression changes at the gene level, there are currently no better options than RNAseq with a high hit rate and low false discovery rate (there are limitations with bioinformatic analyses, e.g. dealing with variability between replicates, etc.). The best available bioinformatics tool for RNA expression analysis is DESeq2, which uses a robust normalization function (*median of ratios*), deals fairly well with variability and is easy-to-use (R-based).

For splicing analyses, short-read RNAseq is more challenging. For classical “Illumina-based” sequencing, extracted RNA is fragmented (~200-500 bp) and then amplified and tagged during library preparation. However, most genes are much longer than this, meaning that a gene is cut in several pieces. With known amplification cycles, the original abundance of each gene can be more or less correctly inferred, but the assignment of a gene fragment to a specific *isoform* is much more difficult. To account for this, the sequencing depth is increased many times over: from around 10 to 20 million

reads to around 60 million reads or more. Nevertheless, bioinformatic tools have struggled to detect and quantify differences in splice site usage with exon usage being the easiest to detect: reads spanning junction-junction boundaries are counted and ratios can be calculated. A variety of bioinformatic tools have been developed to analyze splicing variation and all of them have their strengths and weaknesses, while none is perfect. This makes the use of several splicing analysis tools necessary. A combination of three tools have proven to be most useful: DEXSeq, MAJIQ and isoformSwitchAnalyzerR (ISA).

3.1.1 DEXSeq

DEXSeq (Anders et al. 2012) was written by the same bioinformaticians who released DESeq2 (Love et al. 2014) and focuses on exon usage, for which it yields robust results, and although it is not as simple as DESeq2, the majority of the code is still R-based, which is often the only language that biologists are trained in (and can be used on Windows, Mac and Linux). Reads aligned with STAR (Linux based (Dobin et al. 2013)) can be used, but for the inbuilt counting function of DEXSeq (which is python-based), the reads need to be the same length, so they must be trimmed prior to alignment. Notably, DEXSeq does not use PSI-values (percent spliced in) and delta PSI-values (differential use of exons between two groups) for quantification but logarithmic foldchanges (logFC) like in expression analyses, which most biologists are more familiar with. Other splicing events like intron retention, alternative splice sites (SS), mutually exclusive exons etc. are not covered by DEXSeq and must be addressed by other tools. While the neglect of alternative splice sites is not a problem per se, ignoring intron retention is more worrying: a retained intron is naturally flanked by two exons – which accordingly appear as “used” exons, while the actually interesting event is the intron retention. As a result, DEXseq will always show more differentially used exons than methods that look at different splicing events, such as MAJIQ and rMATS. Another disadvantage of DEXSeq is its in-built exon nomenclature, which is based on bins rather than annotated exons. Only its’ genomic coordinates reveal which exon is affected making third party tools such as Ensembl necessary. Finally, the use of logFCs compared to PSI-values may be advantageous for beginners but makes an interpretation difficult: A logFC of 2 for an exon in a disease condition vs. control condition gives no information of how often this exon is used in general.

3.1.2 MAJIQ

Similar to DESeq2 and DEXSeq, MAJIQ processes BAM-files aligned with STAR (Vaquero-Garcia et al. 2016). rMATS (Shen et al. 2014) and MAJIQ are both event-based tools, but MAJIQs' findings are more robust, and the true hit rate is higher. It can also calculate complex events with multiple event types, which is a result of its philosophy of looking at local splicing variations (LSVs), listing all the events detected in a particular region of a gene rather than focusing on event types. MAJIQ also comes with a visualization tool (voila) that displays splicing events in an easy-to-understand way, whereas rMATS relies on third-party tools. Another useful feature of MAJIQ is a built-in primer design tool specifically for the verification of LSVs by PCR. Notably, both MAJIQ and rMATS are Python-based and run from Linux command line, making basic Linux and Python proficiency a must. The main disadvantage of MAJIQ compared to rMATS is the structure of the output files: the same event can occur several times with different reference exons/introns, making filtering and quantification tedious. Furthermore, the nomenclature in these output files is not explained anywhere, so it is necessary to study them for some time. Notably, an alignment with STAR needs an annotation file, so MAJIQ is not annotation independent. It can still detect novel events, which are of course still annotation dependent.

3.1.3 ISA

Unlike DEXSeq and MAJIQ, ISA (Vitting-Seerup et al. 2019) cannot use STAR-generated BAM files, but requires alignment tools such as Kallisto and Salmon, which allow reconstruction and quantification of full-length transcripts from short-read RNA-seq data. ISA itself is an R package, which integrates several useful online tools: CPC2 for coding potential assessment (Kang et al. 2017), identification of protein domains using Pfam and disordered protein regions using IUPred2A, as well as the analysis of signal peptides through SignalP (Punta et al. 2011; Mészáros et al. 2018; Teufel et al. 2022). It also integrates in-R tools like DEXSeq and SpliceR (Vitting-Seerup et al. 2014). In the end, not only isoform switches are predicted but also their consequences: does the resulting isoform encode a protein or does it introduce a premature stop codon (PTC) leading to nonsense-mediated decay (NMD)? Is there a loss of a specific domain or localization signal? However, as mentioned above, it is computationally difficult to assign small reads to full isoforms and the more exons and isoforms a gene has, the more difficult

it is. Therefore, ISA naturally has a low number of significant results. In addition, isoforms must be annotated to be detected, which limits the power of ISA. It is also time-consuming to use the different online tools, that are not directly accessible from R, especially *Pfam* domain annotations, which can take many hours.

3.2 Comparison of DM1, EDMD and FSHD symptomology

In the results part, I examine two regulatory elements of tissue-specific gene expression: splicing and the nuclear envelope, and place them in the disease context of DM1, EDMD and FSHD. Before discussing these two aspects, I would like to recapitulate the phenotypic similarities between the diseases that define them all as muscular dystrophies: the degeneration and reorganization of skeletal muscle into fat and connective tissue, resulting in muscle weakness and wasting. Although the genetic causes are completely different and unrelated (apparently or actually), they start a cascade that leads to a very similar outcome. So what are the molecular pathways of muscular dystrophy? The synthesis of new proteins is reduced while aged proteins are lysed leading to smaller and weaker muscle fibers, a process which is referred to as muscle atrophy. Notably, the disuse of muscles leads to muscle atrophy as it can be observed in bed-ridden patients, but this non-dystrophic muscle has a functional regenerative potential and will recover when used again (Nunes et al. 2022), whereas dystrophic muscle is impaired in its regenerative potential. It is well known that fast-twitch fibers are more susceptible to atrophy than slow-twitch fibers because they have a much faster metabolism (Wang et al. 2013). How can we explain the predominance of slow-fiber atrophy in DM1 and EDMD mutation carrying EDMD patients? In contrast, FSHD patients tend to have fast fiber atrophy. The atrophy-induced “free space” is then occupied by either hypertrophied fibers and/or fat and connective tissue, the latter being a process commonly referred to as fibrosis. Fibrosis, which is also a hallmark of aging, is an excessive, unnecessary buildup of ECM tissue produced by fibroblasts and an activation of fibroadipogenic progenitors (FAPs), that inhibits muscle function and regeneration (Mahdy 2019). Activation of key signaling pathways such as *Wnt*, *MAPK* and *TNF* is thought to induce fibrosis and has indeed been described in most muscular dystrophies (de Las Heras et al. 2022; Todorow et al. 2021; Schätzl et al. 2021; De Paepe 2020; Guiraud et al. 2015). Although present in most muscular dystrophies, fibrosis is particularly dominant in FSHD compared to DM1 and EDMD. Importantly, muscular dystrophy can involve not only fiber atrophy but also fiber necrosis, which further contributes to muscle wasting, and in the case of the

dystrophies discussed here, FSHD shows higher levels of necrosis, whereas it is rare in DM1 and EDMD.

Another important aspect to bear in mind is the high inter-patient variability of all three diseases, an intriguing but tricky similarity: they are closest in how far apart the individuals are. The conclusion that can be drawn from this is that the etiology of all three dystrophies is highly complex and probably modulated by various factors, internal and external ones such as genetics, epigenetics, metabolism, diet and exercise, to name just a few that come to mind. It is therefore difficult to reduce the pathomechanism behind muscular dystrophies to their genetic cause and indeed, there are still unresolved questions, including why some tissues are affected while others are spared, why muscular dystrophies are progressive and the age of onset is so variable, or why some muscle groups are more affected than others. What is clear, however, is that muscular dystrophy results from the same molecular pathways, whatever the genetic cause. As always, it is necessary to link cause and effect in an unbroken chain to elucidate each dystrophy.

3.3 Splicing in DM1, EDMD and FSHD

Although DM1 is seen as the splicing disease among the muscular dystrophies, I could show that splicing is also affected in EDMD myotubes and FSHD muscle, with many splicing factors misregulated and many genes mis-spliced, which had not been shown or investigated in detail before. Interestingly, many mis-spliced genes and even specific events are shared by at least two diseases, like the NMJ architectural genes *GFPT1* and *MACF1*, while others seem to be disease specific. Of note, myotubes and mature muscle are only partially comparable, and differences between EDMD and the other two diseases can be developmental stage dependent rather than disease driven. Further, genes that are unique to a disease are often involved in commonly misregulated molecular pathways, many of which contribute to the common dystrophic outcome as described above: protein synthesis and degradation, signaling, muscle development and contraction.

The expression patterns of spliceosomal genes for each dystrophy are intriguing: DM1 displays up- and downregulation of splicing factors, while most splicing related genes are downregulated in EDMD and nearly all in FSHD (**Figure 10**). However, downregulation is more severe in DM1 compared to FSHD and EDMD. In this regard, I presume that a mild misregulation of many splicing factors still disrupts the necessary stoichiometry for correct splicing and inhibiting spliceosome assembly and splice site

detection. This would then lead to broader effects than a strong misregulation of one (alternative) splicing factor, but with milder isoform and splicing event shifts (smaller deltaPSI-values).

MBNL1 is upregulated in EDMD, and downregulated in FSHD and DM1. The balance between MBNL1 and CELF1 is highly important in muscle development and *MBNL1* is about 4-fold upregulated in adult muscle compared to embryonic or developing muscle. Accordingly, it was shown that MBNL1 loss-of-function and CELF1 gain-of-function in DM1 leads to embryonic splice patterns. If *MBNL1* upregulation on the transcript level in EDMD is reflected on the protein level, would we see a splicing pattern in EDMD myotubes that resembles mature muscle? I will revisit this topic soon, when looking at specific splice events. Another commonly downregulated splicing factor is *RBM20*, which regulates the splicing of *TTN* and causes dilated cardiomyopathy if mutated. *RBM20* expression differs between muscle-types which is proposed to be causative for different splicing isoforms of *TTN* (Maimaiti et al. 2021). Importantly, *TTN* is heavily mis-spliced in all three muscular dystrophies. Another splicing factor highly upregulated in all three diseases is *NOVA1*, which is described as a neuron-specific splicing factor contributing to correct splicing of the neuronal isoform (Z+ isoform) of *AGRN* which is pivotal for NMJ formation. As our data shows, it is also expressed in skeletal muscle, its exact role however remains to be elucidated.

Strikingly, there are by far more alternative exons used in DM1 compared to EDMD and especially FSHD (**Figure 11**). In contrast, EDMD and FSHD show an increase of exon skipping. Both observations fit to what we see on the splicing factor expression level: increased expression should lead to increased splice site usage (DM1), while decreased expression should result in incomplete spliceosome assembly and thus less splicing in general (EDMD and FSHD). I expect that the trends of up- and downregulation at the transcript level is at least partially reflected at the protein level and that the degree of activity of the splicing factors lead in the same direction. This hypothesis is supported by the fact that FSHD and EDMD have a higher overlap of both mis-spliced genes and splicing events than DM1 has with either of the two.

In my comparative splicing analysis, I showed that DM1 shares 138 LSVs with EDMD and FSHD (**Figure 12**). Some of these events lead to the same splicing outcome in all dystrophies like *GFPT1* and *LMO7*, while others differ in the outcome between DM1 and the other two. Assuming that splicing alteration is a secondary effect in EDMD and FSHD, the data suggests that shared events between DM1 and other non-myotonic

MDs are also secondary events in DM1 due to general dystrophic changes. Importantly, it was shown before, that many “DM1-specific” splice events are shared between DM1 and Duchenne muscular dystrophy (DMD) (Bachinski et al. 2014), among others *MBNL1*, *ATP2A1*, *TTN*, *TNNT2* and *MEF2A/C*. It will be necessary to find more of these shared events to enable distinction between primary and secondary mis-splicing in DM1. Notably, a mis-splicing event that is secondary in other muscular dystrophies could still be primary in DM1.

Interestingly, many genes display splicing variations at the same locus in several dystrophies but while DM1 has a tendency of including exons and introns, both are rather spliced out in FSHD and EDMD, which fits the general expression of splicing factors, as pointed out above (for example see **Figure 13**, events ABCC9, MBNL1 and BIN1). Why are the same genes and loci affected, although different splicing factor expression patterns exist between DM1 and non-DM dystrophies? It is conceivable that additional mechanisms are involved, for example the expression levels of the gene or the accessibility of certain loci. A hint for this is the high number of sarcomeric proteins that are mis-spliced, like *NEB* and *MYH* isoforms: they are highly expressed and highly spliced, which increases the chance of being affected by differentially expressed splicing factors. In other words, a gene that is not expressed cannot be mis-spliced. Additionally, it has been shown that the speed of transcription through PolIII has a great effect on splice site decision, increasing or decreasing the time frame for spliceosome assembly. I thus hypothesize that the splicing outcome depends on splicing factor misregulation, but the loci itself is regulated by third party mechanisms. If this is the case it is also not surprising that the resulting phenotype is so similar: in which way you diverge from the “healthy” isoform stoichiometry, through inclusion or exclusion of alternative events, both lead to altered isoform levels and contribute to the misregulation of the same genes and thus pathways. In this regard, it is also interesting to check which events are present in DM1 but not in other muscular dystrophies. *CLCN1* exon 7a inclusion is one of the most studied mis-splicing events in DM1 and directly linked to myotonia – a symptom that is highly characteristic for myotonic dystrophy and absent in other MDs. Intriguingly, *CLCN1* is not mis-spliced in EDMD or FSHD, nor in Duchenne muscular dystrophy as shown by (Bachinski et al. 2014) who, similar to my findings, reported a high overlap of mis-splicing events in DM and non-DM diseases – except *CLCN1* and a few other events. Further genes uniquely mis-spliced in DM1 in my data include *MYOM1* and *CACNAIS*. The splicing of these genes seems to be specifically dysregulated in DM1, most likely

due to RBP-sequestration, but not due to general (secondary) splicing alterations found in non-DM. Further comparisons with other MDs are necessary to support this theory.

I mentioned above that *MBNL1* is transcriptionally upregulated in EDMD and proposed that this would lead to mature muscle-like splicing patterns. Splicing of *MBNL1* at exon 4 and *SYNE1* exon DV23 clearly lead in this direction, showing that DM1 muscle splicing resembles that of myotubes, while EDMD myotube splicing changes towards mature muscle splicing. Embryonic splice patterns in DM1 adult muscle are well established. For EDMD, a pre-mature splicing pattern has never been described, but importantly, EDMD shows progeroid tendencies and mutations in *LMNA* can lead to Hutchinson-Gilford-Progeria-syndrome, which is characterized by premature aging (Eriksson et al. 2003; De Sandre-Giovannoli et al. 2003). To shed light on this matter it is necessary to investigate splicing in EDMD mature muscle and healthy aged muscle.

If dystrophic, non-myotonic muscle (EDMD, FSHD, DMD) results in similar alternative splicing events as in DM1, how do we know that events we see – and target - in DM1 are due to toxic RNA (and thus primary) and not secondary due to muscle damage? How early is splicing affected in EDMD and FSHD, or in other words: is mis-splicing the cause or result of atrophy/dystrophy? Or just an early contributor? If it is “just” a contributor but not sufficient to produce strong dystrophic effects (maybe as in DM2?), how can we explain the strong misregulation of signaling and transcription in DM1? Is there a shared cascade of events in muscular dystrophies that is set off at different spatial and temporal points but leads to the same downstream effects? With the here investigated material and methods, we cannot answer these questions at this point.

3.4 The nuclear envelope is affected in DM1, EDMD and FSHD

The NE is obviously altered in EDMD since genes of the NE are mutated. It is interesting however, that NETs other than the mutated one are misregulated in EDMD myotubes, for example *TMEM38A*, *PLPP7* and *KLHL31* were all downregulated in group 1 (gp1) and target genes of *TMEM38A* and *PLPP7* were misregulated. This suggests a reciprocal regulation of NETs. Notably, the same muscle-specific NETs are also downregulated in DM1 (which we validated on the protein level) and FSHD and again, many of the target genes were found among the misregulated genes. As mentioned in the introduction, *TMEM38A* and *PLPP7* are muscle-specific NETs with crucial functions in muscle gene expression and development. NETs interact with each other as well as chromatin regions

and thereby regulate the expression of genes, and accordingly, tissue-specific NETs have been shown to regulate the expression of genes relevant for their tissue. For example, Robson et al. demonstrated the regulation of the pleiotrophin encoding gene *PTN* through PLPP7: *PTN* is repositioned to the NE during myogenesis and through PLPP7 overexpression independent of myogenesis, while a PLPP7 knockdown inhibited the repositioning (Robson et al. 2016). *PTN* is an important gene of the NMJ, which is altered in DM1 and FSHD. It is thus likely, that a downregulation of PLPP7 and TMEM38A results in genome organization alterations, also influencing the expression of non-target genes.

Non-muscle NETs like the LINC complex, but also lamins and lamin-associated proteins are misregulated in all three dystrophies, highlighting the importance of the NE in muscle identity and disease. As FSHD and EDMD seem to be secondary splicing diseases, DM1 and FSHD appear to be secondary envelopopathies. *SYNE1* has an especially interesting role: it is one of the main players in myonuclear positioning at the NMJ. The muscle-specific exon DV23 that is skipped in DM1 and healthy myotubes in 50% of the transcripts might encode for the *SYNE1* isoform important for anchoring the nuclei, however, the function of this exon containing isoform is still unknown. It is also intriguing that components of the LINC complex are variably dysregulated between the three dystrophies, potentially resulting in the preferential formation of certain LINC complexes and thus, to distinct patterns of gene expression and mechano-signaling. For example, *SYNE1* is upregulated and *SUN2* downregulated in DM1, while *SUN1* expression is unchanged. This would result in a higher abundance of the *SYNE1/SUN1* containing LINC complex compared to control. Different LINC complexes probably have different functions, but these have never been investigated.

Notably, for DM1 and FSHD, we used additional translational disease information about disease severity (dorsiflexion strength and MRI, respectively) and could show that the misregulation of NETs is correlated with severity (Schätzl et al. and (Todorow et al. 2022)). Together with their dystrophy-specific expression pattern, many of these NETs are predestined to serve as biomarkers for disease progression or even for differential diagnosis: *SYNE1* upregulation for instance, would indicate DM1; *TOR1AIP1* is only upregulated in EDMD but downregulated in DM1 and FSHD, while *LMNA* is mostly upregulated in FSHD. Providing a set of biomarkers that is misregulated not in one but several muscular dystrophies could be helpful for clinicians treating various muscle disorders as tests could be standardized independent of the individual patient and

diagnosis. It would be necessary to explore the expression of NETs in other dystrophies like Duchenne muscular dystrophy (DMD) and limb-girdle-muscular-dystrophy (LGMD). Given the high importance of the NE in muscle-specific gene regulation, it is neither surprising that it is affected in muscle diseases nor is it inconceivable that its misregulation contributes to the downstream effects on gene expression and signaling we observe. But as for splicing, it will be crucial to reveal the mechanistic cause and the timepoint it becomes affected for each disease.

3.5 Final remarks and future directions

This dissertation demonstrates that DM1, EDMD and FSHD do not share a set of symptoms by chance, but because some overlapping cellular pathways are affected. With completely different genetic causes, how do the same events occur? Gene expression, signaling, adhesion, ECM and fibrosis, proliferation, differentiation, muscle contraction, innervation, regeneration, splicing and genes of the nuclear envelope (we could add the last two to the list) – all these molecular pathways are affected in each of the three dystrophies. But where is the starting point for each? Is it the splicing in case of DM1 and the nuclear envelope in EDMD, gene expression through DUX4 in FSHD? Is a misregulation of every single pathway sufficient to induce an imbalance in all others? This would suggest a very strong interconnection and interdependence of all these pathways, comparable to a tightly woven net: no matter where you manipulate the net, you move it as a whole with regions closest (spatial and temporal) to the locus of manipulation showing the strongest movement. Or are we mistaken and the initial events leading to this cascade are much more similar between DM1, EDMD and FSHD than anticipated? DM1 and FSHD are caused by genetic alterations that are very well conceivable to mis-regulate the structure and integrity of the nuclear envelope (which indeed *is* altered) and the NE itself *would* be sufficient to start the above-mentioned cascade as we see in EDMD. On the other hand, is splicing misregulation sufficient to start the cascade? If yes, why is DM2 the much milder disease compared to DM1 although the repeats in the *CNBP* gene are much longer than in *DMPK* (and are verifiably bound by the same splicing factors)? Muscular dystrophies are highly complex and variable and through the combined force of many scientists, a great portion of all pieces of this jigsaw have been identified. But how these pieces fall into place, which piece connects the genetic cause (the only certain fact) with all other pieces, we still cannot say. For all I know, every piece in the jigsaw below (**Figure 15**) could be shuffled around, come up earlier or later in the cascade, influencing, rivaling and deflating each other, while additional pieces could join or replace others, probably individualized for each patient. Similar comparisons with DMD, ALS and especially DM2 will provide more insights.

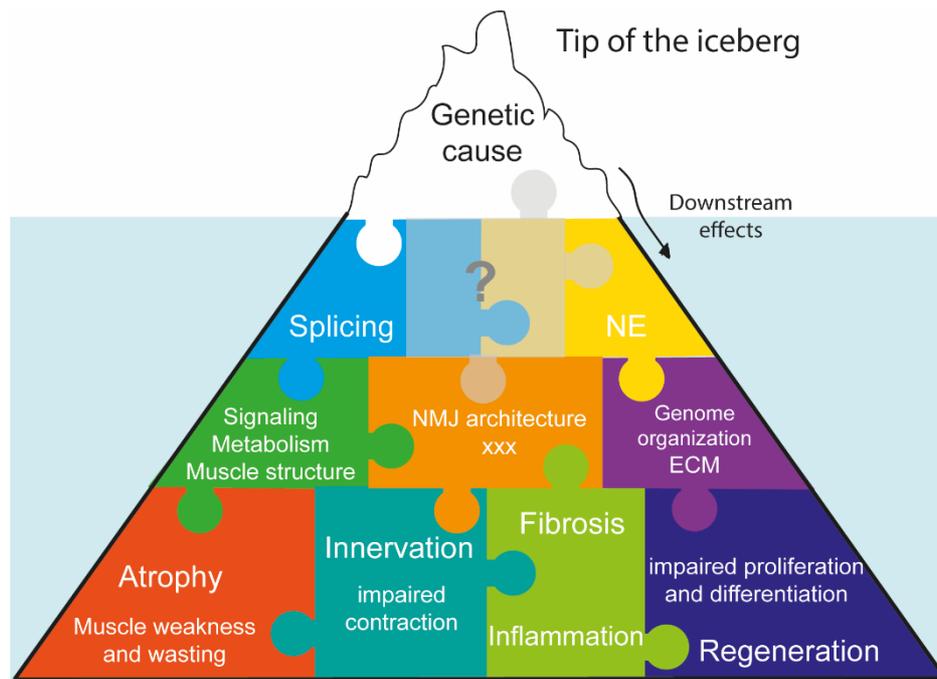


Figure 15: What we know and do not know about muscular dystrophies. The genetic causes are described for DM1, EDMD and FSHD – but this is only the tip of the iceberg. Various pieces of this jigsaw have been identified but how are they connected, when do they fall into place and how exactly does the genetic variation lead to the downstream effects we observe?

Disease research is naturally conducted in the light of possible therapeutic approaches. The comparison of different but similar diseases has thus great potential: we might find a commonality between many dystrophies that can be treated with the same drugs. Each muscular dystrophy is a rare genetic disorder – but taken together, millions of people are suffering from them. For helping them, we need to think out-of-the-box, inclusive and innovative instead of focusing on only one aspect.

4 REFERENCES

- Alam, S. G., D. Lovett, D. I. Kim, K. J. Roux, R. B. Dickinson, and T. P. Lele. 2015. 'The nucleus is an intracellular propagator of tensile forces in NIH 3T3 fibroblasts', *J Cell Sci*, 128: 1901-11.
- Anders, S., A. Reyes, and W. Huber. 2012. 'Detecting differential usage of exons from RNA-seq data', *Genome Res*, 22: 2008-17.
- Astejada, M. N., K. Goto, A. Nagano, S. Ura, S. Noguchi, I. Nonaka, I. Nishino, and Y. K. Hayashi. 2007. 'Emerinopathy and laminopathy clinical, pathological and molecular features of muscular dystrophy with nuclear envelopathy in Japan', *Acta Myol*, 26: 159-64.
- Attarian, S., E. Salort-Campana, K. Nguyen, A. Behin, and J. Andoni Urtizberea. 2012. 'Recommendations for the management of facioscapulohumeral muscular dystrophy in 2011', *Rev Neurol (Paris)*, 168: 910-8.
- Bachinski, L. L., K. A. Baggerly, V. L. Neubauer, T. J. Nixon, O. Raheem, M. Sirito, A. K. Unruh, J. Zhang, L. Nagarajan, L. T. Timchenko, G. Bassez, B. Eymard, J. Gamez, T. Ashizawa, J. R. Mendell, B. Udd, and R. Krahe. 2014. 'Most expression and splicing changes in myotonic dystrophy type 1 and type 2 skeletal muscle are shared with other muscular dystrophies', *Neuromuscul Disord*, 24: 227-40.
- Banerji, C. R. S., and P. S. Zammit. 2021. 'Pathomechanisms and biomarkers in facioscapulohumeral muscular dystrophy: roles of DUX4 and PAX7', *EMBO Mol Med*, 13: e13695.
- Baralle, Francisco E., and Jimena Giudice. 2017. 'Alternative splicing as a regulator of development and tissue identity', *Nature Reviews Molecular Cell Biology*, 18: 437.
- Barbé, L., S. Lanni, A. López-Castel, S. Franck, C. Spits, K. Keymolen, S. Seneca, S. Tomé, I. Miron, J. Letourneau, M. Liang, S. Choufani, R. Weksberg, M. D. Wilson, Z. Sedlacek, C. Gagnon, Z. Musova, D. Chitayat, P. Shannon, J. Mathieu, K. Sermon, and C. E. Pearson. 2017. 'CpG Methylation, a Parent-of-Origin Effect for Maternal-Biased Transmission of Congenital Myotonic Dystrophy', *Am J Hum Genet*, 100: 488-505.
- Berget, S. M., C. Moore, and P. A. Sharp. 1977. 'Spliced segments at the 5' terminus of adenovirus 2 late mRNA', *Proc Natl Acad Sci U S A*, 74: 3171-5.
- Berk, Arnold J. 2016. 'Discovery of RNA splicing and genes in pieces', *Proceedings of the National Academy of Sciences*, 113: 801-05.
- Bione, S., E. Maestrini, S. Rivella, M. Mancini, S. Regis, G. Romeo, and D. Toniolo. 1994. 'Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy', *Nat Genet*, 8: 323-7.
- Blencowe, B. J. 2000. 'Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases', *Trends Biochem Sci*, 25: 106-10.
- Bonne G, Leturcq F, Ben Yaou R. 2004 *Emery-Dreifuss Muscular Dystrophy*. (University of Washington: Seattle).
- Bonne, G., M. R. Di Barletta, S. Varnous, H. M. Bécane, E. H. Hammouda, L. Merlini, F. Muntoni, C. R. Greenberg, F. Gary, J. A. Urtizberea, D. Duboc, M. Fardeau, D. Toniolo, and K. Schwartz. 1999. 'Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy', *Nat Genet*, 21: 285-8.
- Bonne, G., E. Mercuri, A. Muchir, A. Urtizberea, H. M. Bécane, D. Recan, L. Merlini, M. Wehnert, R. Boor, U. Reuner, M. Vorgerd, E. M. Wicklein, B. Eymard, D. Duboc, I.

- Penisson-Besnier, J. M. Cuisset, X. Ferrer, I. Desguerre, D. Lacombe, K. Bushby, C. Pollitt, D. Toniolo, M. Fardeau, K. Schwartz, and F. Muntoni. 2000. 'Clinical and molecular genetic spectrum of autosomal dominant Emery-Dreifuss muscular dystrophy due to mutations of the lamin A/C gene', *Ann Neurol*, 48: 170-80.
- Bosnakovski, D., Z. Xu, E. J. Gang, C. L. Galindo, M. Liu, T. Simsek, H. R. Garner, S. Agha-Mohammadi, A. Tassin, F. Coppée, A. Belayew, R. R. Perlingeiro, and M. Kyba. 2008. 'An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies', *Embo j*, 27: 2766-79.
- Breathnach, R., J. L. Mandel, and P. Chambon. 1977. 'Ovalbumin gene is split in chicken DNA', *Nature*, 270: 314-19.
- Brenner, B., and E. Eisenberg. 1987. 'The mechanism of muscle contraction. Biochemical, mechanical, and structural approaches to elucidate cross-bridge action in muscle', *Basic Res Cardiol*, 82 Suppl 2: 3-16.
- Broers, J. L., F. C. Ramaekers, G. Bonne, R. B. Yaou, and C. J. Hutchison. 2006. 'Nuclear lamins: laminopathies and their role in premature ageing', *Physiol Rev*, 86: 967-1008.
- Buckingham, Margaret. 2017. 'Gene regulatory networks and cell lineages that underlie the formation of skeletal muscle', *Proceedings of the National Academy of Sciences*, 114: 5830.
- Buckley, L., M. Lacey, and M. Ehrlich. 2016. 'Epigenetics of the myotonic dystrophy-associated DMPK gene neighborhood', *Epigenomics*, 8: 13-31.
- Carrell, S. T., E. M. Carrell, D. Auerbach, S. K. Pandey, C. F. Bennett, R. T. Dirksen, and C. A. Thornton. 2016. 'Dmpk gene deletion or antisense knockdown does not compromise cardiac or skeletal muscle function in mice', *Hum Mol Genet*, 25: 4328-38.
- Cartegni, L., S. L. Chew, and A. R. Krainer. 2002. 'Listening to silence and understanding nonsense: exonic mutations that affect splicing', *Nat Rev Genet*, 3: 285-98.
- Castle, J. C., C. Zhang, J. K. Shah, A. V. Kulkarni, A. Kalsotra, T. A. Cooper, and J. M. Johnson. 2008. 'Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines', *Nat Genet*, 40: 1416-25.
- Charlet, B. N., R. S. Savkur, G. Singh, A. V. Philips, E. A. Grice, and T. A. Cooper. 2002. 'Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing', *Mol Cell*, 10: 45-53.
- Chow, L. T., R. E. Gelinis, T. R. Broker, and R. J. Roberts. 1977. 'An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA', *Cell*, 12: 1-8.
- Dahlqvist, J. R., M. C. Ørngreen, N. Witting, and J. Vissing. 2015. 'Endocrine function over time in patients with myotonic dystrophy type 1', *Eur J Neurol*, 22: 116-22.
- Darnell, J. E., L. Philipson, R. Wall, and M. Adesnik. 1971. 'Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA', *Science*, 174: 507-10.
- De Antonio, M., C. Dogan, D. Hamroun, M. Mati, S. Zerrouki, B. Eymard, S. Katsahian, and G. Bassez. 2016. 'Unravelling the myotonic dystrophy type 1 clinical spectrum: A systematic registry-based study with implications for disease classification', *Rev Neurol (Paris)*, 172: 572-80.
- De Conti, L., M. Baralle, and E. Buratti. 2013. 'Exon and intron definition in pre-mRNA splicing', *Wiley Interdiscip Rev RNA*, 4: 49-60.
- de Las Heras, J. I., V. Todorow, L. Krečinić-Balić, S. Hintze, R. Czapiewski, S. Webb, B. Schoser, P. Meinke, and E. C. Schirmer. 2022. 'Metabolic, fibrotic, and splicing

- pathways are all altered in Emery-Dreifuss muscular dystrophy Spectrum patients to differing degrees', *Hum Mol Genet*.
- de Las Heras, J. I., N. Zuleger, D. G. Batrakou, R. Czapiewski, A. R. Kerr, and E. C. Schirmer. 2017. 'Tissue-specific NETs alter genome organization and regulation even in a heterologous system', *Nucleus*, 8: 81-97.
- De Paepe, Boel. 2020. 'Progressive Skeletal Muscle Atrophy in Muscular Dystrophies: A Role for Toll-Like Receptor-Signaling in Disease Pathogenesis', *International Journal of Molecular Sciences*, 21: 4440.
- De Sandre-Giovannoli, A., R. Bernard, P. Cau, C. Navarro, J. Amiel, I. Boccaccio, S. Lyonnet, C. L. Stewart, A. Munnich, M. Le Merrer, and N. Lévy. 2003. 'Lamin a truncation in Hutchinson-Gilford progeria', *Science*, 300: 2055.
- Deak, K. L., R. J. Lemmers, J. M. Stajich, R. Klooster, R. Tawil, R. R. Frants, M. C. Speer, S. M. van der Maarel, and J. R. Gilbert. 2007. 'Genotype-phenotype study in an FSHD family with a proximal deletion encompassing p13E-11 and D4Z4', *Neurology*, 68: 578-82.
- Disset, A., C. F. Bourgeois, N. Benmalek, M. Claustres, J. Stevenin, and S. Tuffery-Giraud. 2006. 'An exon skipping-associated nonsense mutation in the dystrophin gene uncovers a complex interplay between multiple antagonistic splicing elements', *Hum Mol Genet*, 15: 999-1013.
- Dixit, M., E. Anseau, A. Tassin, S. Winokur, R. Shi, H. Qian, S. Sauvage, C. Mattéotti, A. M. van Acker, O. Leo, D. Figlewicz, M. Barro, D. Laoudj-Chenivresse, A. Belayew, F. Coppée, and Y. W. Chen. 2007. 'DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1', *Proc Natl Acad Sci U S A*, 104: 18157-62.
- Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. 2012. 'Topological domains in mammalian genomes identified by analysis of chromatin interactions', *Nature*, 485: 376-80.
- Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, 29: 15-21.
- Douglas, Andrew G. L., and Matthew J. A. Wood. 2011. 'RNA splicing: disease and therapy', *Briefings in Functional Genomics*, 10: 151-64.
- Echenne, B., and G. Bassez. 2013. 'Congenital and infantile myotonic dystrophy', *Handb Clin Neurol*, 113: 1387-93.
- Emery, A. E. 1989. 'Emery-Dreifuss syndrome', *J Med Genet*, 26: 637-41.
- Emery, A. E., and F. E. Dreifuss. 1966. 'Unusual type of benign x-linked muscular dystrophy', *J Neurol Neurosurg Psychiatry*, 29: 338-42.
- Eriksson, M., W. T. Brown, L. B. Gordon, M. W. Glynn, J. Singer, L. Scott, M. R. Erdos, C. M. Robbins, T. Y. Moses, P. Berglund, A. Dutra, E. Pak, S. Durkin, A. B. Csoka, M. Boehnke, T. W. Glover, and F. S. Collins. 2003. 'Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome', *Nature*, 423: 293-8.
- Fawcett, D. W. 1966. 'On the occurrence of a fibrous lamina on the inner aspect of the nuclear envelope in certain cells of vertebrates', *Am J Anat*, 119: 129-45.
- Filipowicz, W., M. Konarska, H. J. Gross, and A. J. Shatkin. 1983. 'RNA 3'-terminal phosphate cyclase activity and RNA ligation in HeLa cell extract', *Nucleic acids research*, 11: 1405-18.

- Fox-Walsh, Kristi L., Yimeng Dou, Bianca J. Lam, She-pin Hung, Pierre F. Baldi, and Klemens J. Hertel. 2005. 'The architecture of pre-mRNAs affects mechanisms of splice-site pairing', *Proceedings of the National Academy of Sciences*, 102: 16176-81.
- Froese, A., S. S. Breher, C. Waldeyer, R. F. Schindler, V. O. Nikolaev, S. Rinné, E. Wischmeyer, J. Schlueter, J. Becher, S. Simrick, F. Vauti, J. Kuhtz, P. Meister, S. Kreissl, A. Torlopp, S. K. Liebig, S. Laakmann, T. D. Müller, J. Neumann, J. Stieber, A. Ludwig, S. K. Maier, N. Decher, H. H. Arnold, P. Kirchhof, L. Fabritz, and T. Brand. 2012. 'Popeye domain containing proteins are essential for stress-mediated modulation of cardiac pacemaking in mice', *J Clin Invest*, 122: 1119-30.
- Fukushige, T., T. M. Brodigan, L. A. Schriefer, R. H. Waterston, and M. Krause. 2006. 'Defining the transcriptional redundancy of early bodywall muscle development in *C. elegans*: evidence for a unified theory of animal muscle development', *Genes Dev*, 20: 3395-406.
- Gabellini, D., M. R. Green, and R. Tupler. 2002. 'Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle', *Cell*, 110: 339-48.
- Gaillard, M. C., N. Broucqsault, J. Morere, C. Laberthonnière, C. Dion, C. Badja, S. Roche, K. Nguyen, F. Magdinier, and J. D. Robin. 2019. 'Analysis of the 4q35 chromatin organization reveals distinct long-range interactions in patients affected with Facio-Scapulo-Humeral Dystrophy', *Sci Rep*, 9: 10327.
- Geng, L. N., Z. Yao, L. Snider, A. P. Fong, J. N. Cech, J. M. Young, S. M. van der Maarel, W. L. Ruzzo, R. C. Gentleman, R. Tawil, and S. J. Tapscott. 2012. 'DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy', *Dev Cell*, 22: 38-51.
- Gilbert, W. 1978. 'Why genes in pieces?', *Nature*, 271: 501.
- Goossens, R., M. L. van den Boogaard, Rjlf Lemmers, J. Balog, P. J. van der Vliet, I. M. Willemsen, J. Schouten, I. Maggio, N. van der Stoep, R. C. Hoeben, S. J. Tapscott, N. Geijsen, Mafv Gonçalves, S. Sacconi, R. Tawil, and S. M. van der Maarel. 2019. 'Intronic SMCHD1 variants in FSHD: testing the potential for CRISPR-Cas9 genome editing', *J Med Genet*, 56: 828-37.
- Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van Steensel. 2008. 'Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions', *Nature*, 453: 948-51.
- Gueneau, L., A. T. Bertrand, J. P. Jais, M. A. Salih, T. Stojkovic, M. Wehnert, M. Hoeltzenbein, S. Spuler, S. Saitoh, A. Verschueren, C. Tranchant, M. Beuvin, E. Lacene, N. B. Romero, S. Heath, D. Zelenika, T. Voit, B. Eymard, R. Ben Yaou, and G. Bonne. 2009. 'Mutations of the FHL1 gene cause Emery-Dreifuss muscular dystrophy', *Am J Hum Genet*, 85: 338-53.
- Guiraud, Simon, Annemieke Aartsma-Rus, Natassia M. Vieira, Kay E. Davies, Gert-Jan B. van Ommen, and Louis M. Kunkel. 2015. 'The Pathogenesis and Therapy of Muscular Dystrophies', *Annual Review of Genomics and Human Genetics*, 16: 281-308.
- Guo, W., S. Schafer, M. L. Greaser, M. H. Radke, M. Liss, T. Govindarajan, H. Maatz, H. Schulz, S. Li, A. M. Parrish, V. Dauksaite, P. Vakeel, S. Klaassen, B. Gerull, L. Thierfelder, V. Regitz-Zagrosek, T. A. Hacker, K. W. Saupe, G. W. Dec, P. T. Ellinor,

- C. A. MacRae, B. Spallek, R. Fischer, A. Perrot, C. Özcelik, K. Saar, N. Hubner, and M. Gotthardt. 2012. 'RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing', *Nat Med*, 18: 766-73.
- Hamel, J., and R. Tawil. 2018. 'Facioscapulohumeral Muscular Dystrophy: Update on Pathogenesis and Future Treatments', *Neurotherapeutics*, 15: 863-71.
- Haque, F., D. J. Lloyd, D. T. Smallwood, C. L. Dent, C. M. Shanahan, A. M. Fry, R. C. Trembath, and S. Shackleton. 2006. 'SUN1 interacts with nuclear lamin A and cytoplasmic nesprins to provide a physical connection between the nuclear lamina and the cytoskeleton', *Mol Cell Biol*, 26: 3738-51.
- Harmon, E. B., M. L. Harmon, T. D. Larsen, A. F. Paulson, and M. B. Perryman. 2008. 'Myotonic dystrophy protein kinase is expressed in embryonic myocytes and is required for myotube formation', *Dev Dyn*, 237: 2353-66.
- Harmon, E. B., M. L. Harmon, T. D. Larsen, J. Yang, J. W. Glasford, and M. B. Perryman. 2011. 'Myotonic dystrophy protein kinase is critical for nuclear envelope integrity', *J Biol Chem*, 286: 40296-306.
- Harris, H., and J. W. Watts. 1962. 'The relationship between nuclear and cytoplasmic ribonucleic acid', *Proc R Soc Lond B Biol Sci*, 156: 109-21.
- Heller, S. A., R. Shih, R. Kalra, and P. B. Kang. 2020. 'Emery-Dreifuss muscular dystrophy', *Muscle Nerve*, 61: 436-48.
- Hetzer, M. W., T. C. Walther, and I. W. Mattaj. 2005. 'Pushing the envelope: structure, function, and dynamics of the nuclear periphery', *Annu Rev Cell Dev Biol*, 21: 347-80.
- Hicks, M. J., B. J. Lam, and K. J. Hertel. 2005. 'Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays', *Methods*, 37: 306-13.
- Himeda, C. L., T. I. Jones, C. M. Virbasius, L. J. Zhu, M. R. Green, and P. L. Jones. 2018. 'Identification of Epigenetic Regulators of DUX4-fl for Targeted Therapy of Facioscapulohumeral Muscular Dystrophy', *Mol Ther*, 26: 1797-807.
- Höger, T. H., K. Zatloukal, I. Waizenegger, and G. Krohne. 1990. 'Characterization of a second highly conserved B-type lamin present in cells previously thought to contain only a single B-type lamin', *Chromosoma*, 99: 379-90.
- Holaska, James M., Soroush Rais-Bahrami, and Katherine L. Wilson. 2006. 'Lmo7 is an emerin-binding protein that regulates the transcription of emerin and many other muscle-relevant genes', *Hum Mol Genet*, 15: 3459-72.
- Horrigan, J., T. B. Gomes, M. Snape, N. Nikolenko, A. McMorn, S. Evans, A. Yaroshinsky, O. Della Pasqua, S. Oosterholt, and H. Lochmüller. 2020. 'A Phase 2 Study of AMO-02 (Tideglusib) in Congenital and Childhood-Onset Myotonic Dystrophy Type 1 (DM1)', *Pediatr Neurol*, 112: 84-93.
- Howe, K. J. 2002. 'RNA polymerase II conducts a symphony of pre-mRNA processing activities', *Biochim Biophys Acta*, 1577: 308-24.
- Huxley, A. F., and R. Niedergerke. 1954. 'Structural changes in muscle during contraction; interference microscopy of living muscle fibres', *Nature*, 173: 971-3.
- Iovino, M., U. Pfisterer, J. L. Holton, T. Lashley, R. J. Swingler, L. Calo, R. Treacy, T. Revesz, M. Parmar, M. Goedert, M. M. Muqit, and M. G. Spillantini. 2014. 'The novel MAPT mutation K298E: mechanisms of mutant tau toxicity, brain pathology and tau expression in induced fibroblast-derived neurons', *Acta Neuropathol*, 127: 283-95.

- Johnson, N. E., R. J. Butterfield, K. Mayne, T. Newcomb, C. Imburgia, D. Dunn, B. Duval, M. L. Feldkamp, and R. B. Weiss. 2021. 'Population-Based Prevalence of Myotonic Dystrophy Type 1 Using Genetic Analysis of Statewide Blood Screening Program', *Neurology*, 96: e1045-e53.
- Jung, Hyun Suk, Stan A. Burgess, Neil Billington, Melanie Colegrave, Hitesh Patel, Joseph M. Chalovich, Peter D. Chantler, and Peter J. Knight. 2008. 'Conservation of the regulated structure of folded myosin 2 in species separated by at least 600 million years of independent evolution', *Proc Natl Acad Sci U S A*, 105: 6022-26.
- Kaliman, P., D. Catalucci, J. T. Lam, R. Kondo, J. C. Gutiérrez, S. Reddy, M. Palacín, A. Zorzano, K. R. Chien, and P. Ruiz-Lozano. 2005. 'Myotonic dystrophy protein kinase phosphorylates phospholamban and regulates calcium uptake in cardiomyocyte sarcoplasmic reticulum', *J Biol Chem*, 280: 8016-21.
- Kalsotra, Auinash, Xinshu Xiao, Amanda J. Ward, John C. Castle, Jason M. Johnson, Christopher B. Burge, and Thomas A. Cooper. 2008. 'A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart', *Proc Natl Acad Sci U S A*, 105: 20333-38.
- Kang, Yu-Jian, De-Chang Yang, Lei Kong, Mei Hou, Yu-Qi Meng, Liping Wei, and Ge Gao. 2017. 'CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features', *Nucleic acids research*, 45: W12-W16.
- Kinniburgh, Alan J., Janet E. Mertz, and Jeffrey Ross. 1978. 'The precursor of mouse β -globin messenger RNA contains two intervening RNA sequences', *Cell*, 14: 681-93.
- Kite, GL. 1913. 'THE RELATIVE PERMEABILITY OF THE SURFACE AND INTERIOR PORTIONS OF THE CYTOPLASM OF ANIMAL AND PLANT CELLS: (A PRELIMINARY PAPER.)', *The Biological Bulletin*, 25: 1-7.
- Klesert, T. R., D. H. Cho, J. I. Clark, J. Maylie, J. Adelman, L. Snider, E. C. Yuen, P. Soriano, and S. J. Tapscott. 2000. 'Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy', *Nat Genet*, 25: 105-9.
- Klooster, Rinse, Kirsten Straasheijm, Bharati Shah, Janet Sowden, Rune Frants, Charles Thornton, Rabi Tawil, and Silvere van der Maarel. 2009. 'Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level', *European Journal of Human Genetics*, 17: 1615-24.
- Knopp, P., Y. D. Krom, C. R. Banerji, M. Panamarova, L. A. Moyle, B. den Hamer, S. M. van der Maarel, and P. S. Zammit. 2016. 'DUX4 induces a transcriptome more characteristic of a less-differentiated cell state and inhibits myogenesis', *J Cell Sci*, 129: 3816-31.
- Koebis, M., N. Ohsawa, Y. Kino, N. Sasagawa, I. Nishino, and S. Ishiura. 2011. 'Alternative splicing of myomesin 1 gene is aberrantly regulated in myotonic dystrophy type 1', *Genes Cells*, 16: 961-72.
- Konarska, M., W. Filipowicz, and H. J. Gross. 1982. 'RNA ligation via 2'-phosphomonoester, 3'5'-phosphodiester linkage: requirement of 2',3'-cyclic phosphate termini and involvement of a 5'-hydroxyl polynucleotide kinase', *Proc Natl Acad Sci U S A*, 79: 1474-8.
- Konarska, M. M., P. J. Grabowski, R. A. Padgett, and P. A. Sharp. 1985. 'Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors', *Nature*, 313: 552-7.

- Korfali, N., G. S. Wilkie, S. K. Swanson, V. Srsen, J. de Las Heras, D. G. Batrakou, P. Malik, N. Zuleger, A. R. Kerr, L. Florens, and E. C. Schirmer. 2012. 'The nuclear envelope proteome differs notably between tissues', *Nucleus*, 3: 552-64.
- Krawczak, M., J. Reiss, and D. N. Cooper. 1992. 'The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences', *Hum Genet*, 90: 41-54.
- Kuyumcu-Martinez, N. M., G. S. Wang, and T. A. Cooper. 2007. 'Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation', *Mol Cell*, 28: 68-78.
- Labeit, S., D. P. Barlow, M. Gautel, T. Gibson, J. Holt, C. L. Hsieh, U. Francke, K. Leonard, J. Wardale, A. Whiting, and et al. 1990. 'A regular pattern of two types of 100-residue motif in the sequence of titin', *Nature*, 345: 273-6.
- Labeit, S., and B. Kolmerer. 1995. 'Titins: giant proteins in charge of muscle ultrastructure and elasticity', *Science*, 270: 293-6.
- Landouzy, Louis, and Joseph Jules Dejerine. 1885. *De la myopathie atrophique progressive: myopathie sans neuropathie débutant d'ordinaire dans l'enfance par la face* (F. Alcan).
- Lavigueur, A., H. La Branche, A. R. Kornblihtt, and B. Chabot. 1993. 'A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding', *Genes Dev*, 7: 2405-17.
- Lehman, W., R. Craig, and P. Vibert. 1994. 'Ca(2+)-induced tropomyosin movement in Limulus thin filaments revealed by three-dimensional reconstruction', *Nature*, 368: 65-7.
- Lemmers, R. J., J. J. Goeman, P. J. van der Vliet, M. P. van Nieuwenhuizen, J. Balog, M. Vos-Versteeg, P. Camano, M. A. Ramos Arroyo, I. Jerico, M. T. Rogers, D. G. Miller, M. Upadhyaya, J. J. Verschuuren, A. Lopez de Munain Arregui, B. G. van Engelen, G. W. Padberg, S. Sacconi, R. Tawil, S. J. Tapscott, B. Bakker, and S. M. van der Maarel. 2015. 'Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2', *Hum Mol Genet*, 24: 659-69.
- Li, Dunhui, Craig Stewart McIntosh, Frank Louis Mastaglia, Steve Donald Wilton, and May Thandar Aung-Htut. 2021. 'Neurodegenerative diseases: a hotbed for splicing defects and the potential therapies', *Translational Neurodegeneration*, 10: 16.
- Liao, Q., Y. Zhang, J. He, and K. Huang. 2022. 'Global Prevalence of Myotonic Dystrophy: An Updated Systematic Review and Meta-Analysis', *Neuroepidemiology*, 56: 163-73.
- Lin, X., J. W. Miller, A. Mankodi, R. N. Kanadia, Y. Yuan, R. T. Moxley, M. S. Swanson, and C. A. Thornton. 2006. 'Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy', *Hum Mol Genet*, 15: 2087-97.
- Lindemann, J. P. 1986. 'Alpha-adrenergic stimulation of sarcolemmal protein phosphorylation and slow responses in intact myocardium', *J Biol Chem*, 261: 4860-7.
- Liu, M., and A. Grigoriev. 2004. 'Protein domains correlate strongly with exons in multiple eukaryotic genomes--evidence of exon shuffling?', *Trends Genet*, 20: 399-403.

- Loescher, Christine M, Anastasia J Hobbach, and Wolfgang A Linke. 2021. 'Titin (TTN): from molecule to modifications, mechanics, and medical significance', *Cardiovascular Research*, 118: 2903-18.
- López Castel, A., M. Nakamori, S. Tomé, D. Chitayat, G. Gourdon, C. A. Thornton, and C. E. Pearson. 2011. 'Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues', *Hum Mol Genet*, 20: 1-15.
- Lorand, L. 1953. 'Adenosine triphosphate-creatine transphosphorylase as relaxing factor of muscle', *Nature*, 172: 1181-3.
- Lorson, C. L., E. Hahnen, E. J. Androphy, and B. Wirth. 1999. 'A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy', *Proc Natl Acad Sci U S A*, 96: 6307-11.
- Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol*, 15: 550.
- Lunt, P. W., and P. S. Harper. 1991. 'Genetic counselling in facioscapulohumeral muscular dystrophy', *J Med Genet*, 28: 655-64.
- Mackels, L., X. Liu, G. Bonne, and L. Servais. 2023. 'TOR1AIP1-Associated Nuclear Envelopathies', *Int J Mol Sci*, 24.
- Mahdy, M. A. A. 2019. 'Skeletal muscle fibrosis: an overview', *Cell Tissue Res*, 375: 575-88.
- Maimaiti, R., C. Zhu, Y. Zhang, Q. Ding, and W. Guo. 2021. 'RBM20-Mediated Pre-mRNA Splicing Has Muscle-Specificity and Differential Hormonal Responses between Muscles and in Muscle Cell Cultures', *Int J Mol Sci*, 22.
- Manilal, S., T. M. Nguyen, C. A. Sewry, and G. E. Morris. 1996. 'The Emery-Dreifuss muscular dystrophy protein, emerin, is a nuclear membrane protein', *Hum Mol Genet*, 5: 801-8.
- Mankodi, A., E. Logigian, L. Callahan, C. McClain, R. White, D. Henderson, M. Krym, and C. A. Thornton. 2000. 'Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat', *Science*, 289: 1769-73.
- Masny, Peter S., On Ying A. Chan, Jessica C. de Greef, Ulla Bengtsson, Melanie Ehrlich, Rabi Tawil, Leslie F. Lock, Jane E. Hewitt, Jennifer Stocksdales, Jorge H. Martin, Silvere M. van der Maarel, and Sara T. Winokur. 2010. 'Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei', *European Journal of Human Genetics*, 18: 448-56.
- McConkie-Rosell, A., A. M. Lachiewicz, G. A. Spiridigliozzi, J. Tarleton, S. Schoenwald, M. C. Phelan, P. Goonewardena, X. Ding, and W. T. Brown. 1993. 'Evidence that methylation of the FMR-I locus is responsible for variable phenotypic expression of the fragile X syndrome', *Am J Hum Genet*, 53: 800-9.
- Meinke, P., A. R. W. Kerr, R. Czapiewski, J. I. de Las Heras, C. R. Dixon, E. Harris, H. Kölbl, F. Muntoni, U. Schara, V. Straub, B. Schoser, M. Wehnert, and E. C. Schirmer. 2020. 'A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism', *EBioMedicine*, 51: 102587.
- Meinke, P., E. Mattioli, F. Haque, S. Antoku, M. Columbaro, K. R. Straatman, H. J. Worman, G. G. Gundersen, G. Lattanzi, M. Wehnert, and S. Shackleton. 2014. 'Muscular dystrophy-associated SUN1 and SUN2 variants disrupt nuclear-

- cytoskeletal connections and myonuclear organization', *PLoS Genet*, 10: e1004605.
- Meinke, Peter. 'Molecular Genetics of Emery–Dreifuss Muscular Dystrophy.' in, *Encyclopedia of Life Sciences*.
- Meinke, Peter, Alexandr A. Makarov, Phú Lê Thành, Daina Sadurska, and Eric C. Schirmer. 2015. 'Nucleoskeleton dynamics and functions in health and disease', *Cell Health and Cytoskeleton*, 7: 55-69.
- Mészáros, Bálint, Gábor Erdős, and Zsuzsanna Dosztányi. 2018. 'IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding', *Nucleic acids research*, 46: W329-W37.
- Michel, L., A. Huguet-Lachon, and G. Gourdon. 2015. 'Sense and Antisense DMPK RNA Foci Accumulate in DM1 Tissues during Development', *PLoS One*, 10: e0137620.
- Miller, J. W., C. R. Urbinati, P. Teng-Umnuay, M. G. Stenberg, B. J. Byrne, C. A. Thornton, and M. S. Swanson. 2000. 'Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy', *Embo j*, 19: 4439-48.
- Minnerop, M., B. Weber, J. C. Schoene-Bake, S. Roeske, S. Mirbach, C. Anspach, C. Schneider-Gold, R. C. Betz, C. Helmstaedter, M. Tittgemeyer, T. Klockgether, and C. Kornblum. 2011. 'The brain in myotonic dystrophy 1 and 2: evidence for a predominant white matter disease', *Brain*, 134: 3530-46.
- Mounsey, J. P., J. E. John, 3rd, S. M. Helmke, E. W. Bush, J. Gilbert, A. D. Roses, M. B. Perryman, L. R. Jones, and J. R. Moorman. 2000. 'Phospholemman is a substrate for myotonic dystrophy protein kinase', *J Biol Chem*, 275: 23362-7.
- Muchir, A., G. Bonne, A. J. van der Kooij, M. van Meegen, F. Baas, P. A. Bolhuis, M. de Visser, and K. Schwartz. 2000. 'Identification of mutations in the gene encoding lamins A/C in autosomal dominant limb girdle muscular dystrophy with atrioventricular conduction disturbances (LGMD1B)', *Hum Mol Genet*, 9: 1453-9.
- Muntoni, F., S. Torelli, and A. Ferlini. 2003. 'Dystrophin and mutations: one gene, several proteins, multiple phenotypes', *Lancet Neurol*, 2: 731-40.
- Nakamori, Masayuki, Krzysztof Sobczak, Araya Puwanant, Steve Welle, Katy Eichinger, Shree Pandya, Jeanne Dekdebrun, Chad R. Heatwole, Michael P. McDermott, Tian Chen, Melissa Cline, Rabi Tawil, Robert J. Osborne, Thurman M. Wheeler, Maurice S. Swanson, Richard T. Moxley, 3rd, and Charles A. Thornton. 2013. 'Splicing biomarkers of disease severity in myotonic dystrophy', *Annals of neurology*, 74: 862-72.
- Newell-Price, J., A. J. Clark, and P. King. 2000. 'DNA methylation and silencing of gene expression', *Trends in endocrinology and metabolism: TEM*, 11: 142-8.
- Novelli, G., M. Gennarelli, G. Zelano, A. Pizzuti, C. Fattorini, C. T. Caskey, and B. Dallapiccola. 1993. 'Failure in detecting mRNA transcripts from the mutated allele in myotonic dystrophy muscle', *Biochem Mol Biol Int*, 29: 291-7.
- Nunes, E. A., T. Stokes, J. McKendry, B. S. Currier, and S. M. Phillips. 2022. 'Disuse-induced skeletal muscle atrophy in disease and nondisease states in humans: mechanisms, prevention, and recovery strategies', *Am J Physiol Cell Physiol*, 322: C1068-c84.
- O'Coilain, D. Fearghas, Carmen Perez-Terzic, Santiago Reyes, Garvan C. Kane, Atta Behfar, Denice M. Hodgson, Jeffrey A. Strommen, Xiao-Ke Liu, Walther van den Broek, Derick G. Wansink, Bé Wieringa, and Andre Terzic. 2004. 'Transgenic overexpression of human DMPK accumulates into hypertrophic cardiomyopathy,

- myotonic myopathy and hypotension traits of myotonic dystrophy', *Hum Mol Genet*, 13: 2505-18.
- Orengo, J. P., A. J. Ward, and T. A. Cooper. 2011. 'Alternative splicing dysregulation secondary to skeletal muscle regeneration', *Ann Neurol*, 69: 681-90.
- Otomo, J., S. Kure, T. Shiba, A. Karibe, T. Shinozaki, T. Yagi, H. Naganuma, F. Tezuka, M. Miura, M. Ito, J. Watanabe, Y. Matsubara, and K. Shirato. 2005. 'Electrophysiological and histopathological characteristics of progressive atrioventricular block accompanied by familial dilated cardiomyopathy caused by a novel mutation of lamin A/C gene', *J Cardiovasc Electrophysiol*, 16: 137-45.
- Palade, G. E. 1955. 'A small particulate component of the cytoplasm', *J Biophys Biochem Cytol*, 1: 59-68.
- Palmer, C. J., B. T. Scott, and L. R. Jones. 1991. 'Purification and complete sequence determination of the major plasma membrane substrate for cAMP-dependent protein kinase and protein kinase C in myocardium', *J Biol Chem*, 266: 11126-30.
- Pan, Q., O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe. 2008. 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nat Genet*, 40: 1413-5.
- Pascual-Gilabert, M., A. López-Castel, and R. Artero. 2021. 'Myotonic dystrophy type 1 drug development: A pipeline toward the market', *Drug Discov Today*, 26: 1765-72.
- Paulson, H. 2018. 'Repeat expansion diseases', *Handb Clin Neurol*, 147: 105-23.
- Phillips-Cremins, J. E., M. E. Sauria, A. Sanyal, T. I. Gerasimova, B. R. Lajoie, J. S. Bell, C. T. Ong, T. A. Hookway, C. Guo, Y. Sun, M. J. Bland, W. Wagstaff, S. Dalton, T. C. McDevitt, R. Sen, J. Dekker, J. Taylor, and V. G. Corces. 2013. 'Architectural protein subclasses shape 3D organization of genomes during lineage commitment', *Cell*, 153: 1281-95.
- Poeta, L., D. Drongitis, L. Verrillo, and M. G. Miano. 2020. 'DNA Hypermethylation and Unstable Repeat Diseases: A Paradigm of Transcriptional Silencing to Decipher the Basis of Pathogenic Mechanisms', *Genes*, 11.
- Preston, MK; Tawil, R; Wang, LH. 1999. *Facioscapulohumeral Muscular Dystrophy*. (Seattle).
- Punta, Marco, Penny C. Coggill, Ruth Y. Eberhardt, Jaina Mistry, John Tate, Chris Bournsnell, Ningze Pang, Kristoffer Forslund, Goran Ceric, Jody Clements, Andreas Heger, Liisa Holm, Erik L. L. Sonnhammer, Sean R. Eddy, Alex Bateman, and Robert D. Finn. 2011. 'The Pfam protein families database', *Nucleic acids research*, 40: D290-D301.
- Rearick, D., A. Prakash, A. McSweeney, S. S. Shepard, L. Fedorova, and A. Fedorov. 2011. 'Critical association of ncRNA with introns', *Nucleic acids research*, 39: 2357-66.
- Reddy, S., D. B. Smith, M. M. Rich, J. M. Leferovich, P. Reilly, B. M. Davis, K. Tran, H. Rayburn, R. Bronson, D. Cros, R. J. Balice-Gordon, and D. Housman. 1996. 'Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy', *Nat Genet*, 13: 325-35.
- Rickard, A. M., L. M. Petek, and D. G. Miller. 2015. 'Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways', *Hum Mol Genet*, 24: 5901-14.
- Robson, M. I., J. I. de Las Heras, R. Czapiewski, Lê Thành P, D. G. Booth, D. A. Kelly, S. Webb, A. R. W. Kerr, and E. C. Schirmer. 2016. 'Tissue-Specific Gene

- Repositioning by Muscle Nuclear Membrane Proteins Enhances Repression of Critical Developmental Genes during Myogenesis', *Mol Cell*, 62: 834-47.
- Rottman, F., A. J. Shatkin, and R. P. Perry. 1974. 'Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing', *Cell*, 3: 197-9.
- Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green. 1984. 'Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro', *Cell*, 38: 317-31.
- Ruskin, Barbara, and Michael R. Green. 1985. 'Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing', *Nature*, 317: 732-34.
- Salsia, V., G. N. A. Vattemi, and R. G. Tupler. 2023. 'The FSHD jigsaw: are we placing the tiles in the right position?', *Curr Opin Neurol*.
- Samaranch, L., O. Lorenzo-Betancor, J. M. Arbelo, I. Ferrer, E. Lorenzo, J. Irigoyen, M. A. Pastor, C. Marrero, C. Isla, J. Herrera-Henriquez, and P. Pastor. 2010. 'PINK1-linked parkinsonism is associated with Lewy body pathology', *Brain*, 133: 1128-42.
- Sanchez de Groot, Natalia, Alexandros Armaos, Ricardo Graña-Montes, Marion Alriquet, Giulia Calloni, R. Martin Vabulas, and Gian Gaetano Tartaglia. 2019. 'RNA structure drives interaction with proteins', *Nature communications*, 10: 3246.
- Schätzl, T., L. Kaiser, and H. P. Daigner. 2021. 'Facioscapulohumeral muscular dystrophy: genetics, gene activation and downstream signalling with regard to recent therapeutic approaches: an update', *Orphanet J Rare Dis*, 16: 129.
- Scherrer, K., H. Latham, and J. E. Darnell. 1963. 'Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells', *Proc Natl Acad Sci U S A*, 49: 240-8.
- Scotti, Marina M., and Maurice S. Swanson. 2016. 'RNA mis-splicing in disease', *Nature Reviews Genetics*, 17: 19-32.
- Sharma, V., S. N. Pandey, H. Khawaja, K. J. Brown, Y. Hathout, and Y. W. Chen. 2016. 'PARP1 Differentially Interacts with Promoter region of DUX4 Gene in FSHD Myoblasts', *J Genet Syndr Gene Ther*, 7.
- Shen, S., J. W. Park, Z. X. Lu, L. Lin, M. D. Henry, Y. N. Wu, Q. Zhou, and Y. Xing. 2014. 'rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data', *Proc Natl Acad Sci U S A*, 111: E5593-601.
- Snider, L., L. N. Geng, R. J. Lemmers, M. Kyba, C. B. Ware, A. M. Nelson, R. Tawil, G. N. Filippova, S. M. van der Maarel, S. J. Tapscott, and D. G. Miller. 2010. 'Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene', *PLoS Genet*, 6: e1001181.
- Sofueva, S., E. Yaffe, W. C. Chan, D. Georgopoulou, M. Vietri Rudan, H. Mira-Bontenbal, S. M. Pollard, G. P. Schroth, A. Tanay, and S. Hadjur. 2013. 'Cohesin-mediated interactions organize chromosomal domain architecture', *Embo j*, 32: 3119-29.
- Sosa, B. A., A. Rothballer, U. Kutay, and T. U. Schwartz. 2012. 'LINC complexes form by binding of three KASH peptides to domain interfaces of trimeric SUN proteins', *Cell*, 149: 1035-47.
- Sposito, R., L. Pasquali, F. Galluzzi, A. Rocchi, B. Solito, D. Soragna, R. Tupler, and G. Siciliano. 2005. 'Facioscapulohumeral muscular dystrophy type 1A in northwestern Tuscany: a molecular genetics-based epidemiological and genotype-phenotype study', *Genet Test*, 9: 30-6.

- Squire, John M. 1997. 'Architecture and function in the muscle sarcomere', *Current Opinion in Structural Biology*, 7: 247-57.
- Starr, D. A., and M. Han. 2003. 'ANChors away: an actin based mechanism of nuclear positioning', *J Cell Sci*, 116: 211-6.
- Statland, J. M., M. P. McDermott, C. Heatwole, W. B. Martens, S. Pandya, E. L. van der Kooi, J. T. Kissel, K. R. Wagner, and R. Tawil. 2013. 'Reevaluating measures of disease progression in facioscapulohumeral muscular dystrophy', *Neuromuscul Disord*, 23: 306-12.
- Staudt, D., and D. Stainier. 2012. 'Uncovering the molecular and cellular mechanisms of heart development using the zebrafish', *Annu Rev Genet*, 46: 397-418.
- Steinbach, P., D. Gläser, W. Vogel, M. Wolf, and S. Schwemmler. 1998. 'The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat', *Am J Hum Genet*, 62: 278-85.
- Strahl, B. D., and C. D. Allis. 2000. 'The language of covalent histone modifications', *Nature*, 403: 41-5.
- Sun, J. H., L. Zhou, D. J. Emerson, S. A. Phyo, K. R. Titus, W. Gong, T. G. Gilgenast, J. A. Beagan, B. L. Davidson, F. Tassone, and J. E. Phillips-Cremins. 2018. 'Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries', *Cell*, 175: 224-38.e15.
- Sun, S., S. C. Ling, J. Qiu, C. P. Albuquerque, Y. Zhou, S. Tokunaga, H. Li, H. Qiu, A. Bui, G. W. Yeo, E. J. Huang, K. Eggan, H. Zhou, X. D. Fu, C. Lagier-Tourenne, and D. W. Cleveland. 2015. 'ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP', *Nat Commun*, 6: 6171.
- Tawil, R., and S. M. Van Der Maarel. 2006. 'Facioscapulohumeral muscular dystrophy', *Muscle Nerve*, 34: 1-15.
- Tawil, Rabi, Silvére M. van der Maarel, and Stephen J. Tapscott. 2014. 'Facioscapulohumeral dystrophy: the path to consensus on pathophysiology', *Skeletal Muscle*, 4: 12.
- Tennyson, C. N., H. J. Klamut, and R. G. Worton. 1995. 'The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced', *Nat Genet*, 9: 184-90.
- Teufel, Felix, José Juan Almagro Armenteros, Alexander Rosenberg Johansen, Magnús Halldór Gíslason, Silas Irby Pihl, Konstantinos D. Tsirigos, Ole Winther, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2022. 'SignalP 6.0 predicts all five types of signal peptides using protein language models', *Nat Biotechnol*, 40: 1023-25.
- Thornton, C. A., J. P. Wymer, Z. Simmons, C. McClain, and R. T. Moxley, 3rd. 1997. 'Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene', *Nat Genet*, 16: 407-9.
- Todorow, V., S. Hintze, A. R. W. Kerr, A. Hehr, B. Schoser, and P. Meinke. 2021. 'Transcriptome Analysis in a Primary Human Muscle Cell Differentiation Model for Myotonic Dystrophy Type 1', *Int J Mol Sci*, 22.
- Todorow, V., S. Hintze, B. Schoser, and P. Meinke. 2022. 'Nuclear envelope transmembrane proteins involved in genome organization are misregulated in myotonic dystrophy type 1 muscle', *Front Cell Dev Biol*, 10: 1007331.
- van den Boogaard, M. L., Rjlf Lemmers, J. Balog, M. Wohlgemuth, M. Auranen, S. Mitsuhashi, P. J. van der Vliet, K. R. Straasheijm, R. F. P. van den Akker, M. Kriek,

- M. E. Y. Laurence-Bik, V. Raz, M. M. van Ostaijen-Ten Dam, K. B. M. Hansson, E. L. van der Kooi, S. Kiuru-Enari, B. Udd, M. J. D. van Tol, I. Nishino, R. Tawil, S. J. Tapscott, B. G. M. van Engelen, and S. M. van der Maarel. 2016. 'Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy', *Am J Hum Genet*, 98: 1020-29.
- van der Ven, P. F., G. Jansen, T. H. van Kuppevelt, M. B. Perryman, M. Lupa, P. W. Dunne, H. J. ter Laak, P. H. Jap, J. H. Veerkamp, H. F. Epstein, and et al. 1993. 'Myotonic dystrophy kinase is a component of neuromuscular junctions', *Hum Mol Genet*, 2: 1889-94.
- van Deutekom, J. C., C. Wijmenga, E. A. van Tienhoven, A. M. Gruter, J. E. Hewitt, G. W. Padberg, G. J. van Ommen, M. H. Hofker, and R. R. Frants. 1993. 'FSD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit', *Hum Mol Genet*, 2: 2037-42.
- van Steensel, B., and A. S. Belmont. 2017. 'Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression', *Cell*, 169: 780-91.
- Vaquero-Garcia, J., A. Barrera, M. R. Gazzara, J. González-Vallinas, N. F. Lahens, J. B. Hogenesch, K. W. Lynch, and Y. Barash. 2016. 'A new view of transcriptome complexity and regulation through the lens of local splicing variations', *eLife*, 5: e11752.
- Verstraeten, V. L., J. L. Broers, F. C. Ramaekers, and M. A. van Steensel. 2007. 'The nuclear envelope, a key structure in cellular integrity and gene expression', *Curr Med Chem*, 14: 1231-48.
- Vigouroux, Corinne, and Gisèle Bonne. 2013. 'Laminopathies: One Gene, Two Proteins, Five Diseases.' in, *Madame Curie Bioscience Database [Internet]* (Landes Bioscience).
- Vitting-Seerup, K., and A. Sandelin. 2019. 'IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences', *Bioinformatics*, 35: 4469-71.
- Vitting-Seerup, Kristoffer, Bo Torben Porse, Albin Sandelin, and Johannes Waage. 2014. 'spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data', *BMC Bioinformatics*, 15: 81.
- Voermans, N. C., C. E. Erasmus, C. W. Ockeloën, B. G. Van Engelen, and C. A. Eggink. 2015. 'Primary cataract as a key to recognition of myotonic dystrophy type 1', *Eur J Ophthalmol*, 25: e46-9.
- Wang, E. T., D. Treacy, K. Eichinger, A. Struck, J. Estabrook, H. Olafson, T. T. Wang, K. Bhatt, T. Westbrook, S. Sedehizadeh, A. Ward, J. Day, D. Brook, J. A. Berglund, T. Cooper, D. Housman, C. Thornton, and C. Burge. 2019. 'Transcriptome alterations in myotonic dystrophy skeletal muscle and heart', *Hum Mol Genet*, 28: 1312-21.
- Wang, Y. H., S. Amirhaeri, S. Kang, R. D. Wells, and J. D. Griffith. 1994. 'Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene', *Science*, 265: 669-71.
- Wang, Y., and J. E. Pessin. 2013. 'Mechanisms for fiber-type specificity of skeletal muscle atrophy', *Curr Opin Clin Nutr Metab Care*, 16: 243-50.
- Watson, M. L. 1955. 'The nuclear envelope; its structure and relation to cytoplasmic membranes', *J Biophys Biochem Cytol*, 1: 257-70.

- Wenninger, Stephan, Federica Montagnese, and Benedikt Schoser. 2018. 'Core Clinical Phenotypes in Myotonic Dystrophies', *Frontiers in neurology*, 9: 303-03.
- Wijmenga, C., R. R. Frants, O. F. Brouwer, P. Moerer, J. L. Weber, and G. W. Padberg. 1990. 'Location of facioscapulohumeral muscular dystrophy gene on chromosome 4', *Lancet*, 336: 651-3.
- Wijmenga, C., J. E. Hewitt, L. A. Sandkuijl, L. N. Clark, T. J. Wright, H. G. Dauwerse, A. M. Gruter, M. H. Hofker, P. Moerer, R. Williamson, and et al. 1992. 'Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy', *Nat Genet*, 2: 26-30.
- Wilhelmsen, K., M. Ketema, H. Truong, and A. Sonnenberg. 2006. 'KASH-domain proteins in nuclear migration, anchorage and other processes', *J Cell Sci*, 119: 5021-9.
- Wilhelmsen, K., S. H. Litjens, I. Kuikman, N. Tshimbalanga, H. Janssen, I. van den Bout, K. Raymond, and A. Sonnenberg. 2005. 'Nesprin-3, a novel outer nuclear membrane protein, associates with the cytoskeletal linker protein plectin', *J Cell Biol*, 171: 799-810.
- Will, Cindy L., and Reinhard Lührmann. 2011. 'Spliceosome structure and function', *Cold Spring Harbor perspectives in biology*, 3: a003707.
- Wilson, K. L., and R. Foisner. 2010. 'Lamin-binding Proteins', *Cold Spring Harb Perspect Biol*, 2: a000554.
- Wu, F., and J. Yao. 2017. 'Identifying Novel Transcriptional and Epigenetic Features of Nuclear Lamina-associated Genes', *Sci Rep*, 7: 100.
- Wu, J. Y., and T. Maniatis. 1993. 'Specific interactions between proteins implicated in splice site selection and regulated alternative splicing', *Cell*, 75: 1061-70.
- Yao, Z., L. Snider, J. Balog, R. J. Lemmers, S. M. Van Der Maarel, R. Tawil, and S. J. Tapscott. 2014. 'DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle', *Hum Mol Genet*, 23: 5342-52.
- Zhang, Q., C. Bethmann, N. F. Worth, J. D. Davies, C. Wasner, A. Feuer, C. D. Ragnauth, Q. Yi, J. A. Mellad, D. T. Warren, M. A. Wheeler, J. A. Ellis, J. N. Skepper, M. Vorgerd, B. Schlotter-Weigel, P. L. Weissberg, R. G. Roberts, M. Wehnert, and C. M. Shanahan. 2007. 'Nesprin-1 and -2 are involved in the pathogenesis of Emery Dreifuss muscular dystrophy and are critical for nuclear envelope integrity', *Hum Mol Genet*, 16: 2816-33.
- Zhang, Q., C. Ragnauth, M. J. Greener, C. M. Shanahan, and R. G. Roberts. 2002. 'The nesprins are giant actin-binding proteins, orthologous to Drosophila melanogaster muscle protein MSP-300', *Genomics*, 80: 473-81.
- Zhao, Shanrong, Wai-Ping Fung-Leung, Anton Bittner, Karen Ngo, and Xuejun Liu. 2014. 'Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells', *PLoS One*, 9: e78644.
- Zuin, J., J. R. Dixon, M. I. van der Reijden, Z. Ye, P. Kolovos, R. W. Brouwer, M. P. van de Corput, H. J. van de Werken, T. A. Knoch, IJcken W. F. van, F. G. Grosveld, B. Ren, and K. S. Wendt. 2014. 'Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells', *Proc Natl Acad Sci U S A*, 111: 996-1001.

5 APPENDICES

5.1 Publication I

Transcriptome Analysis in a Primary Human Muscle Cell Differentiation Model for

Myotonic Dystrophy Type 1

Vanessa Todorow, Stefan Hintze, Alastair. R. W. Kerr, Andreas Hehr, Benedikt Schoser and Peter Meinke (2021)

International Journal of Molecular Sciences, 22, doi: 10.3390/ijms22168607

Contributions

I contributed to the following paper by conducting and interpreting the bioinformatic analysis of the RNAseq data, including visualizing the results and designing the final figures. I further participated in writing of the original draft as well as reviewing and editing the final manuscript.



Article

Transcriptome Analysis in a Primary Human Muscle Cell Differentiation Model for Myotonic Dystrophy Type 1

Vanessa Todorow^{1,†}, Stefan Hintze^{1,†}, Alastair R. W. Kerr², Andreas Hehr³, Benedikt Schoser¹ and Peter Meinke^{1,*}

¹ Department of Neurology, Friedrich-Baur-Institute, LMU Klinikum, Ludwig-Maximilians-University Munich, 80336 Munich, Germany; vanessa.todorow@med.uni-muenchen.de (V.T.); stefan.hintze@med.uni-muenchen.de (S.H.); benedikt.schoser@med.uni-muenchen.de (B.S.)

² Cancer Biomarker Centre, CRUK Manchester Institute, University of Manchester, Manchester SK10 4TG, UK; alastair.kerr@cruk.manchester.ac.uk

³ Centre for Human Genetics, 93047 Regensburg, Germany; andreas.hehr@humangenetik-regensburg.de

* Correspondence: peter.meinke@med.uni-muenchen.de; Tel.: +49-(0)-89-2180-78279

† These authors contributed equally to this work.

Abstract: Myotonic dystrophy type 1 (DM1) is caused by CTG-repeat expansions leading to a complex pathology with a multisystemic phenotype that primarily affects the muscles and brain. Despite a multitude of information, especially on the alternative splicing of several genes involved in the pathology, information about additional factors contributing to the disease development is still lacking. We performed RNAseq and gene expression analyses on proliferating primary human myoblasts and differentiated myotubes. GO-term analysis indicates that in myoblasts and myotubes, different molecular pathologies are involved in the development of the muscular phenotype. Gene set enrichment for splicing reveals the likelihood of whole, differentiation stage specific, splicing complexes that are misregulated in DM1. These data add complexity to the alternative splicing phenotype and we predict that it will be of high importance for therapeutic interventions to target not only mature muscle, but also satellite cells.

Keywords: myotonic dystrophy type 1; human primary muscle cell culture; transcriptomics; splicing



Citation: Todorow, V.; Hintze, S.; Kerr, A.R.W.; Hehr, A.; Schoser, B.; Meinke, P. Transcriptome Analysis in a Primary Human Muscle Cell Differentiation Model for Myotonic Dystrophy Type 1. *Int. J. Mol. Sci.* **2021**, *22*, 8607. <https://doi.org/10.3390/ijms22168607>

Academic Editor: Catherine Coirault

Received: 28 June 2021

Accepted: 6 August 2021

Published: 10 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Myotonic dystrophy type 1 (DM1) is a multisystemic disorder with wide ranging effects starting from skeletal muscle weakness, wasting and myotonia, cardiac arrhythmia, cataracts and insulin resistance, up to central nervous system dysfunctionality, sleep-wake cycle disturbances, endocrine dysfunction, frontal balding and a shortened lifespan [1–4]. Estimations of the prevalence of DM1 are about 1 in 8000 or fewer [5,6]. The predominant muscle involvement makes DM1 one of the most frequent muscular dystrophies in adulthood.

The occurrence, combination and severity of symptoms is highly variable and positively correlated to a pathological CTG-repeat expansion in the 3' UTR of the *DMPK* (DM1 protein kinase) gene [7,8]. Up to 35 CTG triplets in blood derived DNA are considered normal, a repeat length between 35 and 49 is considered to be a premutation. Between 50 and ~150 repeats a mild expression of the phenotype has been observed and ~100 to ~1000 CTG repeats were identified in patients with classical DM. Repeats consisting of more than 1000 CTG-triplets usually result in congenital DM, the most severe expression of the disease [9]. The extended CTG-repeats are unstable and tend to expand further. This results in anticipation, an increase in the CTG-repeat length in offspring paralleled by earlier onset and severer disease symptoms compared to their parents [10–12]. The sexual inheritance also affects the severity of the disease: maternal inheritance results in more severe clinical features than paternal inheritance [13,14]. Furthermore, somatic mosaicism

has been observed in DM1: a tissue-specific and age-dependent length heterogeneity of the *DMPK* CTG-repeat [15,16].

The main pathomechanism described for DM1 is RNA-toxicity caused by the formation of hairpin structures in the *DMPK* RNA transcript due to expanded CUG repeats [17,18]. The RNA gain-of-function paradigm is supported by the expression of CUG expansions independent of the *DMPK* locus in transgenic mice, which cause a severe DM1-like phenotype [19]. Regarding the mechanism by which these repeat expansions result in the DM1 phenotype, there are several proposed theories [20]. These include (i) alternative splicing of several mRNAs [21–24], (ii) altered transcriptional regulation [25–27], (iii) inhibited translation [28,29], (iv) repeat associated non-ATG (RAN) translation resulting in the presence of toxic peptides [30–32], (v) alternative polyadenylation of several mRNAs [33] and (vi) miRNA misregulation [27,34–39]. These mechanisms are not exclusive; however, they do contribute to the complex clinical phenotype.

The mechanism leading to a disbalanced splicing network originates directly from the formation of the hairpin structure in the *DMPK* mRNA, whose CUG expansion-driven secondary structure is recognized by diverse RNA-binding proteins, for instance muscleblind family members (MBNL1-3) [40,41]. The subsequent sequestration of MBNL is tantamount to its depletion from the nucleoplasm and with this, loss of function. At the same time, another splicing factor, CUGBP elav-like family member 1 (CELF1), is hyperphosphorylated and thus stabilized, leading to a gain in function [42,43]. Splicing dysregulation of several key transcripts has been shown to be involved in the development of tissue specific pathologies in DM1. Examples for genes affected by this missplicing include *CLCN1* (chloride voltage-gated channel 1) [44,45], *DMD* (dystrophin) [46], *RYR1* (ryanodine receptor 1) [47], *INSR* (insulin receptor), and *BIN1* (bridging integrator 1) [48]. Interestingly, many of the reported alternative splicing events represent a shift from adult to embryonic splicing patterning [49]. Healthy muscle tissue undergoes an MBNL/CELF-dependent, postnatal switch of alternative splicing patterns necessary to match the increasing demands of adult muscle. While MBNL1 levels are low in early embryonic stages and increase during development, CELF1 expression follows the exact opposite order [49,50]. Although MBNL and CELF1 imbalance accounts for a whole range of missplicing events in DM1, and thus have been a main scientific interest for a long time, the phenotypic presentation is unlikely to be caused entirely by those two factors alone. Indeed, a broad splicing network misregulation appears likely as the basis of DM1. Consistently, more splicing factors have been identified to regulate alternative splicing events in DM1, among them STAU1 and HNRNPA1 [51,52].

Apart from changes to the splicing pattern, which have been subject to numerous studies, there are also indications that gene expression alterations are playing a role in DM1. Examples are MBNL1 itself [53], but also various ion channels like calcium, sodium and potassium channels [53], as well as transcription factors such as MYOG, MYOD and SP1 [54]. MYOG and MYOD are both downregulated in DM1 and thus probably cause the aberrant and delayed differentiation process described in DM1 myoblast cell culture [55,56], possibly explaining poor muscle regeneration potential in patients. Furthermore, several signalling pathways are known to be affected through downregulated activity of kinases such as protein kinase B (AKT), MAPK and ERK, further contributing to the differentiation defect phenotype but also to metabolic dysregulation and autophagy [54]. Compared to the investigation of alternative splicing very little research has been focused on pathways affected by differential gene expression. A large study performed on muscle biopsies from DM1 patients and healthy controls proposed downregulation of cell adhesion and translation, as well as upregulation of mitochondrial metabolic pathways [57].

It will be of high interest for upcoming therapeutic approaches to know which defects are playing an important role in the pathomechanism at which stage of differentiation. Thus, we performed an analysis of gene expression alterations and pathway regulation in proliferating primary myoblasts as well as differentiated myotubes gained from DM1 patients and compared them to materials gained from non-disease controls.

2. Results

2.1. Differential Gene Expression in DM1 Myoblasts and Myotubes Compared to Non-Disease Controls

First, DM1 samples were investigated for the typical DM1 phenotype. Cells of all DM1 tissue cultures did show nuclear MBNL1 foci (Figure S1A) and *DMPK* was slightly upregulated in myoblasts as well as myotubes (Figure S1B). To investigate differential expression of genes between non-disease control and DM1, we performed RNAseq in proliferating myoblasts and post-mitotic myotubes. To avoid contamination of the myotubes by non-differentiated cells we isolated multinucleated myotubes by selective trypsin treatment followed by sedimentation. In total, more than 28,000 genes were identified. Read counts were analysed using DESeq2. Normalized counts are listed Supplemental Table S1. We identified 261 genes in proliferating myoblasts and 195 genes in differentiated myotubes with a log₂ fold change > ±1 (adjusted *p*-value < 0.1) of which 40 genes are shared between proliferating and differentiated cells (Figure 1; Tables S2 and S3), indicating differences in gene expression between non-disease controls and DM1 but also between proliferating myoblasts and post-mitotic myotubes. Gene expression changes were validated using RT-qPCR for selected genes (Figure S2). Subsequently, we performed an extensive pathway enrichment analysis using two approaches: classical gene ontology enrichment as well as gene set enrichment.

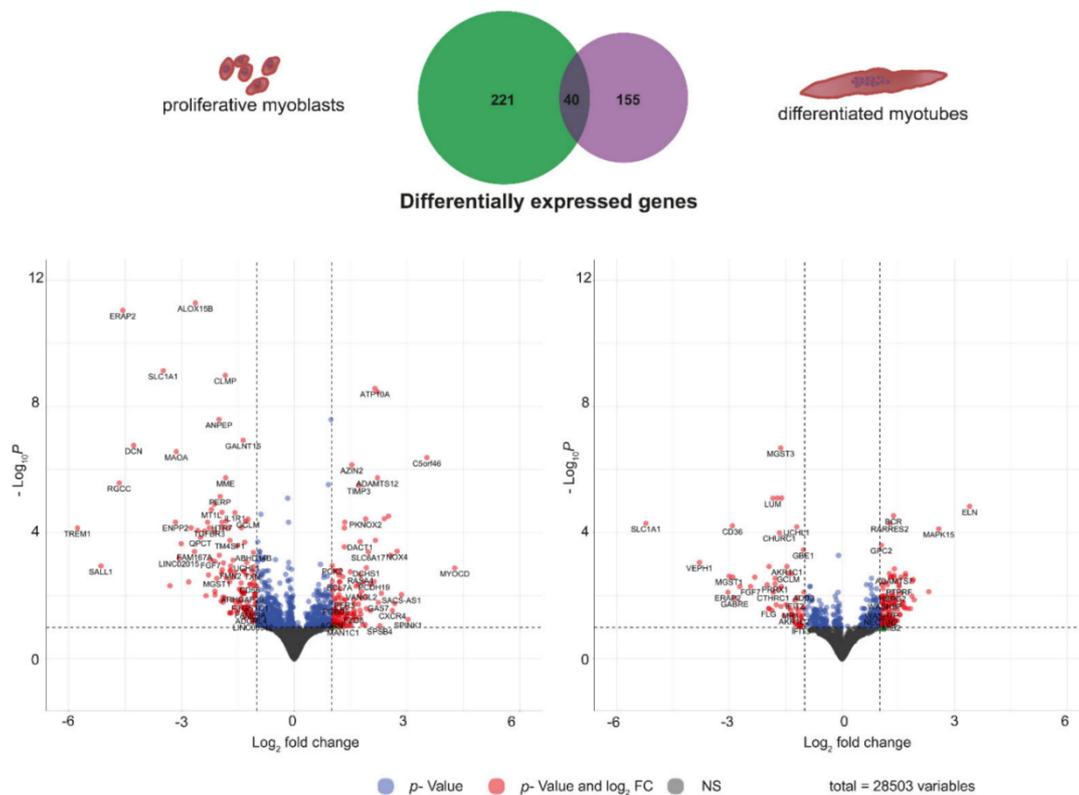


Figure 1. Differential gene expression between non-disease control and DM1 for proliferating myoblasts (left) and post-mitotic myotubes (right) and visualization of the number of genes with a log₂ fold change > ±1 for both conditions (at the top).

2.2. Pathway Enrichment Analysis: Gene Ontology “Molecular Function”

First, we were interested in the molecular functions of the differential expressed genes. DAVID GO molecular function enrichment analysis revealed that myoblasts and myotubes share enrichment of calcium (p -value 0.003 and 0.01) and actin binding (0.09 and 0.002) as well as glutathione transferase activity (0.00004 and 0.04) (Figure 2). Genes found in the calcium binding cluster represent sarcomeric proteins like actinin (*ACTN4*) or calcium homeostasis factors like *ASPH*, but mostly genes related to cell adhesion and the extracellular matrix vital for successful differentiation, development of the neuromuscular junction as well as wound healing and regeneration. This includes cadherines and protocadherines (*CDH15*, *PCDH7/19*), as well as glycoproteins such as agrin (*AGRN*) and proteoglycans (*SPOCK1*, *HSPG2*). Actin binding is more enriched in myotubes than in myoblasts which can be explained by increasing demands of the cellular morphology and the development towards sarcomeric structure. Consistently, genes of the spectrin–actin complex like *SPTBN5* and *ADD3* but also heavy myosin (*MYH9*) can be found in myotubes. Glutathione transferase activity is best known for its function in detoxification and response to reactive oxygen species. Here, several subunits of the glutathione S-transferase (*GST*) are differentially expressed in both myoblasts and myotubes. Further, molecular functions only enriched in myoblasts include PDZ domain binding, indicating a role in signalling pathways, and metal ion binding including genes like Na^+/K^+ ATPases (*ATP1A2*, *ATP1A4*) and RAS p21 protein activator (*RASA4*). In contrast, in DM1 myotubes there is an enrichment of histone binding genes like *H4C11-15* as well as oxidoreductase activity highlighting possible changes in metabolism (Figure 2).

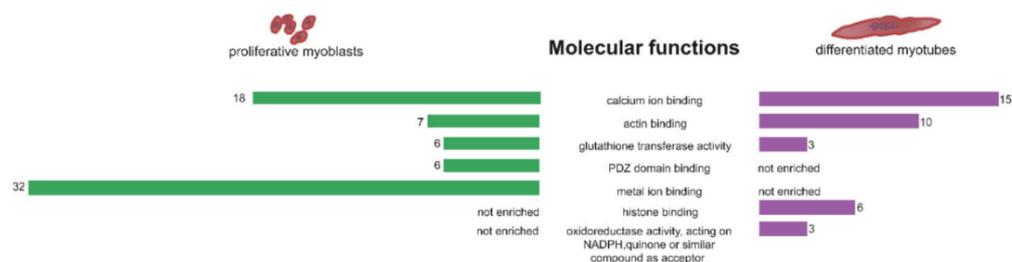


Figure 2. GO enrichment analysis for molecular function of genes differentially expressed between non-disease control and DM1. Significantly enriched molecular functions are shown for proliferating myoblasts (left, green bars) and post-mitotic myotubes (right, purple bars) with the number of genes in the respective molecular function displayed next to the bars.

2.3. Pathway Enrichment Analysis: Gene Ontology “Signalling”

As there are reports of altered signalling in myoblasts gained from embryos with congenital DM1 [54], we were looking directly into signalling alterations. We could identify several genes involved in signalling pathways that are misregulated in DM1. This misregulation was more pronounced in myoblasts than in myotubes (Figure 3). Amongst the identified gene ontology pathways related to signalling pathways were kinase B-, MAPK-, WNT and PI3K signalling—all in myoblasts. Decorin (*DCN*) is downregulated in DM1 myoblasts ($\log_2\text{FC} = -4.3$) and is reported to positively regulate muscle differentiation and regeneration through signalling [58]. The NADPH oxidase (*NOX4*, $\log_2\text{FC} = 2.7$) generates specific reactive oxygen species (ROS) which are important signalling molecules and were shown to influence Ca^{2+} release through the ryanodine receptor (*RYR1*), among others [59]. In myotubes, identified pathways include apoptotic signalling, DNA damage response, beta-catenin TCF complex assembly as well as GTPase activity. Angiopoietin 1 (*ANGPT1*), which is downregulated in DM1 myotubes ($\log_2\text{FC} = -2.9$), plays a role in muscle regeneration, as described in murine muscle [60]. *PLEKHG5*, which is upregulated in DM1 myotubes ($\log_2\text{FC} = 1.2$), is an activator of the nuclear factor kappa B (*NFKB1*)

signalling pathway, which modulates the switch between muscle proliferation and differentiation, and inhibits late-stage differentiation by silencing myofibrillar gene transcription. Misregulated genes also include the transcription factors *MYOCD* and *SALL1* (Figure 3).

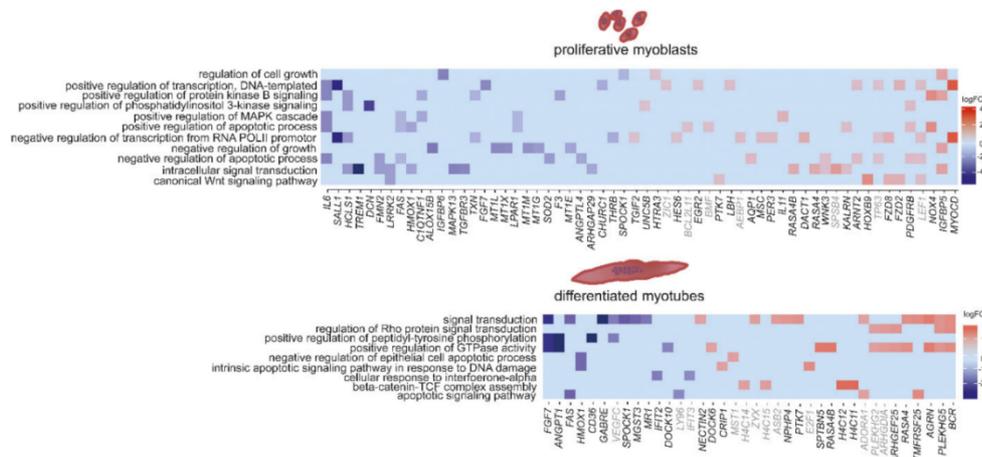


Figure 3. Heat maps showing differences in expression between non-disease controls and DM1 for genes linked to signalling pathways in proliferating myoblasts (**upper panel**) and differentiated myotubes (**lower panel**). Gene names displayed in grey: p -value < 0.1, in black: p -value < 0.05.

2.4. Pathway Enrichment Analysis: Gene Ontology “Others”

Apart from the interest in signalling, we analysed the data according to the top hits in pathway enrichment. This revealed, similar to signalling, similarities and differences between myoblasts and myotubes (Figure 4). In both, extracellular matrix organization is amongst the top hits. The above mentioned *DCN* is a member of the small leucine-rich proteoglycan (SLRP) family regulating collagen fibril assembly and thus, is a key modulator of ECM assembly. The ECM is closely connected to signalling pathways, ultimately activating myogenic factors like *MYOD* and *MYOG*, at the same time however, it is necessary for myoblast fusion. Here, genes like *DCN*, *LAMC2* ($\log_2FC = -2.7$) and *LOX* ($\log_2FC = 1.7$) which build the ECM are downregulated in myoblasts, while genes inducing ECM breakdown like *MMP11* are upregulated ($\log_2FC = 2.4$). DM1 myotubes show decreased levels of lumican (*LUM*), which belongs to the SLRP family and has similar functions to *DCN*. In both myoblasts and myotubes, collagens, cadherines and protocadherines are upregulated, while elastin (*ELN*) is strongly upregulated in myotubes only ($\log_2FC = 3.4$).

In myoblasts there are also pathways linked to mitochondria (protein insertion into mitochondrial membrane involved in apoptotic signal pathways, regulation of mitochondrial depolarization, ATP hydrolysis-coupled proton transport and the oxidation–reduction process), but also pathways linked to wound healing and cytokine-mediated signalling. In myotubes the top hits include chromatin organization and silencing, cellular response to UV-B, actin cytoskeleton organization and inflammatory response (Figure 4).

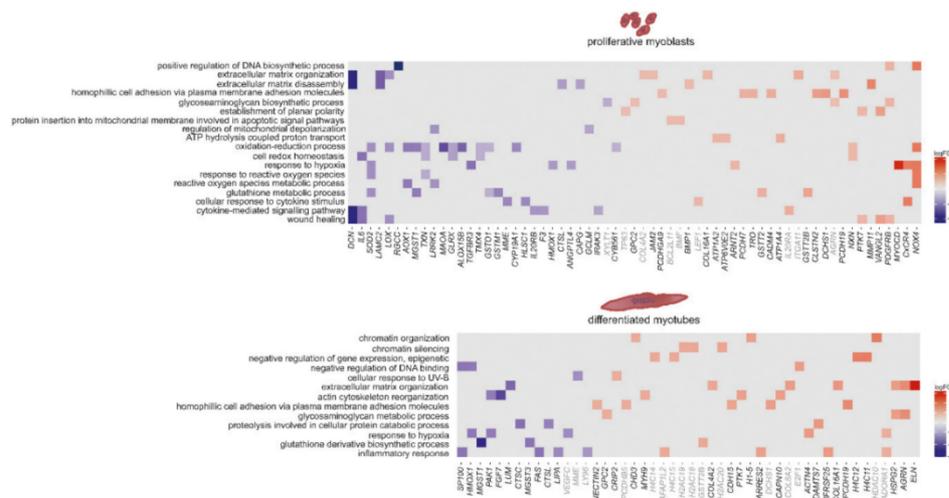


Figure 4. Heat maps showing differences in expression between DM1 and non-disease controls for the top enriched pathways based on GO-terms in proliferating myoblasts (**upper panel**) and differentiated myotubes (**lower panel**). Gene names displayed in grey: p -value < 0.1, in black: p -value < 0.05.

2.5. Splicing as the Basis of Pathway Alterations

Missplicing is well described for DM1. We confirmed the presence of alternative splicing events in our tissue cultures on selected genes, including *MBNL1* and *TNNT2*. *LMO7* and *TEAD1*, both encoding transcription factors, are misspliced preferentially in myotubes respectively myoblasts (Figure S3). With the specific aim of investigating splicing alterations, we performed a test for gene set enrichment (GSEA) using the *fgsea* R package (results in Tables S4 and S5). Splicing and especially alternative splicing is a very sensitive process at the basis of gene expression regulation which has different outcomes depending on spliceosome assembly, splice site recognition and numerous other factors (reviewed in [61]). Consistently, it was shown that expression changes of only twofold of certain splicing factors such as *SRSF1* can implicate strong effects on proliferation and apoptosis of more than twofold in cell culture [62]. We thus expect that small changes of a high number of splicing factors have broad implications, as spliceosome assembly and composition are affected. Accordingly, we analysed splicing associated genes with fold changes between 1.1 and 2/−1.1 and −2 to reveal if specific splicing pathways are affected in DM1 (Figure 5). Our data suggest that genes linked to alternative splicing via spliceosome, spliceosomal complex assembly, and mRNA 5' splice site recognition are misregulated in both DM1 myoblasts and myotubes as well as adult muscle. Further, genes associated with positive regulation of mRNA splicing via spliceosome are downregulated in myoblasts, while in myotubes, pathways linked to mRNA splice site selection, mRNA 3' splice site recognition, and negative regulation of RNA splicing are enriched. In muscle biopsies, we found all of these pathways affected. However, most striking was the clear cut-off between up- and downregulation. Pathways generally linked to constitutive splicing (complex assembly) are downregulated, while alternative splicing is upregulated. Notably, core components of the spliceosome machinery are underrepresented in our data and the muscle biopsy data; most splicing related genes belong to the category of accessory components important for recognition of specific splice sites rather than the splicing reaction itself. This might be a reason for the strong effects of DM1 on specific tissue types and transcripts rather than splicing in general. Members of the RNA-binding motif (RBM) family are upregulated in myoblasts, myotubes and adult muscle and play roles in alternative splicing via splice site

selection. Interestingly, several studies have shown that, for example, RBM5 regulates the cell cycle and apoptosis while RBM24 is important for striated muscle differentiation, thus integrating well in our analysis so far [63,64]. Survival motor neurons 1 and 2 (*SMN1/2*) are downregulated in both proliferating and differentiated muscle cells twofold. Deficiency of *SMN1/2* is usually associated with spinal muscular atrophy; however, it was shown that low levels influence muscle differentiation and maturation through the AKT signalling pathway [65]. Similar involvements in proliferation, apoptosis and differentiation are known for many of the genes displayed in Figure 5.

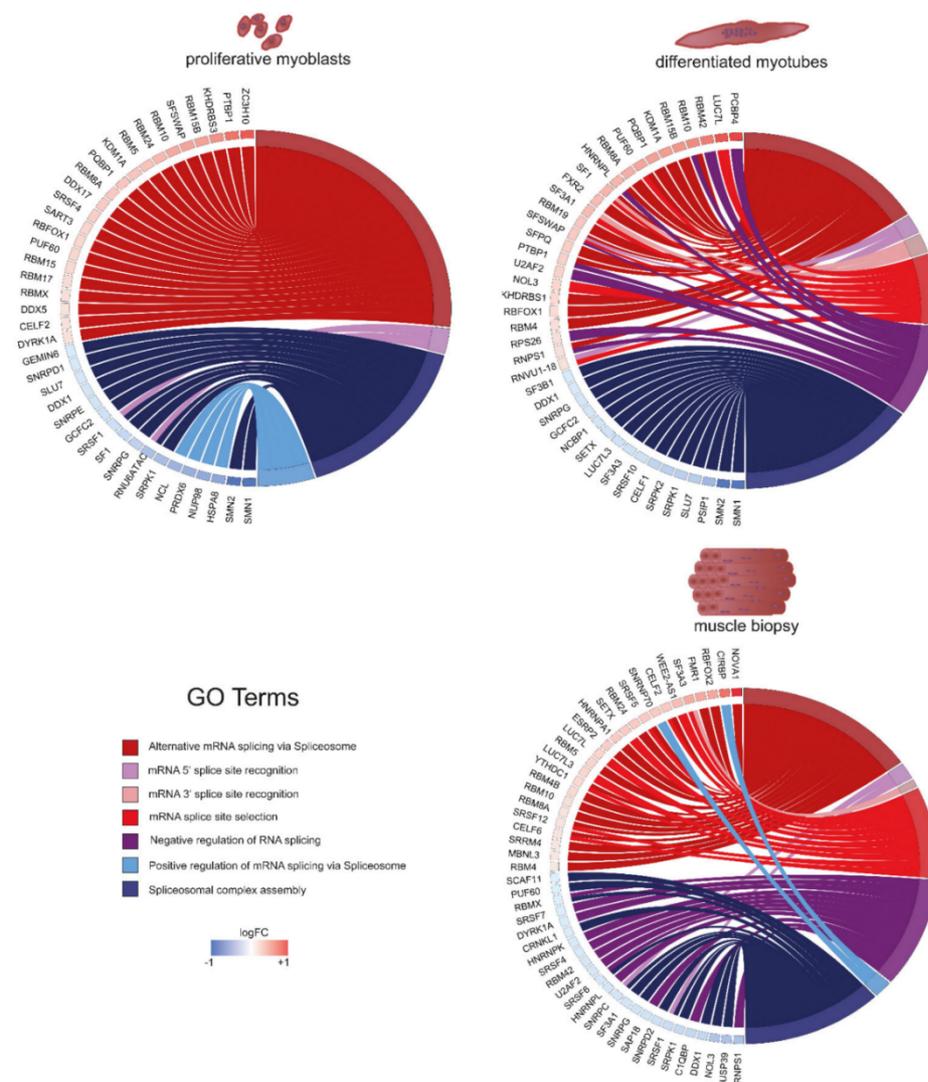


Figure 5. Graphs showing differences in expression between DM1 and non-disease controls for genes linked to splicing in proliferating myoblasts (upper left), differentiated myotubes (upper right), and muscle biopsies (lower right), sequencing data published by [57]. The colour of the boxes next to the gene name shows the level of expression change compared to controls. Genes are grouped depending on their GO-term.

3. Discussion

There is a plethora of data sets regarding missplicing in DM1, yet, it remains elusive which other pathomechanisms may contribute to this complex disease and how they might interconnect. We investigated gene expression in a tissue culture differentiation system using primary human myoblasts collected from DM1 patients and non-disease controls. This is highly promising in providing insight concerning the development of novel treatments and helps to answer the question of whether it will be sufficient to target mature muscle, or if it will be necessary to aim for satellite cells as well.

A general trend we observed, is that there are indeed differences in differentially expressed genes depending on the cellular differentiation status. While this is not surprising, looking specifically into affected pathways provides additional information on the disease etiopathogenesis. A possible contribution to these expression changes could be missplicing of transcription factors. Signalling pathways have been shown to be affected in DM1 before, explicitly in myoblasts gained from embryos with congenital DM1 [54]. Activity of the kinases protein kinase B (AKT), MAPK and ERK was downregulated in those samples. Here, we investigated myoblasts gained from adults with myotonic dystrophy and identified genes involved in the MAPK cascade, protein kinase B (AKT), WNT and PI3K signalling to be misregulated. While this supports results published on congenital DM, the lack of a misregulation of these genes in differentiated myotubes indicates the possibility of a pathomechanism specific for proliferating muscle cells.

Premature senescence has been described for DM1 primary myoblasts [66]. We previously noticed a repeat-length-dependent increase of nuclear envelope invaginations in primary human adult onset DM1 myoblasts, which was accompanied by cell cycle withdrawal of affected cells [67]. We hypothesize that these membrane invaginations are a result of missplicing events in nuclear envelope proteins like nesprin1 [68] and may cause further gene expression changes due to alterations of heterochromatin organization. Conceivably, this contributes to signalling pathways affected in proliferating myoblasts, resulting in a loss of a certain proportion of proliferating cells due to cell cycle defects. In both myoblasts and myotubes, we see strong alterations of genes involved in extracellular matrix (ECM) organization. The ECM is also closely connected to signalling pathways, ultimately activating myogenic factors like MYOD and MYOG, at the same time however, it is necessary for myoblast fusion and thus differentiation into mature myotubes.

Differentiation defects in DM1 cells [56] and patient muscle biopsies suggest a lack of fibre maturation [69]. To investigate this in more detail we decided to isolate differentiated myotubes by minimal trypsin treatment and sedimentation. While there is a potential loss of material, isolating differentiated myotubes has the benefit of avoiding contamination of undifferentiated cells. This seems especially important for a disease with known differentiation defects. The DM1 phenotype correlates approximately with CTG repeat length: the longer the repeat, the more severe the phenotype. We saw the same effect in tissue culture: for the sample with the longest repeat (DM1-3), we failed to isolate enough differentiated myotubes to get sufficient RNAseq reads for analysis. Using the data available for myotubes, we found an enrichment of genes involved in apoptosis, DNA damage and repair, histone binding, chromatin organization, and metabolic pathways. This indicates that the pathological effects in myotubes are different from myoblasts and increased stress is affecting the fitness of the cells.

Notably, we found the expression of genes linked to pathways involved in responses to reactive oxygen and inflammation being affected in proliferating and differentiated DM1 cells. This, if also affected in mature muscle fibres, might contribute to the observed fibrosis in DM1 [70].

The alternative splicing of several genes caused by MBNL/CELF misregulation is well described to drive the pathology of DM1 [71], and other splicing factors seem to be involved as well [51,52]. Importantly, our data suggest a considerably higher complexity of the splicing machinery involved than previously anticipated. Despite being rather mildly misregulated (log₂FC-wise), we can see whole groups of genes linked to alternative

splicing being upregulated, and the spliceosomal complex assembly downregulated in myoblasts as well as myotubes. Yet, there are differences between myoblasts and myotubes regarding genes linked to other splicing related GO-terms. In myotubes, genes involved in 5' and 3' splice site recognition are upregulated. A similar picture can be seen for muscle biopsies [57]. This indicates an altered formation of entire splicing complexes which are assembled specifically depending on the differentiation stage—and in turn result in a differentiation stage dependent specific misregulation in DM1. Additional work is required to further understand the mechanism of this splicing regulation.

In summary, our data suggest that developmental stage-specific missplicing of a multitude of genes affects different pathways in a stage-dependent manner, and moreover, that these pathways are highly interconnected (Figure 6). Based on our observations, we propose that missplicing-induced alterations of gene expression and signalling result in proliferation, ECM and cytoskeleton misregulation in DM1 myoblasts, ultimately leading to both cell cycle and differentiation defects, the first of which again influences the second as fewer myoblasts are present for fusion. In DM1 myotubes, which consequently are smaller and fewer in number compared to their healthy counterparts, we already observe inflammatory responses, metabolic defects and DNA damage responses, forecasting the mature DM1 phenotype. We conclude that it is vital for future therapeutic interventions to target muscle stem cells, which are the precursor cells for myoblasts and myotubes, in addition to mature muscle.

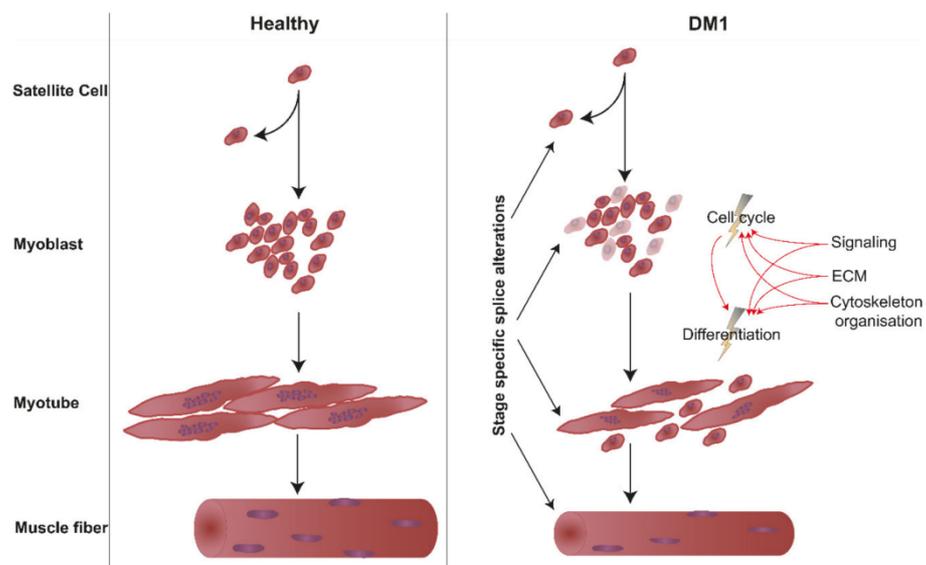


Figure 6. Hypothetical effect of the observed misregulated pathways on the cell cycle and differentiation leading to the DM1 phenotype.

4. Materials and Methods

4.1. Culture of Primary Skeletal Muscle Cells

Primary human myoblasts were cultured in Skeletal Muscle Growth Medium (PELO-Biotech, Munich, Germany) in an incubator at 37 °C and 5% CO₂. The growth medium was supplemented with GlutaMax, 40 U/mL Penicillin, 0.04 mg/mL Streptomycin, and SkMC Supplement (PELOBiotech, Munich, Germany). Myoblasts were kept from reaching confluency to avoid differentiation. Passage numbers were matched for controls and patient cells.

For differentiation, confluent myoblasts were cultivated for 7 days in DMEM containing 5% horse serum under the same environmental conditions.

4.2. Patients and Non-Disease Controls

Human control and patient materials (Table 1) were obtained with written informed consent of the respective donors from the Muscle Tissue Culture Collection (MTCC) at the Friedrich Baur Institute (Department of Neurology, LMU Klinikum, Ludwig Maximilians University, Munich, Germany). Ethical approval for this study was obtained from the ethical review committee at the Ludwig Maximilians University, Munich, Germany (IRB-no. reference 45-14; 4 August 2014).

Table 1. List of patient and control tissue cultures.

	Phenotype	Age at Biopsy	Sex	CTG-Repeat Length
DM1-1	DM1	27	M	400–600
DM1-2	DM1	34	M	240–430
DM1-3	DM1	29	F	800–1500
Control-1	—	32	M	n.d.
Control-2	—	49	F	n.d.
Control-3	—	49	F	n.d.

n.d. = not determined; M = male; F = female.

4.3. Myotube Separation

Myotubes were separated from undifferentiated, mononucleated cells by a minimal trypsin treatment to detach the myotubes. This was followed by gentle centrifugation to sediment myotubes and separate them from remaining mononucleated cells. All steps were controlled by light microscopy to observe the success of the separation.

4.4. RNA Isolation

After two washing steps with 1x PBS Trizol[®] was added to the samples (material gained from myotube separation respectively plates with myoblasts). The Trizol[®]/sample mixture was used directly to isolate the RNA using the Dircet-zol[™] RNA MiniPrep Plus Kit (ZYMO Research, Freiburg, Germany).

4.5. RNA Sequencing

Library preparation was performed using the TruSeq Stranded mRNA Kit (Illumina, San Diego, CA, USA) with the TruSeq RNA Single Index Set A (Illumina) according to the recommended procedure. Quality and size distribution of the generated libraries has been validated using an Agilent 2100 bioanalyzer with high-sensitivity DNA chip (Agilent, Santa Clara, CA, USA) and DNA yield was measured using the Qubit dsDNA HS Assay Kit followed by pooling of the libraries in batches of 12 and sequencing of 1.2 pM pooled library on an Nextseq 500/High Output Flow Cell Cartridge using an paired end, 2 × 76 reads, single index protocol.

4.6. RT-qPCR

Reverse transcription of RNA was performed using the QuantiTect Reverse Transcription Kit (Qiagen, Venlo, Netherlands). For the reaction we used the SYBR[®] Green Master Mix (Bio-Rad, Hercules, CA, USA) and samples were run and measured on CFX Connect[™] (Bio-Rad). As reference gene, *Ap3d1* was used as it has been shown to be a suitable reference gene in normal and dystrophic cell culture models of myogenesis [72]. Samples were analysed using the delta Ct method and samples were normalized to the mean of the controls. Primer sequences are listed in Table S6.

4.7. Immunofluorescence Staining and Microscopy

Myoblasts were fixed with $-20\text{ }^{\circ}\text{C}$ cold methanol. Following primary antibodies were used for staining: MBNL1 (HPA035098, atlas antibodies), Emerin 5D10 (provided by Glenn E. Morris). Secondary antibodies were Alexa Fluor conjugated and generated in donkey with minimal species cross-reactivity. For DNA visualization DAPI (4',6-diamidino-2-phenylindole) was used.

Images were obtained using an Olympus IX83 inverted microscope equipped with a $100\times$ objective and a digital camera (UC90, Olympus, Shinjuku, Japan).

4.8. Quantification of Splicing Alterations

Splice events covering exon–exon junctions were visualized using Sashimi plots in the IGV genome browser. Reads spanning exon–exon junctions for the respective splice events were counted for controls and DM1 samples and visualized as percentage.

4.9. Bioinformatics

Reads were aligned to human genome assembly hg38 using STAR (version 2.6.1b) [73] and processed using deeptools (version 3.0.2) [74]. All subsequent analysis steps were conducted in R 4.0.4. BAM files were counted using featureCounts and then analysed with DESeq2 [75,76]. Normalized reads were used for quality control checks (Figures S4 and S5) which highlighted the relatively high variance of biological replicates, especially those of non-disease controls, ultimately leading to a lower number of genes with high significance and with this, a loss of biologically relevant information. Principal component analysis (PCA, Figure S4) further shows that there are more differences between undifferentiated and differentiated than between disease and non-disease samples, which is to be expected. Moreover, biological replicates more-or-less group together as expected. We decided to screen for biological meaningful results using standard approaches (DESeq2 + DAVID GO enrichment analysis) as well as an internal ranking as used in gene set enrichment analysis (GSEA, Tables S6 and S7). For both, we used the \log_2 fold changes generated in DESeq2 transformed with lfcShrink to account for high \log_2 fold changes [77]. For standard analysis, we set $\log_2\text{FC} > \pm 1$ and $p\text{-value} < 0.1$ to be significantly differentially expressed between samples, the latter of which can be reasoned by using biological rather than technical replicates resulting in higher variance, naturally. For splicing analysis however, we used $\log_2\text{FCs}$ between -1 and 1 , but higher than 0.1 . Visualization was performed using packages listed in Supplemental Table S7.

4.10. Muscle Biopsy Data Analysis

Muscle biopsy data for DM1 and controls was previously published by [57]. Raw sequencing data for four controls and three DM1 samples was downloaded from GEO (GSE86356) and analysed by following the same workflow for myoblasts and myotubes described above, including all parameters and cut-offs. Accordingly, data quality was confirmed through dispersion estimation and PCA (Figures S6 and S7). Additionally, normalized count reads are listed in Table S8, while DESeq2 and fgsea results can be found in Tables S9 and S10.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms22168607/s1>.

Author Contributions: S.H. performed the experiments, A.H. the sequencing and V.T. and A.R.W.K. the bioinformatical analyses. S.H., V.T., P.M. and B.S. were involved in conception and design of the experiments, in analysis and interpretation of the data and contributed to the writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the generosity of Dr. Metin Colpan and by the FöFoLe Reg.-Nr. 959 (LMU).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the ethical review committee at the Ludwig Maximilians University, Munich, Germany (IRB-no. reference 45-14; 4 August 2014).

Informed Consent Statement: Informed written consent was obtained from all subjects involved in the study.

Data Availability Statement: Processed data on DM1 and control myoblasts and myotubes can be found in the Supplemental Materials. Raw data are available on request from the corresponding author. Raw data on DM1 and control muscle biopsies are publicly available at GEO (GSE86356).

Acknowledgments: The Muscular Dystrophy Association (USA) supports monoclonal antibody development in the laboratory of Glenn E. Morris, whom we thank for providing the emerin antibody.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Thornton, C.A. Myotonic dystrophy. *Neurol. Clin.* **2014**, *32*, 705–719. [[CrossRef](#)] [[PubMed](#)]
- Day, J.W.; Roelofs, R.; Leroy, B.; Pech, I.; Benzow, K.; Ranum, L.P. Clinical and genetic characteristics of a five-generation family with a novel form of myotonic dystrophy (DM2). *Neuromuscul. Disord.* **1999**, *9*, 19–27. [[CrossRef](#)]
- Ricker, K.; Koch, M.C.; Lehmann-Horn, F.; Pongratz, D.; Speich, N.; Reiners, K.; Schneider, C.; Moxley, R.T. Proximal myotonic myopathy. Clinical features of a multisystem disorder similar to myotonic dystrophy. *Arch. Neurol.* **1995**, *52*, 25–31. [[CrossRef](#)]
- Wenninger, S.; Montagnese, F.; Schoser, B. Core Clinical Phenotypes in Myotonic Dystrophies. *Front. Neurol.* **2018**, *9*, 303. [[CrossRef](#)]
- Wheeler, T.M. Myotonic dystrophy: Therapeutic strategies for the future. *Neurotherapeutics* **2008**, *5*, 592–600. [[CrossRef](#)] [[PubMed](#)]
- Faustino, N.A.; Cooper, T.A. Pre-mRNA splicing and human disease. *Genes Dev.* **2003**, *17*, 419–437. [[CrossRef](#)]
- Fu, Y.H.; Pizzuti, A.; Fenwick, R.G.; King, J.; Rajnarayan, S.; Dunne, P.W.; Dubel, J.; A Nasser, G.; Ashizawa, T.; De Jong, P.; et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **1992**, *255*, 1256–1258. [[CrossRef](#)]
- Brook, J.D.; McCurrach, M.E.; Harley, H.G.; Buckler, A.J.; Church, D.; Aburatani, H.; Hunter, K.; Stanton, V.P.; Thirion, J.-P.; Hudson, T.; et al. Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **1992**, *68*, 799–808. [[CrossRef](#)]
- De Antonio, M.; Dogan, C.; Hamroun, D.; Mati, M.; Zerrouki, S.; Eymard, B.; Katsahian, S.; Bassez, G.; French Myotonic Dystrophy Clinical Network. Unravelling the myotonic dystrophy type 1 clinical spectrum: A systematic registry-based study with implications for disease classification. *Rev. Neurol.* **2016**, *172*, 572–580. [[CrossRef](#)]
- De Temmerman, N.; Sermon, K.; Seneca, S.; De Rycke, M.; Hilven, P.; Lissens, W.; Van Steirteghem, A.; Liebaers, I. Intergenerational instability of the expanded CTG repeat in the DMPK gene: Studies in human gametes and preimplantation embryos. *Am. J. Hum. Genet.* **2004**, *75*, 325–329. [[CrossRef](#)] [[PubMed](#)]
- Redman, J.B.; Fenwick, R.G.; Fu, Y.H.; Pizzuti, A.; Caskey, C.T. Relationship between parental trinucleotide GCT repeat length and severity of myotonic dystrophy in offspring. *JAMA* **1993**, *269*, 1960–1965. [[CrossRef](#)] [[PubMed](#)]
- Jansen, G.; Willems, P.; Coerwinkel, M.; Nillesen, W.; Smeets, H.; Vits, L.; Höweler, C.; Brunner, H.; Wieringa, B. Gonosomal mosaicism in myotonic dystrophy patients: Involvement of mitotic events in (CTG)_n repeat variation and selection against extreme expansion in sperm. *Am. J. Hum. Genet.* **1994**, *54*, 575–585. [[PubMed](#)]
- Rakočević-Stojanović, V.; Savic, D.; Pavlović, S.; Lavrnić, D.; Stević, Z.; Basta, I.; Romac, S.; Apostolski, S. Intergenerational changes of CTG repeat depending on the sex of the transmitting parent in myotonic dystrophy type 1. *Eur. J. Neurol.* **2005**, *12*, 236–237. [[CrossRef](#)] [[PubMed](#)]
- Martorell, L.; Cobo, A.M.; Baiget, M.; Naudó, M.; Poza, J.J.; Parra, J. Prenatal diagnosis in myotonic dystrophy type 1. Thirteen years of experience: Implications for reproductive counselling in DM1 families. *Prenat. Diagn.* **2006**, *27*, 68–72. [[CrossRef](#)] [[PubMed](#)]
- Thornton, C.A.; Johnson, K.; Moxley, R.T. Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann. Neurol.* **1994**, *35*, 104–107. [[CrossRef](#)]
- Ishii, S.; Nishio, T.; Sunohara, N.; Yoshihara, T.; Takemura, K.; Hikiji, K.; Tsujino, S.; Sakuragawa, N. Small increase in triplet repeat length of cerebellum from patients with myotonic dystrophy. *Qual. Life Res.* **1996**, *98*, 138–140. [[CrossRef](#)]
- Napierala, M.; Krzyzosiak, W.J. CUG repeats present in myotonin kinase RNA form metastable “slippery” hairpins. *J. Biol. Chem.* **1997**, *272*, 31079–31085. [[CrossRef](#)]
- Dere, R.; Napierala, M.; Ranum, L.P.W.; Wells, R.D. Hairpin Structure-forming Propensity of the (CCTG-CAGG) Tetranucleotide Repeats Contributes to the Genetic Instability Associated with Myotonic Dystrophy Type 2. *J. Biol. Chem.* **2004**, *279*, 41715–41726. [[CrossRef](#)]
- Mankodi, A.; Logigian, E.; Callahan, L.; McClain, C.; White, R.; Henderson, D.; Krym, M.; Thornton, C.A. Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* **2000**, *289*, 1769–1773. [[CrossRef](#)] [[PubMed](#)]
- Pettersson, O.J.; Aagaard, L.; Jensen, T.G.; Damgaard, C.K. Molecular mechanisms in DM1—A focus on foci. *Nucleic Acids Res.* **2015**, *43*, 2433–2441. [[CrossRef](#)]

21. Ho, T.; Bundman, D.; Armstrong, D.L.; Cooper, T.A. Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Hum. Mol. Genet.* **2005**, *14*, 1539–1547. [CrossRef]
22. López-Martínez, A.; Soblechero-Martín, P.; de-la-Puente-Ovejero, L.; Nogales-Gadea, G.; Arechavala-Gomez, V. An Overview of Alternative Splicing Defects Implicated in Myotonic Dystrophy Type I. *Genes* **2020**, *11*, 1109. [CrossRef]
23. Savkur, R.S.; Philips, A.V.; Cooper, T.A. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet.* **2001**, *29*, 40–47. [CrossRef]
24. Du, H.; Cline, M.S.; Osborne, R.J.; Tuttle, D.L.; Clark, T.A.; Donohue, J.P.; Hall, M.P.; Shiue, L.; Swanson, M.S.; Thornton, C.A. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat. Struct. Mol. Biol.* **2010**, *17*, 187–193. [CrossRef]
25. Ebralidze, A.; Wang, Y.; Petkova, V.; Ebralidze, K.; Junghans, R.P. RNA Leaching of Transcription Factors Disrupts Transcription in Myotonic Dystrophy. *Science* **2004**, *303*, 383–387. [CrossRef]
26. Osborne, R.J.; Lin, X.; Welle, S.; Sobczak, K.; O'Rourke, J.R.; Swanson, M.; Thornton, C.A. Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy. *Hum. Mol. Genet.* **2009**, *18*, 1471–1481. [CrossRef]
27. Kalsotra, A.; Singh, R.; Gurha, P.; Ward, A.J.; Creighton, C.J.; Cooper, T.A. The Mef2 Transcription Network Is Disrupted in Myotonic Dystrophy Heart Tissue, Dramatically Altering miRNA and mRNA Expression. *Cell Rep.* **2014**, *6*, 336–345. [CrossRef]
28. Huichalaf, C.; Sakai, K.; Jin, B.; Jones, K.; Wang, G.; Schoser, B.; Schneider-Gold, C.; Sarkar, P.; Pereira-Smith, O.M.; Timchenko, N.; et al. Expansion of CUG RNA repeats causes stress and inhibition of translation in myotonic dystrophy 1 (DM1) cells. *FASEB J.* **2010**, *24*, 3706–3719. [CrossRef]
29. Meola, G.; Jones, K.; Wei, C.; Timchenko, L.T. Dysfunction of protein homeostasis in myotonic dystrophies. *Histol. Histopathol.* **2013**, *28*. Available online: <http://hdl.handle.net/10201/60219> (accessed on 7 August 2021).
30. Cleary, J.D.; Ranum, L.P. Repeat associated non-ATG (RAN) translation: New starts in microsatellite expansion disorders. *Curr. Opin. Genet. Dev.* **2014**, *26*, 6–15. [CrossRef]
31. Gudde, A.; van Heeringen, S.J.; de Oude, A.I.; van Kessel, I.D.G.; Estabrook, J.; Wang, E.T.; Wieringa, B.; Wansink, D.G. Antisense transcription of the myotonic dystrophy locus yields low-abundant RNAs with and without (CAG)_n repeat. *RNA Biol.* **2017**, *14*, 1374–1388. [CrossRef]
32. Zu, T.; Gibbens, B.; Doty, N.S.; Gomes-Pereira, M.; Huguet, A.; Stone, M.D.; Margolis, J.; Peterson, M.; Markowski, T.W.; Ingram, M.A.; et al. Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 260–265. [CrossRef]
33. Batra, R.; Charizanis, K.; Manchanda, M.; Mohan, A.; Li, M.; Finn, D.J.; Goodwin, M.; Zhang, C.; Sobczak, K.; Thornton, C.A.; et al. Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol. Cell* **2014**, *56*, 311–322. [CrossRef]
34. Krol, J.; Fiszler, A.; Mykowska, A.; Sobczak, K.; de Mezer, M.; Krzyzosiak, W.J. Ribonuclease Dicer Cleaves Triplet Repeat Hairpins into Shorter Repeats that Silence Specific Targets. *Mol. Cell* **2007**, *25*, 575–586. [CrossRef] [PubMed]
35. Rau, F.; Freyermuth, F.; Fugier, C.; Villemin, J.-P.; Fischer, M.-C.; Jost, B.; Dembélé, D.; Gourdon, G.; Nicole, A.; Duboc, D.; et al. Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat. Struct. Mol. Biol.* **2011**, *18*, 840–845. [CrossRef] [PubMed]
36. Shen, X.; Xu, F.; Li, M.; Wu, S.; Zhang, J.; Wang, A.; Xu, L.; Liu, Y.; Zhu, G. miR-322/-503 rescues myoblast defects in myotonic dystrophy type 1 cell model by targeting CUG repeats. *Cell Death Dis.* **2020**, *11*, 891. [CrossRef]
37. Cappella, M.; Perfetti, A.; Cardinali, B.; Garcia-Manteiga, J.M.; Carrara, M.; Provenzano, C.; Fuschi, P.; Cardani, R.; Renna, L.V.; Meola, G.; et al. High-throughput analysis of the RNA-induced silencing complex in myotonic dystrophy type 1 patients identifies the dysregulation of miR-29c and its target ASB2. *Cell Death Dis.* **2018**, *9*, 729. [CrossRef]
38. Thong, M.-K.; Ambrose, K.K.; Ishak, T.; Lian, L.-H.; Goh, K.-J.; Wong, K.-T.; Ahmad-Annuar, A. Deregulation of microRNAs in blood and skeletal muscles of myotonic dystrophy type 1 patients. *Neurol. India* **2017**, *65*, 512–517. [CrossRef]
39. Fritegotto, C.; Ferrati, C.; Pegoraro, V.; Angelini, C. Micro-RNA expression in muscle and fiber morphometry in myotonic dystrophy type 1. *Neurol. Sci.* **2017**, *38*, 619–625. [CrossRef]
40. Fardaei, M.; Rogers, M.T.; Thorpe, H.M.; Larkin, K.; Hamshere, M.G.; Harper, P.S.; Brook, J.D. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* **2002**, *11*, 805–814. [CrossRef]
41. Fardaei, M.; Larkin, K.; Brook, J.D.; Hamshere, M.G. In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res.* **2001**, *29*, 2766–2771. [CrossRef]
42. Philips, A.V.; Timchenko, L.T.; Cooper, T.A. Disruption of Splicing Regulated by a CUG-Binding Protein in Myotonic Dystrophy. *Science* **1998**, *280*, 737–741. [CrossRef]
43. Kuyumcu-Martinez, N.M.; Wang, G.-S.; Cooper, T.A. Increased Steady-State Levels of CUGBP1 in Myotonic Dystrophy 1 Are Due to PKC-Mediated Hyperphosphorylation. *Mol. Cell* **2007**, *28*, 68–78. [CrossRef]
44. Mankodi, A.; Takahashi, M.P.; Jiang, H.; Beck, C.L.; Bowers, W.J.; Moxley, R.T.; Cannon, S.C.; Thornton, C.A. Expanded CUG repeats trigger aberrant splicing of CIC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol. Cell* **2002**, *10*, 35–44. [CrossRef]
45. Charlet, N.; Savkur, R.S.; Singh, G.; Philips, A.V.; Grice, E.; Cooper, T.A. Loss of the Muscle-Specific Chloride Channel in Type 1 Myotonic Dystrophy Due to Misregulated Alternative Splicing. *Mol. Cell* **2002**, *10*, 45–53. [CrossRef]

46. Yamashita, Y.; Matsuura, T.; Shinmi, J.; Amakusa, Y.; Masuda, A.; Ito, M.; Kinoshita, M.; Furuya, H.; Abe, K.; Ibi, T.; et al. Four parameters increase the sensitivity and specificity of the exon array analysis and disclose 25 novel aberrantly spliced exons in myotonic dystrophy. *J. Hum. Genet.* **2012**, *57*, 368–374. [[CrossRef](#)] [[PubMed](#)]
47. Kimura, T.; Nakamori, M.; Lueck, J.D.; Pouliquin, P.; Aoike, F.; Fujimura, H.; Dirksen, R.T.; Takahashi, M.P.; Dulhunty, A.; Sakoda, S. Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca²⁺-ATPase in myotonic dystrophy type 1. *Hum. Mol. Genet.* **2005**, *14*, 2189–2200. [[CrossRef](#)]
48. Fugier, C.; Klein, A.; Hammer, C.; Vassilopoulos, S.; Ivarsson, Y.; Toussaint, A.; Tosch, V.; Vignaud, A.; Ferry, A.; Messaddeq, N.; et al. Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nat. Med.* **2011**, *17*, 720–725. [[CrossRef](#)]
49. Lin, X.; Miller, J.W.; Mankodi, A.; Kanadia, R.N.; Yuan, Y.; Moxley, R.T.; Swanson, M.S.; Thornton, C.A. Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.* **2006**, *15*, 2087–2097. [[CrossRef](#)]
50. Castle, J.C.; Zhang, C.; Shah, J.K.; Kulkarni, A.V.; Kalsotra, A.; Cooper, T.A.; Johnson, J.M. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat. Genet.* **2008**, *40*, 1416–1425. [[CrossRef](#)] [[PubMed](#)]
51. Bondy-Chorney, E.; Crawford Parks, T.E.; Ravel-Chapuis, A.; Klinck, R.; Rocheleau, L.; Pelchat, M.; Chabot, B.; Jasmin, B.J.; Côté, J. Staufen1 Regulates Multiple Alternative Splicing Events either Positively or Negatively in DM1 Indicating Its Role as a Disease Modifier. *PLoS Genet.* **2016**, *12*, e1005827. [[CrossRef](#)] [[PubMed](#)]
52. Li, M.; Zhuang, Y.; Batra, R.; Thomas, J.D.; Li, M.; Nutter, C.A.; Scotti, M.M.; Carter, H.A.; Wang, Z.J.; Huang, X.-S.; et al. HNRNPA1-induced spliceopathy in a transgenic mouse model of myotonic dystrophy. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 5472–5477. [[CrossRef](#)]
53. Botta, A.; Vallo, L.; Rinaldi, F.; Bonifazi, E.; Amati, F.; Biancolella, M.; Gambardella, S.; Mancinelli, E.; Angelini, C.; Meola, G. Gene expression analysis in myotonic dystrophy: Indications for a common molecular pathogenic pathway in DM1 and DM2. *Gene Expr.* **2007**, *13*, 339–351. [[CrossRef](#)] [[PubMed](#)]
54. Beffy, P.; Del Carratore, R.; Masini, M.; Furling, D.; Puymirat, J.; Masiello, P.; Simili, M. Altered signal transduction pathways and induction of autophagy in human myotonic dystrophy type 1 myoblasts. *Int. J. Biochem. Cell Biol.* **2010**, *42*, 1973–1983. [[CrossRef](#)]
55. Mastroyiannopoulos, N.P.; Chrysanthou, E.; Kyriakides, T.C.; Uney, J.B.; Mahadevan, M.S.; Phylactou, L.A. The effect of myotonic dystrophy transcript levels and location on muscle differentiation. *Biochem. Biophys. Res. Commun.* **2008**, *377*, 526–531. [[CrossRef](#)] [[PubMed](#)]
56. Furling, D.; Coiffier, L.; Mouly, V.; Barbet, J.P.; Guily, J.L.S.; Taneja, K.; Gourdon, G.; Junien, C.; Butler-Browne, G.S. Defective satellite cells in congenital myotonic dystrophy. *Hum. Mol. Genet.* **2001**, *10*, 2079–2087. [[CrossRef](#)]
57. Wang, E.T.; Treacy, D.; Eichinger, K.; Struck, A.; Estabrook, J.; Olafson, H.; Wang, T.T.; Bhatt, K.; Westbrook, T.; Sedehizadeh, S.; et al. Transcriptome alterations in myotonic dystrophy skeletal muscle and heart. *Hum. Mol. Genet.* **2018**, *28*, 1312–1321. [[CrossRef](#)]
58. Li, Y.; Li, J.; Zhu, J.; Sun, B.; Branca, M.; Tang, Y.; Foster, W.; Xiao, X.; Huard, J. Decorin gene transfer promotes muscle cell differentiation and muscle regeneration. *Mol. Ther.* **2007**, *15*, 1616–1622. [[CrossRef](#)]
59. Cully, T.R.; Rodney, G.G. Nox4-RyR1-Nox2: Regulators of micro-domain signaling in skeletal muscle. *Redox Biol.* **2020**, *36*, 101557. [[CrossRef](#)] [[PubMed](#)]
60. Mofarrah, M.; McClung, J.M.; Kontos, C.D.; Davis, E.C.; Tappuni, B.; Moroz, N.; Pickett, A.E.; Huck, L.; Harel, S.; Danialou, G.; et al. Angiopoietin-1 enhances skeletal muscle regeneration in mice. *Am. J. Physiol. Integr. Comp. Physiol.* **2015**, *308*, R576–R589. [[CrossRef](#)]
61. Chen, M.; Manley, J.L. Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 741–754. [[CrossRef](#)] [[PubMed](#)]
62. Anczukow, O.; Rosenberg, A.; Akerman, M.; Das, S.; Zhan, L.; Karni, R.; Muthuswamy, S.; Krainer, A.R. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.* **2012**, *19*, 220–228. [[CrossRef](#)]
63. Mourtada-Maarabouni, M.; Williams, G.T. RBM5/LUCA-15—tumour suppression by control of apoptosis and the cell cycle? *Sci. World J.* **2002**, *2*, 1885–1890. [[CrossRef](#)]
64. Grifone, R.; Saquet, A.; Desgres, M.; Sangiorgi, C.; Gargano, C.; Li, Z.; Coletti, D.; Shi, D.-L. Rbm24 displays dynamic functions required for myogenic differentiation during muscle regeneration. *Sci. Rep.* **2021**, *11*, 1–15. [[CrossRef](#)]
65. Ando, S.; Tanaka, M.; Chinen, N.; Nakamura, S.; Shimazawa, M.; Hara, H. SMN Protein Contributes to Skeletal Muscle Cell Maturation Via Caspase-3 and Akt Activation. *In Vivo* **2020**, *34*, 3247. [[CrossRef](#)] [[PubMed](#)]
66. Bigot, A.; Klein, A.; Gasnier, E.; Jacquemin, V.; Ravassard, P.; Butler-Browne, G.; Mouly, V.; Furling, D. Large CTG Repeats Trigger p16-Dependent Premature Senescence in Myotonic Dystrophy Type 1 Muscle Precursor Cells. *Am. J. Pathol.* **2009**, *174*, 1435–1442. [[CrossRef](#)]
67. Meinke, P.; Hintze, S.; Limmer, S.; Schoser, B. Myotonic Dystrophy-A Progeroid Disease? *Front. Neurol.* **2018**, *9*, 601. [[CrossRef](#)] [[PubMed](#)]
68. Hintze, S.; Knaier, L.; Limmer, S.; Schoser, B.; Meinke, P. Nuclear Envelope Transmembrane Proteins in Myotonic Dystrophy Type 1. *Front. Physiol.* **2018**, *9*, 1532. [[CrossRef](#)] [[PubMed](#)]
69. Vattemi, G.; Tomelleri, G.; Filosto, M.; Savio, C.; Rizzuto, N.; Tonin, P. Expression of late myogenic differentiation markers in sarcoplasmic masses of patients with myotonic dystrophy. *Neuropathol. Appl. Neurobiol.* **2005**, *31*, 45–52. [[CrossRef](#)]

70. Meola, G.; Cardani, R. Myotonic dystrophies: An update on clinical aspects, genetic, pathology, and molecular pathomechanisms. *Biochim. et Biophys. Acta (BBA)—Mol. Basis Dis.* **2015**, *1852*, 594–606. [[CrossRef](#)] [[PubMed](#)]
71. Konieczny, P.; Stepniak-Konieczna, E.; Sobczak, K. MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Res.* **2014**, *42*, 10873–10887. [[CrossRef](#)]
72. Hildyard, J.C.; Wells, D.J. Identification and Validation of Quantitative PCR Reference Genes Suitable for Normalizing Expression in Normal and Dystrophic Cell Culture Models of Myogenesis. *PLoS Curr.* **2014**, *6*. [[CrossRef](#)]
73. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2012**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
74. Grüning, B.; Diehl, S.; Ramirez, F.; Dündar, F. deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **2014**, *42*, W187–W191. [[CrossRef](#)]
75. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2013**, *30*, 923–930. [[CrossRef](#)]
76. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
77. Zhu, A.; Ibrahim, J.G.; I Love, M. Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics* **2018**, *35*, 2084–2092. [[CrossRef](#)] [[PubMed](#)]

5.2 Publication II

Nuclear envelope transmembrane proteins involved in genome organization are misregulated in myotonic dystrophy type 1 muscle

Vanessa Todorow, Stefan Hintze, Benedikt Schoser, and Peter Meinke (2023)

Frontiers in Cellular and Developmental Biology, Jan 9:10:1007331, doi:

10.3389/fcell.2022.1007331

Contributions

My contributions to the following paper are the selection and analysis of a publicly available RNAseq data set of DM1 patients, including raw data processing like trimming and mapping of reads, quality control, expression, splicing and isoform usage analyses. I further helped in the design and assembly of all figures, as well as in reviewing and editing the final manuscript.



OPEN ACCESS

EDITED BY
Joanna M. Bridger,
Brunel University London, United Kingdom

REVIEWED BY
Ping Hu,
Shanghai Institute of Biochemistry and Cell
Biology (CAS), China
Ian Holt,
Robert Jones and Agnes Hunt
Orthopaedic Hospital NHS Trust,
United Kingdom

*CORRESPONDENCE
Peter Meinke,
✉ Peter.Meinke@med.uni-muenchen.de

SPECIALTY SECTION
This article was submitted to Nuclear
Organization and Dynamics,
a section of the journal
Frontiers in Cell and Developmental
Biology

RECEIVED 30 July 2022
ACCEPTED 27 December 2022
PUBLISHED 09 January 2023

CITATION
Todorow V, Hintze S, Schoser B and
Meinke P (2023), Nuclear envelope
transmembrane proteins involved in
genome organization are misregulated in
myotonic dystrophy type 1 muscle.
Front. Cell Dev. Biol. 10:1007331.
doi: 10.3389/fcell.2022.1007331

COPYRIGHT
© 2023 Todorow, Hintze, Schoser and
Meinke. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Nuclear envelope transmembrane proteins involved in genome organization are misregulated in myotonic dystrophy type 1 muscle

Vanessa Todorow, Stefan Hintze, Benedikt Schoser and
Peter Meinke*

Friedrich-Baur-Institute at the Department of Neurology, University Hospital, LMU, Munich, Germany

Myotonic dystrophy type 1 is a multisystemic disorder with predominant muscle and neurological involvement. Despite a well described pathomechanism, which is primarily a global missplicing due to sequestration of RNA-binding proteins, there are still many unsolved questions. One such question is the disease etiology in the different affected tissues. We observed alterations at the nuclear envelope in primary muscle cell cultures before. This led us to reanalyze a published RNA-sequencing dataset of DM1 and control muscle biopsies regarding the misregulation of NE proteins. We could identify several muscle NE protein encoding genes to be misregulated depending on the severity of the muscle phenotype. Among these misregulated genes were NE transmembrane proteins (NETs) involved in nuclear-cytoskeletal coupling as well as genome organization. For selected genes, we could confirm that observed gene-misregulation led to protein expression changes. Furthermore, we investigated if genes known to be under expression-regulation by genome organization NETs were also misregulated in DM1 biopsies, which revealed that misregulation of two NETs alone is likely responsible for differential expression of about 10% of all genes being differentially expressed in DM1. Notably, the majority of NETs identified here to be misregulated in DM1 muscle are mutated in Emery-Dreifuss muscular dystrophy or clinical similar muscular dystrophies, suggesting a broader similarity on the molecular level for muscular dystrophies than anticipated. This shows not only the importance of muscle NETs in muscle health and disease, but also highlights the importance of the NE in DM1 disease progression.

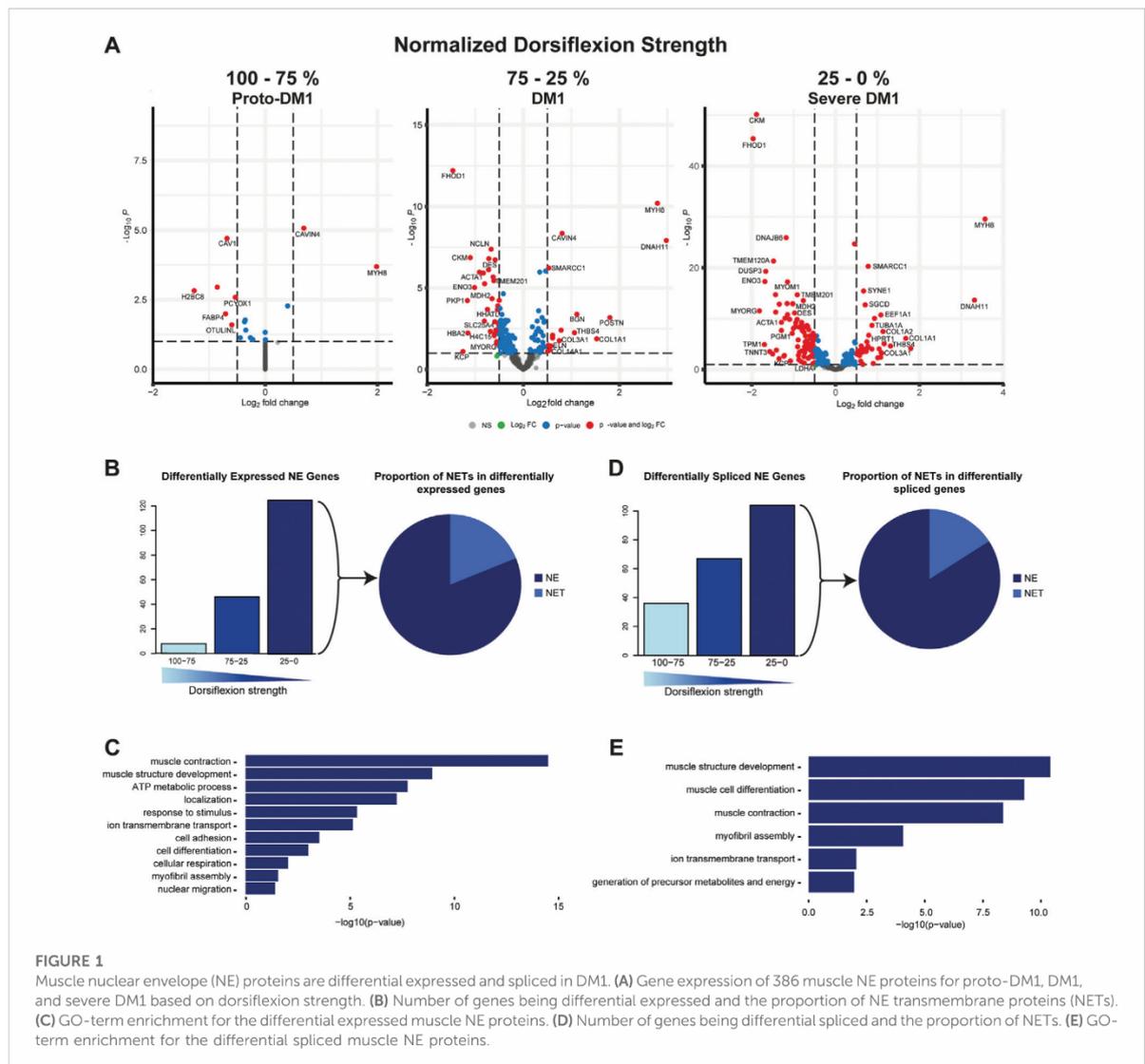
KEYWORDS

nuclear envelope, myotonic dystrophy type 1, genome organization, muscle biopsy, nuclear envelope transmembrane proteins

Introduction

Myotonic dystrophy type 1 (DM1) is clinically characterized by multisystemic involvement with skeletal muscle and brain being the primarily affected organs. Clinical symptoms include myotonia, skeletal muscle weakness and wasting, cardiac arrhythmia, cataracts and insulin resistance, endocrine dysfunction, frontal balding and a shortened lifespan (Udd and Krahe, 2012; Thornton, 2014; Wenninger et al., 2018). An estimated prevalence of about one in 8,000 and the predominant muscle involvement make DM1 one of the most frequent muscular dystrophies in adulthood (Faustino and Cooper, 2003; Wheeler, 2008).

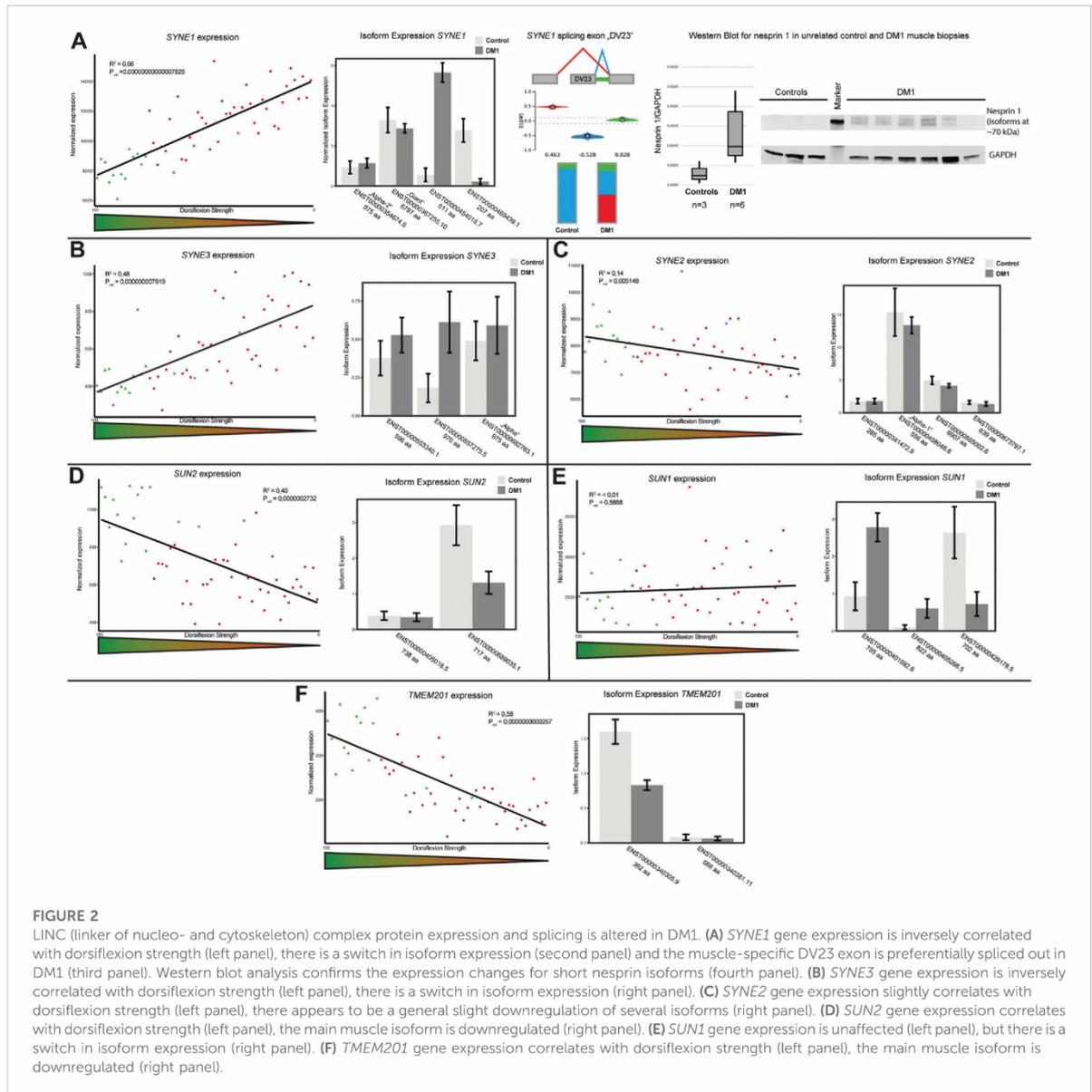
Genetically, DM1 is caused by a pathological CTG-repeat expansion in the 3'UTR of the *DMPK* (dystrophin myotonia protein kinase) gene (Fu et al., 1992). The extended repeat is



unstable, up to 35 CTG-repeats are found in healthy individuals, and between 35 and 49 repeats are considered to be a premutation (Udd and Krahe, 2012). The longer the repeat, the more severe the clinical presentation: between 50 and ~150 repeats usually result in a mild phenotype, a range from ~100 to ~1,000 repeats has been identified in patients with classical DM1, and more than 1000 CTG-triplets usually result in congenital DM, the most severe form of the disease. This rough correlation between repeat length and severity of the disease is non-linear (De Antonio et al., 2016), and there are other factors contributing to the clinical presentation. Maternal inheritance results in more severe symptoms than paternal inheritance, which may be due to an increased greater instability of mutant alleles in female meiosis or maternal-biased CpG methylation of the *DMPK* locus (Rakocevic-Stojanovic et al., 2005; Martorell et al., 2007; Barbé et al., 2017). The extended repeats are somatically unstable, usually resulting in increase of repeat length during the lifetime of an affected individual and

somatic mosaicism (Monckton et al., 1995; Wong et al., 1995). Especially for skeletal muscle it has been shown that repeats can be three- and 25-fold longer than in leukocytes (Thornton et al., 1994; Nakamori et al., 2013).

The mechanisms proposed to contribute to the DM1 phenotype include alternative splicing of several mRNAs (Ho et al., 2005; López-Martínez et al., 2020), altered transcriptional regulation (Ebralidze et al., 2004; Osborne et al., 2009), miRNA misregulation (Rau et al., 2011; Kalsotra et al., 2014; Shen et al., 2020) and inhibited translation (Huichalaf et al., 2010; Meola et al., 2013). The most intensively investigated mechanism is probably alternative splicing, caused by the formation of hairpin structures in the extended CUG-repeat containing *DMPK* RNA transcripts (Napierała and Krzyzosiak, 1997). These secondary structures sequester several RNA-binding proteins, with muscle-blind proteins (MBNL1-3) being the most prominent ones (Fardaei et al., 2001; Fardaei et al., 2002). This



results in a nucleoplasmic depletion of MBLN and therefore loss of function. Another splicing factor, CUGBP elav-like family member 1 (CELF1), gets stabilized in parallel by hyperphosphorylation causing a gain of function (Philips et al., 1998; Kuyumcu-Martinez et al., 2007). In total, this leads to a misbalance of splicing and a shift towards an embryonic splicing pattern. Missplicing of a set of muscle-specific genes including *TTN* (titin), *DMD* (dystrophin) (Yamashita et al., 2012), *CLCN1* (chloride voltage-gated channel 1) (Charlet et al., 2002; Mankodi et al., 2002), and *RYR1* (ryanodine receptor 1) (Kimura et al., 2005), among others, can be directly linked to specific DM1 symptoms.

Despite all this information, it is still unclear which mechanism is contributing to which extent, and if yet unknown factors add to the

development of this complex disease—especially in the different tissues affected. Intriguingly, alterations to the nuclear envelope (NE) structure and expression changes of NE transmembrane proteins (NETs) have been observed in primary DM1 myoblast and myotube cultures (Hintze et al., 2018; Meinke et al., 2018) as well as in patient fibroblasts (Rodríguez et al., 2015; Viegas et al., 2022). NE proteins are linked to a wide range of disorders, including myopathies and neuropathies. Cellular functions of the NE include the organization, regulation, and repair of the genome, signaling, and cellular mechanics (Meinke and Schirmer, 2016). The composition of the NE is at least partially tissue specific (Korfali et al., 2012), and the identification of NE proteins in skeletal muscle (Wilkie et al., 2011) allows to investigate the NE role in DM1.

Here we reanalyzed RNA-sequencing data from deep sequencing of DM1 and control muscle biopsies (Wang et al., 2019) regarding muscle NE proteins to gain some insight in the role of the NE in DM1 and its contribution to the phenotype.

Methods

Sequencing data

The transcriptomes of 44 DM1 and 11 control tibialis biopsies are publicly available in FASTQ format at GEO (GSE86356). Sample processing has been described in (Wang et al., 2019). One DM1 sample was excluded from further analysis due to insufficient quality as assessed with fastqc. Anonymized patient information can be found in the supplementary data of the original publication and includes the evaluation of the normalized dorsiflexion strength in percent with healthy individuals corresponding to 100% of strength. For the subsequent analysis, either all samples or subgroups according to dorsiflexion strength were used. The subgroups are as following: healthy/proto DM1 (dorsiflexion strength 100%–75%), DM1 (dorsiflexion strength 75%–25%), and severe DM1 (dorsiflexion strength 25%–0%).

Bioinformatical analyses

Alignment

Reads were either mapped with STAR v2.7 (Dobin et al., 2013) or Kallisto v0.46.0 (Bray et al., 2016) to the GRCh38 human reference genome. STAR generated BAM files were used for DESeq2 (Love et al., 2014), DEXSeq (Anders et al., 2012) and MAJIQ v2.3 (Vaquero-Garcia et al., 2016), Kallisto counts were used for isoformSwitchAnalyzer (Vitting-Seerup and Sandelin, 2019).

DESeq2

Aligned reads were counted using featureCounts and analyzed with a standard DESeq2 workflow in R v4.2 using the built-in normalization method (median of ratios). Principal component analysis (PCA) was used to plot the samples according to the two main parameters of variability PC1 and PC2, which showed that samples from healthy individuals clustered together, while DM1 patients are scattered along PC1, consistent with disease severity (Supplementary Figure S1). Genes with log₂ foldchanges of > |0.5| and *p*-values < 0.05 have been set to be significantly changed. Gprofiler2 was used for GO analysis. Volcanoplots were generated with EnhancedVolcano, other plots have been generated with ggplot2. For the expression scatter plots of selected nuclear envelope transmembrane proteins in Figure 2 and Figure 3, samples were ordered after the normalized dorsiflexion strength. Additionally, the analysis has been run separately for the above determined three subgroups to find NE associated proteins. All results are in Supplementary Table S1.

DEXSeq

Mapped reads were counted using the in-built python script of DEXSeq with python v3.9. Standard DEXSeq workflow in R v4.2 was followed and exons with less than 40 counts for all

samples filtered out. Exons with a log₂FC of > |0.5| and *p*-value < 0.05 have been set to be significantly changed. Here as well, analysis has been run separately for the above determined three subgroups to find NE associated proteins. All results are in Supplementary Table S2.

Isoformswitchanalyzer

Isoform counts generated by Kallisto were imported in R and abundance values were normalized *via* edgeR. Normalized isoform expressions were used to generate bar charts *via* the in-built isoformSwitchAnalyzer function switchPlotIsoExp (). For this, we focused on the severe DM1 group and compared it to healthy controls. All results are in Supplementary Table S3.

MAJIQ

Alternative splicing events were analyzed using MAJIQ in python v2.7, providing STAR generated BAM files and a GRCh38 gff3 file. The in-built deltapsi script was used to determine significantly altered splice events between DM1 and control with a confidence interval of .9 and percent-spliced-in (psi) values of > |0.1|. MAJIQ Voila was used to visualize the splice graphs. Exon cassette results in Supplementary Table S4.

Western blot

Whole protein extracts were generated from 10 μm muscle sections using RIPA buffer and an ultrasonic sonicator with a MS73 tip (Bandelin Sonopuls) to lyse the sections. The proteins were separated by SDS gel electrophoresis using 4%–15% TGX gels (BioRad #456–8,087) and 10% TGX gels (BioRad #456–8,034). Western blotting was performed using the Trans-Blot[®] Turbo™ system (BioRad). Proteins were transferred to nitrocellulose membranes (Trans-Blot[®] Turbo™ RTA Transfer Kit #170–4,270). Membranes were blocked with 5% skim milk in 1xTBS/0.1% Tween[®] 20. Following primary antibodies were used: nesprin1 (provided by Didier Hodzic (Razafsky et al., 2013)), Tmem38a (Merck Millipore #06–1,005), Plpp7 (Proteintech #20635-1-AP). For quantification mouse antiGAPDH (Milipore #MAB374) was used. As secondary antibodies we used donkey anti-mouse IRDye 680RD and donkey anti-rabbit IRDye 800 CW. All western blot images were obtained using a Licor FC. Quantification was done using the Licor ImageStudio Software. Western blots were repeated at least three times to confirm the results. Full blots are shown in Supplementary Figure S2.

Muscle biopsies

Muscle biopsies were obtained from the Muscle Tissue Culture Collection (MTCC) at the Friedrich-Baur-Institute (Department of Neurology, LMU Klinikum, Ludwig-Maximilians-University, Munich, Germany). All materials were obtained with written informed consent of the donor. Ethical approval for this study was obtained from the ethical review committee at the Ludwig-Maximilians-University, Munich, Germany (reference 45–14).

Results

To analyze muscle NE protein expression and splicing in DM1 biopsies we used a list of 386 proteins identified by mass

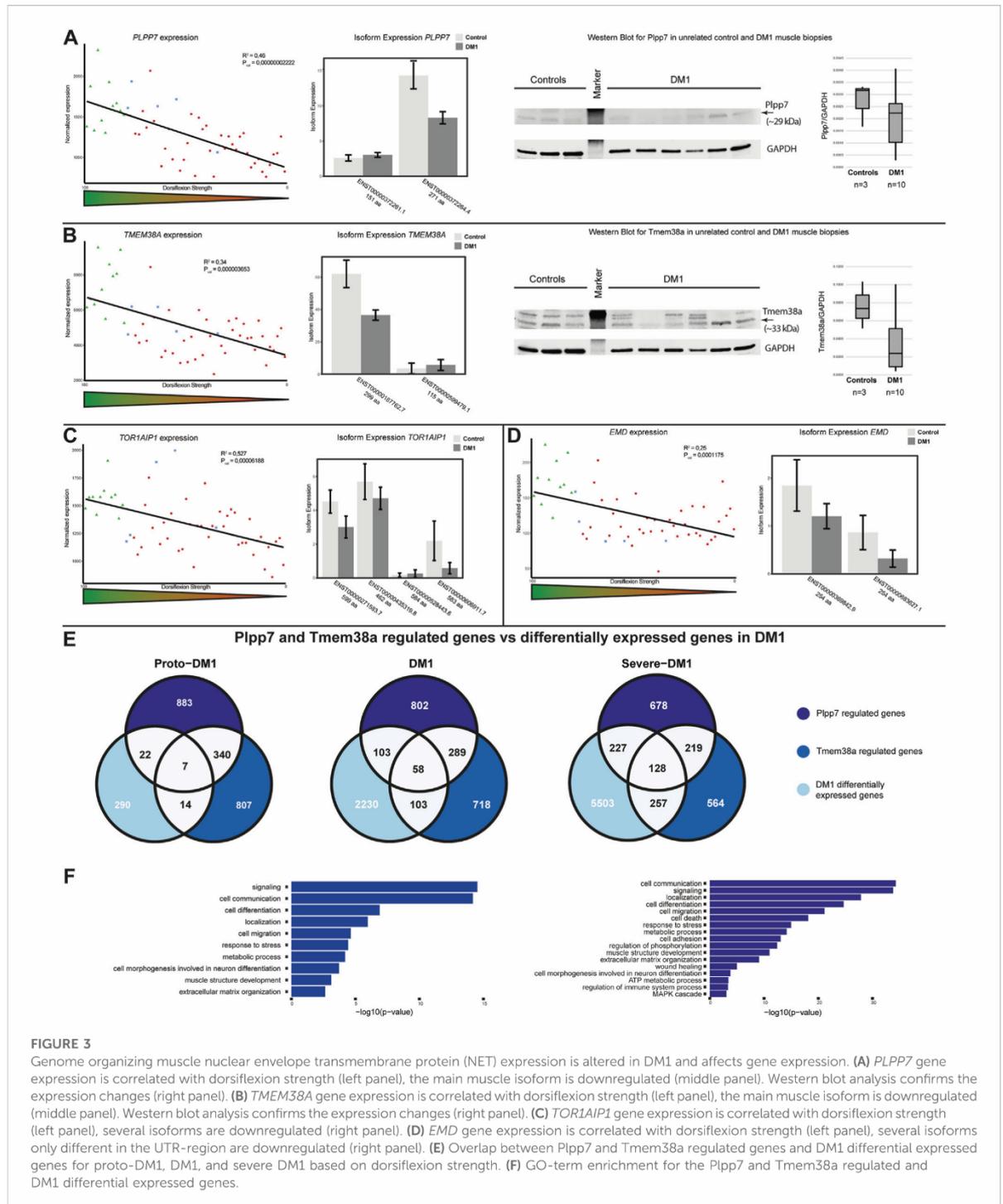


FIGURE 3

Genome organizing muscle nuclear envelope transmembrane protein (NET) expression is altered in DM1 and affects gene expression. (A) *PLPP7* gene expression is correlated with dorsiflexion strength (left panel), the main muscle isoform is downregulated (middle panel). Western blot analysis confirms the expression changes (right panel). (B) *TMEM38A* gene expression is correlated with dorsiflexion strength (left panel), the main muscle isoform is downregulated (middle panel). Western blot analysis confirms the expression changes (right panel). (C) *TOR1AIP1* gene expression is correlated with dorsiflexion strength (left panel), several isoforms are downregulated (right panel). (D) *EMD* gene expression is correlated with dorsiflexion strength (left panel), several isoforms only different in the UTR-region are downregulated (right panel). (E) Overlap between Pipp7 and Tmem38a regulated genes and DM1 differential expressed genes for proto-DM1, DM1, and severe DM1 based on dorsiflexion strength. (F) GO-term enrichment for the Pipp7 and Tmem38a regulated and DM1 differential expressed genes.

spectrometry of isolated muscle NEs (Wilkie et al., 2011; Korfali et al., 2012) (Supplemental Table S5). The genes encoding these 386 proteins were analyzed for alterations in expression or splicing in a published transcriptome dataset of 54 tibialis anterior muscle

biopsies (Wang et al., 2019). We decided to use the datasets of tibialis anterior muscles for our analyses as this muscle is predominantly affected in DM1 (Harper, 2001). These 54 tibialis anterior muscle biopsies originated from 11 unaffected individuals and

43 DM1 patients, all characterized for ankle dorsiflexion strength to quantify how much the muscle was affected. Based on these measurements the DM1 patients were characterized as proto-DM1, DM1 or severe DM1 (Wang et al., 2019).

Differential expression of muscle nuclear envelope proteins

First, read counts were analyzed using DEseq2. We identified two genes up and six genes being downregulated in proto-DM1, while in DM1 14 genes were up and 32 genes downregulated. In severe DM1, there was a further increase of NE-protein encoding genes being differentially expressed, 34 genes were up and 91 genes down (Figure 1A). The total number of genes encoding muscle NE proteins was accordingly increasing with loss of dorsiflexion strength (8, 46, 125; Figure 1B, left panel). Among these differentially expressed genes, the percentage of genes encoding proteins with a transmembrane domain was 11.2% (Figure 1B, right panel). Next, we were interested in which biological functions the protein products of these genes were involved. Pathway analysis revealed functions in muscle relevant processes like muscle contraction, muscle structure development, response to stimulus, and metabolic processes (Figure 1C; Supplementary Table S6).

Splicing alterations of muscle nuclear envelope proteins

Apart from differential expression, splicing alteration can impact the function of translated proteins—especially when considering that the main pathomechanism described in DM1 is an increase in alternative splicing. Similar to differential gene expression the number of genes affected by splice alterations did increase with reduced dorsiflexion strength. In proto-DM1, 36 genes were affected while in DM1 this number increased to 67 and in severe DM1 to 104 genes (Figure 1D, left panel). Among these differentially spliced genes, the percentage of genes encoding proteins with a transmembrane domain was 8.7% (Figure 1D, right panel). Pathway analysis of these alternatively spliced genes also revealed functions in muscle contraction, muscle structure development, and metabolic processes (Figure 1E; Supplementary Table S7).

Cytoskeletal associated NETs

We did describe alterations of the NE in primary DM1 myoblasts and myotubes before (Hintze et al., 2018; Meinke et al., 2018). There, we observed NE invaginations which indicated altered nuclear-cytoskeletal coupling and accordingly identified altered expression of several nesprin isoforms. Based on these data we screened differentially expressed and spliced genes for genes encoding LINC complex (linker of nucleoskeleton and cytoskeleton) and LINC-associated proteins. We identified the expression of *SYNE1*, encoding nesprin 1, to be inverse correlated with dorsiflexion strength (Figure 2A, left panel). As the *SYNE1* gene is giving rise to multiple nesprin isoforms by alternative splicing, we performed an isoform expression analysis. This showed that the expression changes were not caused by

alterations of the giant or muscle specific alpha-2 isoforms. Instead, there was a misregulation of other short isoforms, as illustrated for two isoforms containing neither the KASH nor the actin-binding domain. While a 207 amino acid (aa) isoform was downregulated a 511 aa isoform was strongly upregulated (Figure 2A, second panel). The *SYNE1* gene also came up in the MAJIQ analysis, with a preferential exclusion of a specific exon. This 69 nucleotide exon was identified in an early study (Apel et al., 2000) and later named Δ SR (Simpson and Roberts, 2008) and DV23 (Duong et al., 2014). It is evolutionary conserved and highly muscle-specific (Simpson and Roberts, 2008; Duong et al., 2014). We found this exon to be spliced out in about 50% of the transcripts in DM1 biopsies while it was almost exclusively spliced in in controls (Figure 2A, third panel). To verify the RNAseq data on protein level we performed Western blot on a set of unrelated control and DM1 muscle biopsies. An increased signal of several bands between 70 and 260 kDa in DM1 patients muscle indicates an upregulation of short nesprin 1 isoforms on protein level (Figure 2A, fourth panel; Supplementary Figure S2).

Similar to *SYNE1* the expression of *SYNE3*, encoding nesprin 3, was also inverse correlated with dorsiflexion strength (Figure 2B, left panel). Here the increased expression appears to originate from an upregulation of a 970 aa isoform, which differs from the “alpha isoform” (975 aa) by the loss of the amino acids 793 to 797 due to the usage of an alternative splice site (Figure 2B, right panel). Considering the differential expression of the nesprins 1 and 3 we decided to look also at *SYNE2*, but here we found only a very mild trend for a correlation of gene expression and dorsiflexion strength which may be caused by changes to the expression of the muscle isoform “alpha-1” (Figure 2C).

The nuclear side of the LINC complex consists of SUN proteins. The expression of *SUN2* was strongly correlated with dorsiflexion strength (Figure 2D, left panel). In terms of isoform expression, this seems to originate from a downregulation of the 717 aa isoform (Figure 2D, right panel). We also looked at expression of *SUN1*, but could not find clear correlation with dorsiflexion strength (Figure 2E, left panel). However, looking at the isoform expression we could see several alterations which seem to level out the total gene expression. While a 785 aa and a 822 aa isoform were upregulated, a 702 aa isoform was strongly downregulated (Figure 2E, right panel).

Samp1, which is encoded by the *TMEM201* gene, is functional associated to the LINC complex (Gudise et al., 2011). We found expression of Samp1 to strongly correlate with dorsiflexion strength (Figure 2F, left panel). The reduced expression is due to downregulation of the shorter isoform (392 aa), with the longer isoform (666 aa) being affected very little (Figure 2F, right panel).

Genome organizing mNETs

Samp1 has not only been described to be involved in the nucleoskeleton-cytoskeletal coupling via the LINC complex, but has also been shown to be involved in genome organization (Zuleger et al., 2013). This in addition to observed general gene expression changes here as well as in DM1 tissue culture systems (Todorow et al., 2021) prompted us to investigate muscle specific NETs involved in genome organization in more detail. We found in addition to *TMEM201* the expression of *PLPP7*, *TMEM38A*, *TOR1AIP1* and *EMD* to be altered.

For *PLPP7* we found a positive correlation of gene expression and dorsiflexion strength (Figure 3A, left panel). This was caused by

downregulation of the main isoform of the protein (271 aa) (Figure 3A, middle panel). We proceeded to confirm these expression changes on protein level by Western blot, which showed downregulation of Ppp7 in unrelated DM1 muscle biopsies (Figure 3A, right panel). Tmem38a expression was correlating in a similar manner as *PLPP7* with dorsiflexion strength (Figure 3B, left panel). Here the expression changes also seemingly originated from the main isoform (299 aa) (Figure 3B, middle panel). We could also confirm these results on protein level in unrelated samples (Figure 3B, right panel). We looked at two additional NETs known to be involved in genome organization, LAP1 (encoded by *TORIAIP1*), and emerin (encoded by *EMD*). For both we found a clear correlation with dorsiflexion strength originating from a downregulation of all isoforms (Figures 3C, D).

Apart from the effect on expression and splicing of muscle NE proteins we were also interested in possible functional consequences. Ppp7 and Tmem38a are muscle specific NETs involved in genome organization, and the genes they contribute to regulate in C2C12 myotubes (which partially overlap) have been identified (Robson et al., 2016). To investigate whether the observed expression changes in DM1 muscle biopsies do have any functional relevance we proceeded to test the expression of these Ppp7 and Tmem38a co-regulated genes in the three subgroups. We could indeed find an overlap between genes regulated by both proteins in mouse myotubes and DM1 patient biopsies: in proto-DM1, there was an overlap of 43 genes, in DM1 264 genes, and in severe DM1 612 genes (Figure 3E). This made up 13, 11, and 10% of the overall differentially expressed genes in the DM1 samples, respectively. Next, we were interested in the biological functions of the genes under Ppp7 or Tmem38a control. Considering the number of genes, this analysis was possible for the DM1 and severe DM1 groups. The main enriched pathways were signaling, cell communication, cell migration, localization, response to stress and metabolic process (Figure 3F, Supplementary Table S8).

Discussion

The missplicing in DM1 is well investigated and there are many target genes of this missplicing described, which are contributing or likely contributing to the disease pathology. Yet, it still remains elusive which additional pathomechanisms are contributing to the development in DM1, and to which extent, especially in the different tissues affected. The NE has been shown to be much more than just a barrier separating the genome from the rest of the cell (de Las Heras et al., 2013), it hosts a tissue specific proteome and tissue specific as well as ubiquitously expressed NETs have been shown to be involved in controlling the intranuclear positioning and thus expression of genes, often in a tissue specific manner (Zuleger et al., 2011). We could previously identify NE alterations in muscle tissue culture systems of DM1, with likely effects on cell cycle control and differentiation (Hintze et al., 2018; Meinke et al., 2018). Investigating the involvement of the NE in DM1 mature muscle was therefore the logical follow up to unravel its role in the DM1 pathology.

The set of NE genes we investigated contained genes with and without transmembrane domains, as we did not want to exclude a possible contribution of NE-associated proteins. We could indeed find for both NE and NET encoding genes a high percentage of

differential expression and differential splicing. This highlighted the likelihood of an important role of NE proteins in DM1 as the GO-term analysis revealed that the most enriched processes of these differentially regulated genes are all relevant for muscle function.

We wanted to follow up on specific aspects of NE function. Considering the misregulation of nesprin proteins in DM1 muscle cell cultures (Hintze et al., 2018), which is a possible explanation for observed NE invaginations (Meinke et al., 2018), and the identification of mutations in *SYNE* and *SUN* in a clinically similar disease, EDMD (Zhang et al., 2007; Meinke et al., 2014), we looked at all components of the LINC complex. We could identify isoform-specific alterations in the expression of the *SYNE1*, *SYNE3*, *SUN1*, and *SUN2* genes—all core components of LINC complexes. Although there was no apparent change in the expression of the muscle-specific nesprin 1 isoform “alpha-2”, in about half of these transcripts a 23 aa exon (DV23) was spliced out in DM1 patients. As this exon has been shown to be included in 94% muscle *SYNE1* transcripts (Duong et al., 2014) this could indicate a loss of a muscle-specific nesprin 1 function. Furthermore, for Samp1, which has been identified as a LINC complex associated protein, there was also a downregulation of the major muscle isoform. This clearly indicates a likely weakening of the nuclear-cytoskeletal connection in DM1 muscle, which is going to impact on mechanotransduction as well as nuclei positioning. This is in line with observed missplicing of the myc box-dependent-interacting protein 1 (Bin1), which is involved in the formation of tubular invaginations of the plasma membrane that function in depolarization-contraction coupling (Fugier et al., 2011).

Another important aspect of the NE is the tissue specific regulation of gene expression by the recruitment of specific genes to the NE by tissue specific NETs. Examples for this are Ppp7 and Tmem38a, which have been shown to have important muscle functions (Robson et al., 2016). It is important to note that these proteins appear to have an additive effect, a knockdown of more than one resulted in stronger effects than single knockdowns (Robson et al., 2016). Thus, it is likely that a reduced expression of several NETs is also adding up to result in phenotypical consequences. We found both proteins to be downregulated on RNA and protein level, and by comparing DM1 differentially expressed genes to the genes identified under their expression control in mouse myoblasts, we could prove that we have similar effects in DM1. Intriguingly, mutations in *PLPP7*, *TMEM38A*, and *TMEM201* have been identified in muscular dystrophy patients with an EDMD-like phenotype (Meinke et al., 2020), which highlights the importance of these proteins in muscle disease. Misregulation of the two muscle NETs Ppp7 and Tmem38a alone does indeed account for about 10% of all differentially expressed genes in DM1 muscle. Since there are additional NETs misregulated in DM1 the actual effect is probably even more profound. It has been shown that Samp1 can also reposition chromosomes (Zuleger et al., 2013) and emerin directly binds histone deacetylase 3 (Demmerle et al., 2012), while LAP1 binds indirectly to chromatin (Foisner and Gerace, 1993). Notably, mutations in the genes encoding emerin and LAP1 also cause EDMD (Bione et al., 1994) respectively a very similar muscular dystrophy (Kayman-Kurekci et al., 2014). All in all, our data suggests that DM1 and EDMD share a broader common ground also on the cellular level rather than only in the symptomology.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Ethics statement

The studies involving human participants were reviewed and approved by Ethical review committee at the Ludwig-Maximilians-University, Munich, Germany. The patients/participants provided their written informed consent to participate in this study.

Author contributions

VT, SH, BS, and PM contributed to the conception and design of the experiments. PM and BS wrote the manuscript, SH performed the experiments, VT performed the bioinformatical analyses.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 470092532.

References

- Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Res.* 22 (10), 2008–2017. doi:10.1101/gr.133744.111
- Apel, E. D., Lewis, R. M., Grady, R. M., and Sanes, J. R. (2000). Syne-1, a dystrophin- and Klarsicht-related protein associated with synaptic nuclei at the neuromuscular junction. *J. Biol. Chem.* 275 (41), 31986–31995. doi:10.1074/jbc.M004775200
- Barbé, L., Lanni, S., López-Castel, A., Franck, S., Spits, C., Keymolen, K., et al. (2017). CpG methylation, a parent-of-origin effect for maternal-biased transmission of congenital myotonic dystrophy. *Am. J. Hum. Genet.* 100 (3), 488–505. doi:10.1016/j.ajhg.2017.01.033
- Bione, S., Maestrini, E., Rivella, S., Mancini, M., Regis, S., Romeo, G., et al. (1994). Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat. Genet.* 8 (4), 323–327. doi:10.1038/ng1294-323
- Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Erratum: Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34 (8), 888. doi:10.1038/nbt0816-888d
- Charlet, B. N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A., and Cooper, T. A. (2002). Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell* 10 (1), 45–53. doi:10.1016/s1097-2765(02)00572-5
- De Antonio, M., Dogan, C., Hamroun, D., Mati, M., Zerrouki, S., Eymard, B., et al. (2016). Unravelling the myotonic dystrophy type 1 clinical spectrum: A systematic registry-based study with implications for disease classification. *Rev. Neurol. Paris.* 172 (10), 572–580. doi:10.1016/j.neurol.2016.08.003
- de Las Heras, J. I., Meinke, P., Batrakou, D. G., Srsen, V., Zuleger, N., Kerr, A. R., et al. (2013). Tissue specificity in the nuclear envelope supports its functional complexity. *Nucleus* 4 (6), 460–477. doi:10.4161/nucl.26872
- Demmerle, J., Koch, A. J., and Holaska, J. M. (2012). The nuclear envelope protein emerin binds directly to histone deacetylase 3 (HDAC3) and activates HDAC3 activity. *J. Biol. Chem.* 287 (26), 22080–22088. doi:10.1074/jbc.M111.325308
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). Star: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635
- Duong, N. T., Morris, G. E., Lam, T., Zhang, Q., Sewry, C. A., Shanahan, C. M., et al. (2014). Nesprins: Tissue-specific expression of epsilon and other short isoforms. *PLoS One* 9 (4), e94380. doi:10.1371/journal.pone.0094380
- Ebralidze, A., Wang, Y., Petkova, V., Ebralidze, K., and Junghans, R. P. (2004). RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science* 303 (5656), 383–387. doi:10.1126/science.1088679
- Fardaei, M., Larkin, K., Brook, J. D., and Hamshere, M. G. (2001). *In vivo* co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res.* 29 (13), 2766–2771. doi:10.1093/nar/29.13.2766
- Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., et al. (2002). Three proteins, MBNL, MBLL and MBXL, co-localize *in vivo* with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* 11 (7), 805–814. doi:10.1093/hmg/11.7.805
- Faustino, N. A., and Cooper, T. A. (2003). Pre-mRNA splicing and human disease. *Genes Dev.* 17 (4), 419–437. doi:10.1101/gad.1048803
- Foisner, R., and Gerace, L. (1993). Integral membrane proteins of the nuclear envelope interact with lamins and chromosomes, and binding is modulated by mitotic phosphorylation. *Cell* 73 (7), 1267–1279. doi:10.1016/0092-8674(93)90355-t
- Fu, Y. H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., et al. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 255 (5049), 1256–1258. doi:10.1126/science.1546326
- Fugier, C., Klein, A. F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A., et al. (2011). Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nat. Med.* 17 (6), 720–725. doi:10.1038/nm.2374
- Gudise, S., Figueroa, R. A., Lindberg, R., Larsson, V., and Hallberg, E. (2011). Samp1 is functionally associated with the LINC complex and A-type lamina networks. *J. Cell Sci.* 124, 2077–2085. doi:10.1242/jcs.078923
- Harper, P. S. (2001). in *Myotonic dystrophy—the clinical picture Myotonic dystrophy*. Editor W. B Saunders. third edition ed London, 17–46.
- Hintze, S., Knaier, L., Limmer, S., Schoser, B., and Meinke, P. (2018). Nuclear envelope transmembrane proteins in myotonic dystrophy type 1. *Front. Physiol.* 9, 1532. doi:10.3389/fphys.2018.01532
- Ho, T. H., Bundman, D., Armstrong, D. L., and Cooper, T. A. (2005). Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Hum. Mol. Genet.* 14 (11), 1539–1547. doi:10.1093/hmg/ddi162
- Huichalaf, C., Sakai, K., Jin, B., Jones, K., Wang, G. L., Schoser, B., et al. (2010). Expansion of CUG RNA repeats causes stress and inhibition of translation in myotonic dystrophy 1 (DM1) cells. *FASEB J.* 24 (10), 3706–3719. doi:10.1096/fj.09-151159
- Kalsotra, A., Singh, R. K., Gurha, P., Ward, A. J., Creighton, C. J., and Cooper, T. A. (2014). The MeZ transcription network is disrupted in myotonic dystrophy heart tissue, dramatically altering miRNA and mRNA expression. *Cell Rep.* 6 (2), 336–345. doi:10.1016/j.celrep.2013.12.025

Acknowledgments

We thank Prof. Eric C. Schirmer and Didier Hodzic for the antibodies provided to us.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2022.1007331/full#supplementary-material>

- Kayman-Kurekci, G., Talim, B., Korkusuz, P., Sayar, N., Sarioglu, T., Oncel, I., et al. (2014). Mutation in TOR1AIP1 encoding LAPIB in a form of muscular dystrophy: A novel gene related to nuclear envelopathies. *Neuromuscul. Disord.* 24 (7), 624–633. doi:10.1016/j.nmd.2014.04.007
- Kimura, T., Nakamori, M., Lueck, J. D., Pouliquin, P., Aoike, F., Fujimura, H., et al. (2005). Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca²⁺-ATPase in myotonic dystrophy type 1. *Hum. Mol. Genet.* 14 (15), 2189–2200. doi:10.1093/hmg/ddi223
- Korfali, N., Wilkie, G. S., Swanson, S. K., Srsen, V., de Las Heras, J., Batrakou, D. G., et al. (2012). The nuclear envelope proteome differs notably between tissues. *Nucleus* 3 (6), 552–564. doi:10.4161/nucl.22257
- Kuyumcu-Martinez, N. M., Wang, G. S., and Cooper, T. A. (2007). Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol. Cell* 28 (1), 68–78. doi:10.1016/j.molcel.2007.07.027
- López-Martínez, A., Soblechero-Martín, P., de-la-Puente-Ovejero, L., Nogales-Gadea, G., and Arechavala-Gomez, V. (2020). An overview of alternative splicing defects implicated in myotonic dystrophy type 1. *GenesGenes (Basel)* 11 (9), 1109. doi:10.3390/genes11091109
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Mankodi, A., Takahashi, M. P., Jiang, H., Beck, C. L., Bowers, W. J., Moxley, R. T., et al. (2002). Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol. Cell* 10 (1), 35–44. doi:10.1016/s1097-2765(02)00563-4
- Martorell, L., Cobo, A. M., Baiget, M., Naudo, M., Poza, J. J., and Parra, J. (2007). Prenatal diagnosis in myotonic dystrophy type 1. Thirteen years of experience: Implications for reproductive counselling in DMI families. *Prenat. Diagn.* 27 (1), 68–72. doi:10.1002/pd.1627
- Meinke, P., Hintze, S., Limmer, S., and Schoser, B. (2018). Myotonic dystrophy-A progeroid disease? *Front. Neurol.* 9, 601. doi:10.3389/fneur.2018.00601
- Meinke, P., Kerr, A. R. W., Czapiewski, R., de Las Heras, J. I., Dixon, C. R., Harris, E., et al. (2020). A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism. *EBioMedicine* 51, 102587. doi:10.1016/j.ebiom.2019.11.048
- Meinke, P., Mattioli, E., Haque, F., Antoku, S., Columbaro, M., Straatman, K. R., et al. (2014). Muscular dystrophy-associated SUN1 and SUN2 variants disrupt nuclear-cytoskeletal connections and myonuclear organization. *PLoS Genet.* 10 (9), e1004605. doi:10.1371/journal.pgen.1004605
- Meinke, P., and Schirmer, E. C. (2016). The increasing relevance of nuclear envelope myopathies. *Curr. Opin. Neurol.* 29 (5), 651–661. doi:10.1097/wco.0000000000000359
- Meola, G., Jones, K., Wei, C., and Timchenko, L. T. (2013). Dysfunction of protein homeostasis in myotonic dystrophies. *Histol. Histopathol.* 28 (9), 1089–1098. doi:10.14670/HH-28.1089
- Monckton, D. G., Wong, L. J., Ashizawa, T., and Caskey, C. T. (1995). Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: Small pool PCR analyses. *Hum. Mol. Genet.* 4 (1), 1–8. doi:10.1093/hmg/4.1.1
- Nakamori, M., Sobczak, K., Puwanant, A., Welle, S., Eichinger, K., Pandya, S., et al. (2013). Splicing biomarkers of disease severity in myotonic dystrophy. *Ann. Neurol.* 74 (6), 862–872. doi:10.1002/ana.23992
- Napierala, M., and Krzyzosiak, W. J. (1997). CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J. Biol. Chem.* 272 (49), 31079–31085. doi:10.1074/jbc.272.49.31079
- Osborne, R. J., Lin, X., Welle, S., Sobczak, K., O'Rourke, J. R., Swanson, M. S., et al. (2009). Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy. *Hum. Mol. Genet.* 18 (8), 1471–1481. doi:10.1093/hmg/ddp058
- Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998). Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* 280 (5364), 737–741. doi:10.1126/science.280.5364.737
- Rakocevic-Stojanovic, V., Savic, D., Pavlovic, S., Iavrnica, D., Stevic, Z., Basta, L., et al. (2005). Intergenerational changes of CTG repeat depending on the sex of the transmitting parent in myotonic dystrophy type 1. *Eur. J. Neurol.* 12 (3), 236–237. doi:10.1111/j.1468-1331.2004.01075.x
- Rau, F., Freyermuth, F., Fugier, C., Villemin, J. P., Fischer, M. C., Jost, B., et al. (2011). Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat. Struct. Mol. Biol.* 18 (7), 840–845. doi:10.1038/nsmb.2067
- Razafsky, D. S., Ward, C. L., Kolb, T., and Hodzic, D. (2013). Developmental regulation of linkers of the nucleoskeleton to the cytoskeleton during mouse postnatal retinogenesis. *Nucleus* 4 (5), 399–409. doi:10.4161/nucl.26244
- Robson, M. I., de Las Heras, J. I., Czapiewski, R., P. L. T., Booth, D. G., Kelly, D. A., et al. (2016). Tissue-specific gene repositioning by muscle nuclear membrane proteins enhances repression of critical developmental genes during myogenesis. *Mol. Cell* 62 (6), 834–847. doi:10.1016/j.molcel.2016.04.035
- Rodríguez, R., Hernández-Hernández, O., Magaña, J. J., González-Ramírez, R., García-López, E. S., and Cisneros, B. (2015). Altered nuclear structure in myotonic dystrophy type 1-derived fibroblasts. *Mol. Biol. Rep.* 42 (2), 479–488. doi:10.1007/s11033-014-3791-4
- Shen, X., Xu, F., Li, M., Wu, S., Zhang, J., Wang, A., et al. (2020). miR-322/-503 rescues myoblast defects in myotonic dystrophy type 1 cell model by targeting CUG repeats. *Cell Death Dis.* 11 (10), 891. doi:10.1038/s41419-020-03112-6
- Simpson, J. G., and Roberts, R. G. (2008). Patterns of evolutionary conservation in the nesprin genes highlight probable functionally important protein domains and isoforms. *Biochem. Soc. Trans.* 36, 1359–1367. doi:10.1042/bst0361359
- Thornton, C. A., Johnson, K., and Moxley, R. T., 3rd (1994). Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann. Neurol.* 35 (1), 104–107. doi:10.1002/ana.410350116
- Thornton, C. A. (2014). Myotonic dystrophy. *Neurol. Clin.* 32 (3), 705–719. viii. doi:10.1016/j.ncl.2014.04.011
- Todorow, V., Hintze, S., Kerr, A. R. W., Hehr, A., Schoser, B., and Meinke, P. (2021). Transcriptome analysis in a primary human muscle cell differentiation model for myotonic dystrophy type 1. *Int. J. Mol. Sci.* 22 (16), 8607. doi:10.3390/ijms22168607
- Udd, B., and Krahe, R. (2012). The myotonic dystrophies: Molecular, clinical, and therapeutic challenges. *Lancet Neurol.* 11 (10), 891–905. doi:10.1016/S1474-4422(12)70204-1
- Vaquero-García, J., Barrera, A., Gazzara, M. R., González-Vallinas, J., Lahens, N. F., Hogenesch, J. B., et al. (2016). A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* 5, e11752. doi:10.7554/eLife.11752
- Viegas, D., Pereira, C. D., Martins, F., Mateus, T., da Cruz, E. S. O. A. B., Herdeiro, M. T., et al. (2022). Nuclear envelope alterations in myotonic dystrophy type 1 patient-derived fibroblasts. *Int. J. Mol. Sci.* 23 (1), 522. doi:10.3390/ijms23010522
- Vitting-Seerup, K., and Sandelin, A. (2019). IsoformSwitchAnalyzeR: Analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics* 35 (21), 4469–4471. doi:10.1093/bioinformatics/btz247
- Wang, E. T., Treacy, D., Eichinger, K., Struck, A., Estabrook, J., Olafson, H., et al. (2019). Transcriptome alterations in myotonic dystrophy skeletal muscle and heart. *Hum. Mol. Genet.* 28 (8), 1312–1321. doi:10.1093/hmg/ddy432
- Weninger, S., Montagnese, F., and Schoser, B. (2018). Core clinical phenotypes in myotonic dystrophies. *Front. Neurol.* 9, 303. doi:10.3389/fneur.2018.00303
- Wheeler, T. M. (2008). Myotonic dystrophy: Therapeutic strategies for the future. *Neurotherapeutics* 5 (4), 592–600. doi:10.1016/j.nurt.2008.08.001
- Wilkie, G. S., Korfali, N., Swanson, S. K., Malik, P., Srsen, V., Batrakou, D. G., et al. (2011). Several novel nuclear envelope transmembrane proteins identified in skeletal muscle have cytoskeletal associations. *Mol. Cell Proteomics* 10 (1), M110.003129. doi:10.1074/mcp.M110.003129
- Wong, L. J., Ashizawa, T., Monckton, D. G., Caskey, C. T., and Richards, C. S. (1995). Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am. J. Hum. Genet.* 56 (1), 114–122.
- Yamashita, Y., Matsuura, T., Shinmi, J., Amakusa, Y., Masuda, A., Ito, M., et al. (2012). Four parameters increase the sensitivity and specificity of the exon array analysis and disclose 25 novel aberrantly spliced exons in myotonic dystrophy. *J. Hum. Genet.* 57 (6), 368–374. doi:10.1038/jhg.2012.37
- Zhang, Q., Bethmann, C., Worth, N. F., Davies, J. D., Wasner, C., Feuer, A., et al. (2007). Nesprin-1 and -2 are involved in the pathogenesis of Emery Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. *Hum. Mol. Genet.* 16 (23), 2816–2833. doi:10.1093/hmg/ddm238
- Zuleger, N., Boyle, S., Kelly, D. A., de las Heras, J. I., Lazou, V., Korfali, N., et al. (2013). Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. *Genome Biol.* 14 (2), R14. doi:10.1186/gb-2013-14-2-r14
- Zuleger, N., Robson, M. I., and Schirmer, E. C. (2011). The nuclear envelope as a chromatin organizer. *Nucleus* 2 (5), 339–349. doi:10.4161/nucl.2.5.17846

5.3 Publication III

Metabolic, fibrotic, and splicing pathways are all altered in Emery-Dreifuss muscular dystrophy Spectrum patients to differing degrees

Jose de Las Heras, **Vanessa Todorow**, Leijla Krečinić-Balić, Stefan Hintze, Rafal Czapiewski, Shaun Webb, Benedikt Schoser, Peter Meinke, and Eric C.

Schirmer (2022)

Human Molecular Genetics, Mar 6;32(6):1010-1031, doi: 10.1093/hmg/ddac264

Contributions

In the following publication, I conducted a splicing analysis of an RNAseq data set generated in a collaborating lab and assembled my findings in Figure 5. I also described and interpreted the results of the splicing analysis in the section “Splicing pathways uniformly altered in EDMD patients yield loss of muscle-specific splice variants”.

Metabolic, fibrotic and splicing pathways are all altered in Emery-Dreifuss muscular dystrophy spectrum patients to differing degrees

Jose I. de las Heras¹, Vanessa Todorow², Lejla Krečinić-Balić², Stefan Hintze², Rafal Czapiewski¹, Shaun Webb³, Benedikt Schoser², Peter Meinke² and Eric C. Schirmer^{1,*}

¹Institute of Cell Biology, University of Edinburgh, Edinburgh, UK

²Friedrich-Baur-Institute, Department of Neurology, LMU Clinic, Ludwig-Maximilians-University, Munich, Germany

³Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, UK

*To whom correspondence should be addressed at: Institute of Cell Biology; University of Edinburgh, Kings Buildings, Michael Swann Building, Room 5.22 Max Born Crescent, Edinburgh EH9 3BF, UK. Tel: +44 1316507075; Email: e.schirmer@ed.ac.uk

Abstract

Emery-Dreifuss muscular dystrophy (EDMD) is a genetically and clinically variable disorder. Previous attempts to use gene expression changes to find its pathomechanism were unavailing, so we engaged a functional pathway analysis. RNA-Seq was performed on cells from 10 patients diagnosed with an EDMD spectrum disease with different mutations in seven genes. Upon comparing to controls, the pathway analysis revealed that multiple genes involved in fibrosis, metabolism, myogenic signaling and splicing were affected in all patients. Splice variant analysis revealed alterations of muscle-specific variants for several important muscle genes. Deeper analysis of metabolic pathways revealed a reduction in glycolytic and oxidative metabolism and reduced numbers of mitochondria across a larger set of 14 EDMD spectrum patients and 7 controls. Intriguingly, the gene expression signatures segregated the patients into three subgroups whose distinctions could potentially relate to differences in clinical presentation. Finally, differential expression analysis of miRNAs changing in the patients similarly highlighted fibrosis, metabolism and myogenic signaling pathways. This pathway approach revealed a transcriptome profile that can both be used as a template for establishing a biomarker panel for EDMD and direct further investigation into its pathomechanism. Furthermore, the segregation of specific gene changes into distinct groups that appear to correlate with clinical presentation may template development of prognostic biomarkers, though this will first require their testing in a wider set of patients with more clinical information.

Introduction

Emery-Dreifuss muscular dystrophy (EDMD) is a genetically heterogeneous neuromuscular orphan spectrum disease affecting ~0.3–0.4 in 100 000 people (1), with clinical variability presenting even in family members carrying the same mutation (2–5). EDMD patients present typically in mid to late childhood with early contractures of elbows and Achilles' tendons and progressive wasting of the lower leg and upper arm muscles. Cardiac involvement is also highly characteristic but tends to appear later in development and quite variably in time, though it tends to be reasonably uniform in the form it takes of cardiac conduction defects and dilated cardiomyopathy (6). Other features vary considerably in clinical presentation, leading to the usage of 'Emery-Dreifuss-like syndromes' (7,8): patients from the same pedigree can show remarkable phenotypic variation (2). The genetic variability is underscored by several confirmed linked genes and several additional candidate genes, although there are still some cases where no confirmed or candidate disease allele has been identified (9–11). The lack of large pedigrees in combination with its genetic heterogeneity, clinical variability, already some known modifier genes and limited patient numbers makes solving its pathomechanism difficult.

The original genes linked to EDMD, EMD encoding emerin and LMNA encoding lamin A/C, have both cytoskeletal and gene regulation roles leading to strong arguments for either function being responsible for the EDMD pathomechanism (12,13). The subsequent linking of nesprin and Sun proteins to EDMD (14,15) failed to lend clarity since they function in mechanosignal transduction (16). However, several recently linked genes have clear roles in genome organization and regulation (10), suggesting that this is the pathomechanism. These genes encode proteins that, such as emerin, are nuclear envelope transmembrane proteins (NETs) and seem to function by fine-tuning muscle gene expression by promoting the release of pro-myogenic genes from the nuclear periphery to enhance their activation while concomitantly recruiting metabolism genes (many from the alternative differentiation pathway of adipogenesis) to the nuclear envelope to better repress them (17–19). EDMD mutations were found in five muscle-specific NETs with this genome organization function, PLPP7 (also known as NET39), WFS1, TMEM38A, TMEM201 and TMEM214, and each tested had some specificity in the sets of genes that they target, though there was also some overlap (17). These studies together with the wide range of lamin and emerin gene regulatory activities led us to the non-traditional hypothesis

Received: July 1, 2022. Revised: September 16, 2022. Accepted: October 20, 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

for the EDMD pathomechanism whereby moderate reductions in many genes could yield the same phenotype as shutting down a central gene of a particular pathway. Accordingly, we considered that searching for uniformity among patients in altered pathways might be more revealing than searching for uniformity in expression changes of particular genes.

The only previous study, to our knowledge, using gene expression changes to identify critical misregulated genes underlying EDMD pathophysiology, focused on the identification of genes altered specifically in EDMD compared with a set of 10 other muscular dystrophies (20). This study only considered eight total LMNA- or EMD-linked cases of EDMD, but EDMD now has many more genes and modifiers linked to it and, moreover, there is a wider clinical spectrum of EDMD-like phenotypes (11). Their analysis indicated potential abnormalities in the regulation of cell cycle and myogenic differentiation, associated with perturbations in the pRB/MYOD/LMNA hub, which were consistent with changes in an *Emd*^{-/-} mouse model (21). Roughly a fifth each of EDMD mutations occurs in LMNA and EMD while another 5–6% are collectively caused by four other widely expressed nuclear envelope proteins nesprin 1 (encoded by *SYNE1*), nesprin 2 (encoded by *SYNE2*), Sun1 (encoded by *SUN1*) and FHL1 (encoded by *FHL1*) (14,15,22–24). Another approximately 20% of EDMD mutations were accounted for by muscle-specific NETs that regulate muscle-specific genome organization (10). These include NET39 (encoded by *PLPP7*), TMEM38A (encoded by *TMEM38A*), WFS1 (encoded by *WFS1*), NET5 (encoded by *TMEM201*) and TMEM214 (encoded by *TMEM214*) that affect 3D gene positioning with corresponding effects on expression (17,19). Accordingly, we sought to search for commonly affected pathways from a much wider range of EDMD-linked genes including LMNA, EMD, FHL1, SUN1, SYNE1, PLPP7 and TMEM214 alleles on the expectation that, covering an even wider genetic and clinical spectrum, the most important pathways for EDMD pathophysiology would be highlighted.

Results

RNA-Seq analysis of EDMD patient cells

We performed RNA-Seq on myotubes differentiated *in vitro* from myoblasts isolated from 10 unrelated clinically diagnosed EDMD patients with distinct mutations in seven different genes to sample the genetic diversity of EDMD (Fig. 1). Patient mutations were TMEM214 p.R179H, PLPP7/NET39 p.M92K, Sun1 p.G68D/G388S, Nesprin 1 p.S6869*, Emerin p.S58Sfs*1, FHL1 mutations c.688 + 1G > A, p.C224W and p.V280M, Lamin A/C mutation p.T528K and Lamin A mutation p.R571S (this last mutation occurs in an exon absent from the lamin C splice variant). These mutations covered a wide range of clinical phenotypes with the age of onset ranging from early childhood to adult life and associated pathology ranging from no reported contractures to rigid spine (Table 1). Myoblast isolation followed by *in vitro* differentiation was chosen over isolating mRNA directly from the tissue samples in order to try to capture the earliest changes in gene expression due to the disease mutations *i.e.* the most likely to initiate the pathomechanism. Differentiating isolated myoblasts cells also reduces tertiary effects and variation from the age of the patients, range in time from onset to when biopsies were taken and differences in biopsy site (Table 1). These patient variables could also affect the efficiency of myotube differentiation; so, to ensure that different percentages of undifferentiated cells in the population did not impact measuring gene expression changes, the myotubes were specifically isolated by short trypsinization thus removing all myoblast contamination

(Fig. 1A and Supplementary Material, Fig. S1). To define the baseline for comparison, two age-matched healthy controls were similarly analyzed. Samples from all patients and controls all yielded high-quality reads ranging between 56 and 94 million paired-end reads (Supplementary Material, Table S1).

First, we compared each individual patient against the controls. Compared with the controls, each individual patient had between 310 and 2651 upregulated genes and between 429 and 2384 downregulated genes with a false discovery rate (FDR) of 5% (Fig. 1B). The large difference in the number of differentially expressed (DE) genes between patients suggested large heterogeneity. When we calculated the intersection of DE genes in all patients, only three genes were similarly downregulated (*MTCO1P12*, *HLA-H*, *HLA-C*), and one upregulated (*MYH14*) at 5% FDR, indicating a high degree of variation between patients (Supplementary Material, Fig. S2A). *MTCO1P12* is a mitochondrially encoded pseudogene that has been reported to be severely downregulated in inflammatory bowel disease, associated with reduced mitochondrial energy production (25). *HLA-C* is a member of the MHC class I and is involved in interferon gamma signaling, while *HLA-H* is a pseudogene derived from *HLA-A* which may function in autophagy (26–28). Mutations in non-muscle myosin gene *MYH14* appear to be associated with hearing loss rather than muscle defects (29,30), although it has also been recently linked to mitochondrial fission defects (31).

The small number of intersecting genes was expected, given the heterogeneity and the number of samples. However, this kind of comparison underestimates the underlying similarity since all it takes is one patient for any given gene to miss the arbitrary 5% FDR cutoff and the gene would not be selected. Because the goal of this study is to identify common features among a sample of EDMD spectrum patients, instead of focusing on individual patients we next compared all patients together as a single group against the controls, and we denoted this analysis as G1 (one single group) throughout the text (Fig. 1). The G1 analysis revealed a set of 1127 DE genes (894 upregulated and 233 downregulated) with a 5% FDR cutoff. Preliminary examination of this set of genes across the 10 patients suggested that they might fall into two or three broadly distinct profiles with a main group comprising half of the patients that includes the patient with the classical emerin mutation. Around 60% of the genes were altered in the same direction in all 10 patients (Fig. 1 and Supplementary Material, Fig. S2B and C). Thus, while the majority of the 1127 genes behave in a similar fashion (albeit with differences in gene expression levels), there are substantial underlying differences and apparent subgroups of patients with more similar expression profiles that might reflect clinical variation in EDMD.

Hierarchical clustering identified one main subgroup comprising half of the patients that includes the EMD mutation, with the remaining patients falling more loosely into two smaller groups (Fig. 1C and D). While it is tempting to suggest these groups may represent separate EDMD spectrum subcategories, the small cohort employed precludes this conclusion. However, we reasoned that the distinction may bear clinical relevance and provide proof of principle that transcriptome analysis could assist clinicians in diagnosis and prognosis of EDMD spectrum patients. The three groupings were: group 1—Emerin p.S58Sfs*1, Tmem214 p.R179H, NET39 p.M92K, Lamin A p.R571S, FHL1 p.C224W; group 2—Sun1 p.G68D/G388S, FHL1 c.688 + 1G > A, FHL1 p.V280M; group 3—Nesprin 1 p.S6869*, Lamin A/C p.T528K (Fig. 1D). The same groups were independently identified using a principal component analysis (PCA) (Supplementary Material, Fig. S2D). Of particular note, both PCA and t-distributed stochastic neighbor embedding

Table 1. Patients and controls used in this study. List of patients and controls including gene, mutation, age at biopsy, age of onset, sex, creatine kinase (CK) levels, clinical features and the muscle groups from whence biopsies were obtained. Note that the normal range for CK levels is 25–200 U/l.

ID	Gene	Mutation	RNAseq	Metabolic analysis	Age of onset	Age	Sex	CK(U/l)	Muscle weakness	Contractions	Cardiac defects	Family history	Neurological defects	Biopsy location	Other clinical information
C1	Control		Yes	Yes	35		M		No	No	No			Gastrocnemius caput laterale	
C2	Control		Yes	Yes	13		M		No	No	No			Biceps brachii	
C3	Control		Yes	Yes	43		M		No	No	No			Biceps brachii	
C4	Control		Yes	Yes	36		F		No	No	No			Vastus lateralis	
C5	Control		Yes	Yes	49		F		No	No	No			Vastus lateralis	
C6	Control		Yes	Yes	49		F		No	No	No			Vastus lateralis	
C7	Control		Yes	Yes	53		M		No	No	No			Vastus lateralis	
P1	TMEM214	p.179H	Yes	Yes	48	28	F	350	Limb girdle weakness	No	No	Sister, father pacemaker		Rectus femoris	LGMD proxis, type 2 fiber atrophy, Trendelenburg sign, Gowers sign
P2	PLPP7	p.M92K	Yes	Yes	34	16	F	400	Distal myopathy pattern	Yes	No	No		Vastus lateralis	Distal myopathy, Trendelenburg sign calf atrophy, tibial muscle more affected, Achilles contracture
P3	SUN1	p.G68D	Yes	Yes	9	10	M	>3000	Yes	Yes	Yes			Quadriceps femoris	Growth deficit, ataxia
P4	SYNE1	p.G388S	Yes	Yes	15		F	347	Yes	Yes	Yes	Father, uncle and grandmother	yes	Quadriceps femoris	Scapula alata scoliosis, mild proximal paresis, calf hypertrophy, Achilles contracture
P5	EMD	p.S585fs*1	Yes	Yes	17	12	F	278	Limb girdle weakness	Yes	Yes		no	Biceps brachii	Moderate proximal weakness, neck flexor contracture, Achilles tendon axial atrophy
P6	FHL1	c.688 + 1G > A	Yes	Yes	28		M	1000	Limb girdle weakness	Yes	Yes	Mother and cousin	no	Biceps brachii	Moderate proximal weakness, neck flexor contracture, Achilles tendon axial atrophy
P7	FHL1	p.C224W	Yes	Yes	46		M	300–1000	Limb girdle weakness	Yes	Yes	Mother and brother	no	Tibialis anterior	Moderate proximal weakness, neck flexor contracture, Achilles tendon axial atrophy
P8	FHL1	p.V280M	Yes	Yes	21		M	231	Limb girdle weakness	Yes	Yes	Brother	no	Gastrocnemius caput laterale	Rigid spine, moderate proximal weakness, neck flexor contracture, axial atrophy
P9	LMNA	p.T528K	Yes	Yes	2	9	M		Yes	Limb girdle weakness	No	No	no	Vastus lateralis	Degenerative myopathy, mild proximal weakness
P10	LMNA	p.R571S	Yes	Yes	24		F	up to 12 000	Limb girdle weakness	No	No	No	no	Vastus lateralis	Degenerative myopathy, mild proximal weakness
P11	LMNA	p.R545C	Yes	Yes	18		M		Yes	Yes	Yes			Vastus lateralis	Acrogeria
P12	LMNA	p.R543W	Yes	Yes	12		F		Yes	Yes	Yes			Tibialis anterior	Degenerative myopathy
P13	Unknown	Unknown	Yes	Yes	14		M	2800	Yes	Yes	Yes			Vastus lateralis	Myofibrillar myopathy
P14	FHL1	p.C224W	Yes	Yes	42		M	1700	Yes	Yes	Yes			Vastus lateralis	

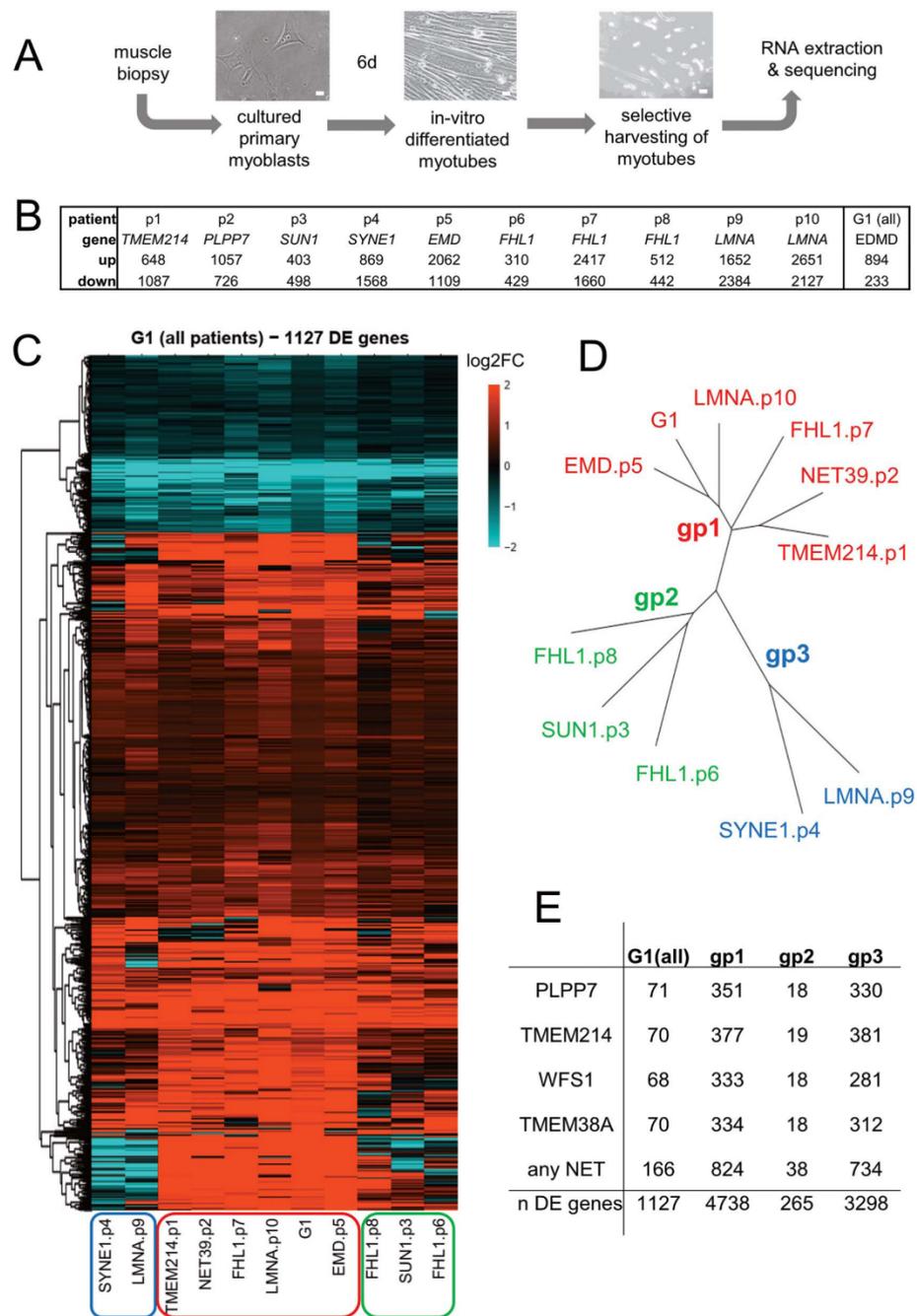


Figure 1. RNA-Seq for EDMD. (A) Workflow. Muscle biopsies were taken from the regions described in Table 1 and myoblasts recovered. These were then differentiated *in vitro* into myotubes, and myotubes selectively recovered by partial trypsinization. Myotube RNA was extracted and used for sequencing. Scale bar 20 μ m. (B) Number of genes differentially expressed (FDR 5%) for each individual patient and for all patients considered together as a single group (denoted as 'G1'). (C) Heatmap of log₂FC values for genes changing expression in EDMD patients compared with healthy controls (G1). Red is upregulated and blue is downregulated. Black indicates no change. The G1 lane is the averaged data across all patients. (D) Dendrogram showing the relationships between patients, which fall into three broad groups. (E) Overlaps with genome-organizing NETs gene targets. The number of genes altered by knockdown of muscle-specific genome-organizing NETs Wfs1, Tmem38a, NET39 or Tmem214 that were also altered in the EDMD patient cells is given. G1 refers to the analysis of all 10 patients as a single group against the healthy controls, while gp1–3 refers to the three subgroups identified. Total numbers of differentially expressed (DE) genes are given for each.

analyses revealed that clustering was independent of parameters such as patient gender or age or myotube enrichment differences (Supplementary Material, Fig. S3). Moreover, the several FHL1 and lamin mutations tested segregated into different expression subgroups. At the same time, the more recently identified EDMD mutations in TMEM214 and NET39 segregated with more classic emerlin, FHL1 and lamin A mutations, further indicating the likelihood that their genome organizing functions could mediate core EDMD pathophysiology.

The hypothesis that EDMD is a disease of genome organization misregulation is underscored by the fact that 15% of the genes changing expression in EDMD patient cells were altered by knockdown of at least one of the four muscle-specific genome-organizing NETs that we previously tested (17) (Fig. 1E). Interestingly, in that study most of the genes altered by knockdown of NET39, TMEM38A and WFS1 were non-overlapping, while those altered by knockdown of TMEM214 exhibited considerable overlap with the sets altered by each of the other NETs (17). However, here there were roughly 70 DE genes overlapping with the sets of genes altered by knockdown of each individual NET while the total number of DE genes under the regulation of any of the four NETs was 165, indicating an enrichment in the EDMD DE set for genes influenced by multiple NETs (Fig. 1E). Thus, it is not surprising that NET39 and TMEM214 were both segregated together. Another interesting observation is that the number of NET-regulated genes overlapping with group 1 and group 3 was similar, but much fewer were overlapping for group 2. This suggests that gene misregulation in groups 1 and 3 might be more strongly mediated by the muscle-specific NET-gene tethering complexes than in group 2.

Functional pathway analysis of gene expression changes in EDMD patient cells

The primary aims of this study were to determine whether a functional pathway analysis would be more effective at revealing the likely underlying EDMD pathomechanism than just looking for uniformly altered genes and, if so, to identify candidate biomarkers from the affected pathways, though these would require subsequent validation due to the limited number of patient cells available for analysis. Before using this approach with our wider set of EDMD alleles, we applied a pathway analysis to the data from the previous microarray study by Bakay and colleagues where just LMNA and EMD mutations were considered (20). We reanalyzed Bakay's EDMD data and extracted the subset of DE genes with FDR of 5% (1349 and 1452 upregulated and downregulated genes, respectively). In order to identify enriched functional categories within each set of DE genes, we used g:Profiler (32). This tool calculates the expected number of genes to be identified for any given functional category by chance and compares it to the number of genes observed. We selected categories that were significantly enriched with an FDR of 5%. The resulting list was then summarized by selecting representative classes using a similar approach to Revigo (33), but extended to other functional category databases in addition to gene ontology (GO) terms. Briefly, similarity matrices were generated by calculating pairwise Jaccard similarity indices between categories and applying hierarchical clustering to group together similar functional categories based on the genes identified. Redundancy was then reduced by choosing a representative category from each group.

The functional categories enriched in the set of EDMD-upregulated genes revealed defects in cytokine signaling, organization of the extracellular matrix (ECM), and various signaling pathways important for muscle differentiation and

function (e.g. PI3K-Akt, TGF-beta, SMADs). In addition, there was an aberrant upregulation of alternative differentiation pathways, notably adipogenesis but also angiogenesis and osteogenesis. The functions highlighted among the downregulated genes were largely related to metabolism, mitochondrial especially, as well as ribosome biogenesis, muscle contraction and myofibril assembly (Fig. 2A). Applying the same methodology to our wider set of patient alleles highlighted fewer pathways than what we observed in the Bakay EDMD data, an expected outcome for the hypothesis that sampling the wider genetic variation in the disorder might hone in on the most central pathways for the pathomechanism. Among the upregulated categories, neurogenesis and ECM-related functions stood out, as well as MAPK signaling, lipid transport and TAP binding which are linked to interferon-gamma signaling (34). One category stands out among the downregulated genes: RNA splicing (Fig. 2B). The data above used g:Profiler which is very sensitive to the number of DE genes identified because it looks for statistical overrepresentation of genes belonging to specific functional categories among a set of previously identified DE genes. By contrast, gene set enrichment analysis (GSEA) (35) does not prefilter the data and instead ranks all genes according to the difference in expression between the two conditions tested: controls and EDMD. Next, it determines whether the distribution in the ranked list for any given functional category is random or significantly enriched statistically at either end of the ranked list. This method is especially sensitive for detecting functional categories where many genes are altered by a small amount and does not consider individual gene P-values. Therefore, we also applied GSEA to our data, querying several functional genesets within the Reactome, KEGG and WikiPathways databases, found in the Molecular Signatures Database (MSigDB) (36). This approach identified a larger set of functional categories that generally expanded on those identified by g:Profiler and matched better what we observed from the Bakay EDMD geneset, with strong links to ECM organization and cytokine signaling that may be relevant to fibrosis, differentiation, metabolism and splicing (Fig. 2C).

An expansion of categories for metabolic functions included specific categories for diabetes mellitus, adipogenesis, white and brown fat differentiation, nitrogen metabolism fatty acid metabolism, retinol metabolism and many others. Similarly, there was an expansion of cytokines supporting inflammation for the fibrotic pathways and proteoglycans and elastin adding to the previous emphasis on collagens for ECM defects. Among the differences between our data and Bakay's EDMD data, two categories stand out: RNA splicing and calcium signaling, which were only observed in our data. It is unclear how much this reflects using terminally differentiated muscle material versus early stages of differentiation *in vitro*, or a factor of microarray versus RNAseq analysis. In some cases, this is most likely due to the different transcriptome platform used. For example, applying GSEA to genomic positional genesets revealed near uniform upregulation of all mitochondrially encoded genes (Figs 2D and 3A). This could not be observed on Bakay's data because the microarrays did not contain probes for mitochondrially encoded genes. The upregulation of mitochondrial transcripts could lead to increased oxidative stress (37). This finding provides yet another mechanism that could lead to metabolic dysregulation on top of the alterations already indicated by the nuclear genome transcript changes. This further underscored the need to test for actual metabolic deficits in the patient cells themselves as well as to further investigate the other functional pathways highlighted by this analysis.

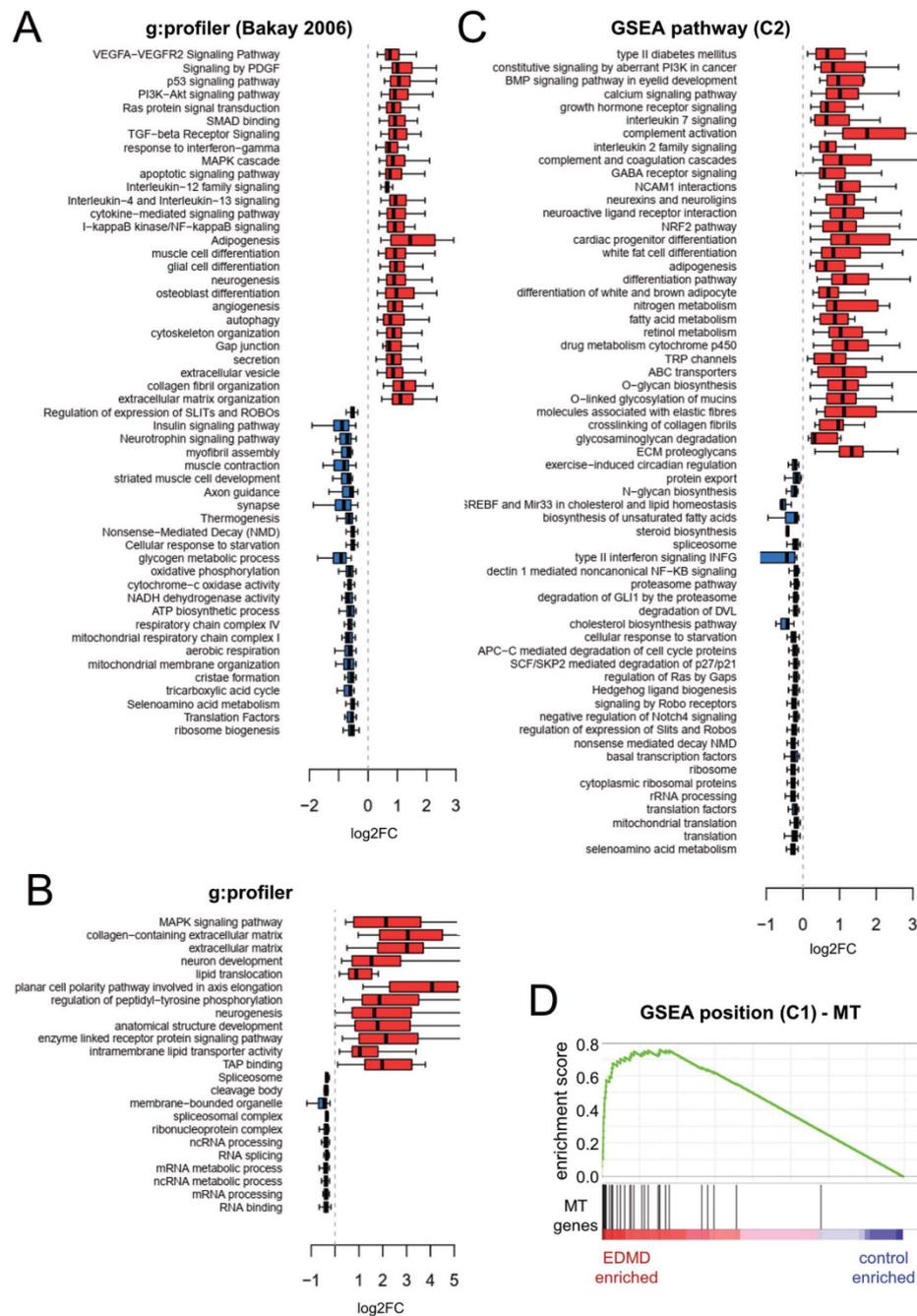


Figure 2. Search for functional categories changing in EDMD patients. (A) Box plot of log₂FC values for differentially expressed genes within significantly enriched functional categories in EDMD patients analyzed in the Bakay study using g:profiler. (B) Box plot of log₂FC values for differentially expressed genes within significantly enriched functional categories in EDMD patients analyzed in this study using g:profiler. (C) Box plot of log₂FC values for leading edge genes within significantly enriched functional categories in EDMD patients analyzed in this study using GSEA pathway (C2) geneset collection: canonical pathways). (D) GSEA enrichment plot for mitochondrially encoded genes. GSEA analysis using the C1 geneset (positional) revealed an upregulation of mitochondrially encoded genes. See Supplementary Material, Table S3 for further details.

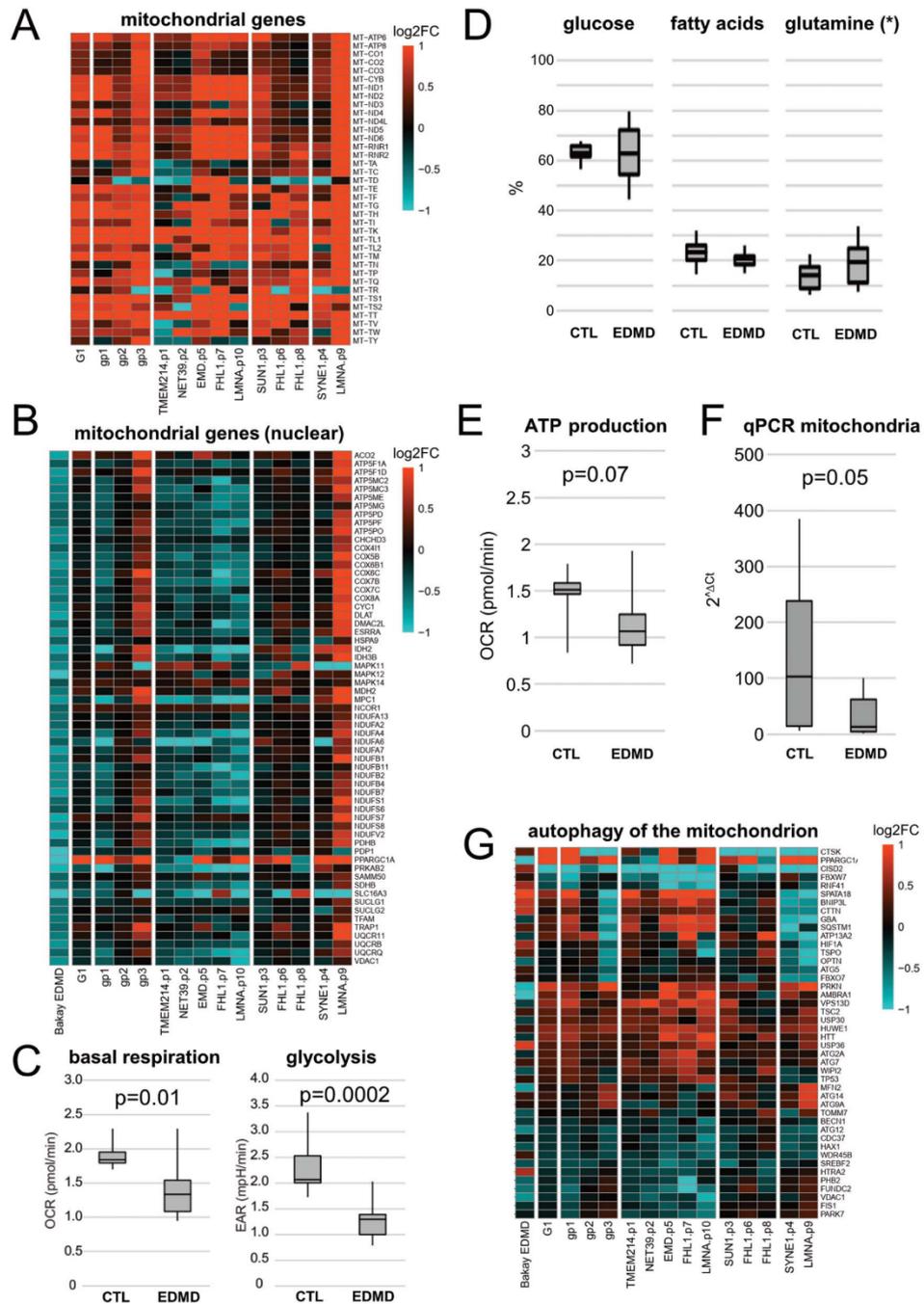


Figure 3. Metabolism changes confirmed in patients. **(A)** Heatmap of all mitochondria-encoded genes showing almost uniform upregulation in EDMD patients. **(B)** Heatmap of nuclear-encoded downregulated genes in Bakay's study that supported mitochondria function. Downregulation is also generally observed in patients belonging to group 1, but not group 2 or 3. Functional analyses performed in primary patient ($n=12$) and control ($n=7$) myoblast cultures for **(C)** glycolysis and basal respiration show a significant reduction of both in EDMD myoblast cultures. **(D)** Testing for fuel dependency we could not observe any changes (* glucose and fatty acids have been measured; glutamine calculated). **(E)** ATP production is reduced in EDMD samples. All functional experiments have been repeated in at least three independent experiments. **(F)** qPCR of mitochondria shows a reduction of mitochondria in EDMD myoblasts. **(G)** Heatmap of mitophagy genes (GO-BP, GO:0000422, Autophagy of the mitochondrion) that are differentially expressed in at least one of Bakay's EDMD, G1 analysis (all 10 patients analyzed as a single group) or G3 analysis (each of the three subgroups analyzed separately).

Detailed analysis of metabolic pathways uniformly altered in EDMD patients

Since metabolic disruption has been previously reported to affect muscle differentiation/myoblast fusion (38), we decided to investigate this further. While we identified a general upregulation of mitochondrially encoded genes, Bakay's data showed a downregulation of several classes related to mitochondrial function (Fig. 2A) which was due entirely to nuclear-encoded genes, as there were no mitochondrial genes represented in the microarrays. When we checked the behavior of those genes in our data, we did not observe the same downregulation when considering all 10 patients as a single group (Fig. 3B). However, this is largely due to variability among the patient subgroups identified earlier, suggesting a mechanistic breakdown between them. Group 1 which contained half of our patients, including emerlin and lamin A mutations, exhibited the same general downregulation of the nuclear-encoded mitochondrial genes. In contrast, group 2 displayed no alteration in gene expression, while group 3 showed upregulation although this was driven mostly by patient 9 (*LMNA*) with the other patient in the group, patient 4 (*SYNE1*), displaying very few changes. While no single gene was uniformly altered in the same direction for all patients, several genes from glycolytic and oxidative metabolism pathways, typically encoding components of mitochondrial complexes, were altered in all tested patients. Other non-mitochondrial metabolic pathways were also altered such as lipid translocation (Fig. 2C and Supplementary Material, Table S3). Interestingly, downregulation of nuclear-encoded mitochondrial genes was also generally observed in other muscular dystrophies included in the study by Bakay and colleagues (Supplementary Material, Fig. S4).

To investigate the relevance of these gene changes to cellular metabolism, we performed real-time metabolic analysis using the Seahorse XFp Extracellular Flux Analyzer. Myoblasts isolated from the above patients plus several additional EDMD patients and controls were tested, so that we had a total of 14 EDMD patients and 8 controls for this analysis (Table 1). Probing for glycolysis, a significant reduction of the extracellular acidification rate in the EDMD samples was observed (Fig. 3C). Next, we investigated mitochondrial function. When testing for basal respiration there was also a significant reduction of the oxygen consumption rate in the EDMD samples (Fig. 3C). There were no significant differences in fuel dependency, but ATP production was considerably reduced in the EDMD samples (Fig. 3D and E). The significant reduction in mitochondrial respiration raised another possibility to investigate that the absolute number of mitochondria might also be down due to problems in mitochondria biogenesis. Therefore, we quantified relative mitochondria numbers by qPCR. This revealed a clear reduction in mitochondria numbers (Fig. 3F), which with the generally elevated mitochondrial genome transcripts would suggest that a reduction in mitochondria numbers resulted in an overcompensation of expression which in turn could have resulted in inhibiting mitochondrial fission and repair. Thus, we also investigated whether genes in pathways associated with mitophagy were altered in the patients. Indeed, multiple mitophagy pathway genes were altered in all patients (Fig. 3G). Although no one individual gene was altered in all the patients, it is worth noting *CISD2* is significantly downregulated in most patients. Reduction of *CISD2* has been linked to degeneration of skeletal muscles, misregulated Ca^{2+} homeostasis and abnormalities in mitochondrial morphology in mouse (39), as well as cardiac dysfunction in humans (40).

Detailed analysis of other pathways uniformly altered in EDMD patients

Several studies suggest that the timing of several aspects of myotube fusion could underlie some of the aberrancies observed in patient muscle (41) and, though it is unclear whether fibrosis drives the pathology or is a consequence of the pathology, fibrosis has been generally observed in EDMD patient biopsies. Contributing to these processes could be several subpathways that fall variously under the larger pathways for ECM/fibrosis, cell cycle regulation and signaling/differentiation (Fig. 4A). As for the metabolic analysis, no individual genes were altered in cells from all patients, but every patient had some genes altered that could affect ECM through changes in collagen deposition (Fig. 4B). For example, 35 out of 46 collagen genes exhibited changes in at least one comparison (Bakay EDMD, G1 or one of the subgroups gp1, gp2 and gp3) and all patients had multiple of these genes altered (Supplementary Material, Fig. S5). Note that it often appears visually that the Bakay data in the first column has little change when viewing the cluster analysis, but when looking at the full set of genes listed in the matching supplemental figures there are definitely some genes strongly changing, just not necessarily the same ones. This may be due to differences in the myogenic state of the material studied: while Bakay and colleagues used muscle biopsies containing terminally differentiated muscle fibers, we focused on the earlier stages of myogenesis by *in vitro* differentiating cultured myoblasts obtained from muscle biopsies. Despite this, it is important to note that while different genes may be affected, most of the same pathways were highlighted in both Bakay's and our study. Collagens COL6A1, COL6A2, COL6A3 and COL12A1 are linked to Bethlem muscular dystrophy (42–45) and, interestingly, all these collagens were upregulated in group 1 patient cells and downregulated in group 3 patient cells (Supplementary Material, Fig. S5). Matrix metalloproteinases, which participate in the degradation and remodeling of the ECM, were also altered with 13 out of 28 matrix metalloproteinases exhibiting changes in at least one of the comparisons and all patients had multiple of these genes altered (Fig. 4C and Supplementary Material, Fig. S5). Notably the metalloproteinase MMP1 (collagenase 1), which has been proposed to resolve fibrotic tissue (46), was downregulated in all but one patient, as well as in Bakay EDMD samples. Likewise, multiple genes associated with fibrosis from FibroAtlas (Fig. 4D and Supplementary Material, Fig. S6) and with inflammation that would support fibrosis such as cytokine (Fig. 4E and Supplementary Material, Fig. S7) and INF-gamma signaling (Fig. 4F and Supplementary Material, Fig. S8) were affected in all patients. In fact, out of 941 genes in FibroAtlas there were 542 altered between all the patients. Heatmaps of gene clusters with similar expression patterns are shown in Figure 4, but more detailed individual panels with all gene names listed are shown in Supplementary Material, Figures S5–11. A few genes that stand out for their functions within the INF-gamma signaling pathway include IRF4 that is a regulator of exercise capacity through the PTG/glycogen pathway (47) and ILB1 that helps maintain muscle glucose homeostasis (48) such that both could also feed into the metabolic pathways altered.

Another subpathway critical for myogenesis and the timing and integrity of myotube fusion is cell cycle regulation. Cell cycle defects could lead to spontaneous differentiation and were previously reported in myoblasts from EDMD patients and in tissue culture cell lines expressing emerlin carrying EDMD mutations which could lead to depletion of the stem cell population (41,49). All tested EDMD patients exhibited downregulation of multiple genes

involved in the degradation of cell cycle proteins (Figs 2C and 4G, and Supplementary Material, Fig. S8) which could indicate an uncoupling of the joint regulation of cell cycle and myogenesis program (50), for example cells starting to fuse when they should still be dividing or *vice versa*.

Other pathways in addition to ECM deposition directly associated with myogenic differentiation, myoblast fusion and muscle regeneration were also altered in all patients (Fig. 4H–J and Supplementary Material, Figs S9 and S10), though, again, no single gene in these pathways was altered in the same way in all patients' cells. Poor differentiation and myotubes with nuclear clustering were observed in differentiated EDMD myoblast cultures (14) and in the mouse C2C12 differentiation system when EDMD-linked NETs were knocked down (17).

Previous work using C2C12 cells identified six genes whose products are required in the early differentiation stages and were under the regulation of muscle-specific genome-organizing NETs (17). These genes (*NID1*, *VCAM1*, *PTN*, *HGF*, *EFNA5* and *BDNF*) are critical for the timing and integrity of myotube fusion and need to be expressed early in myoblast differentiation but shut down later or they inhibit myogenesis (51–55). All six genes were misregulated in at least five but none were affected in all patients (Fig. 4K). In general terms, these genes were upregulated in group 1, downregulated in group 3 and mixed in group 2. All six genes were upregulated in Bakay's EDMD data, although only *NID1* and *HGF* were statistically significant at 5% FDR. Both were upregulated only in group 1 and downregulated in group 3. *PTN* showed a similar pattern of expression as *HGF* although the only statistically significant changes were for upregulation in group 1.

Several myogenic signaling pathways were altered such as MAPK, PI3K, BMP and Notch signaling, and several alternate differentiation pathways were de-repressed such as adipogenesis that could disrupt myotube formation and function (Fig. 2 and Supplementary Material, Table S3). Myogenesis and adipogenesis are two distinct differentiation routes from the same progenitor cells and whichever route is taken the other becomes repressed during normal differentiation (56,57). We previously showed that knock-out of fat- or muscle-specific genome organizing NETs yield depression of the alternate differentiation pathway (17,58) and the Collas lab showed that Lamin A/C lipodystrophy point mutations yield de-repression of muscle differentiation genes in adipocytes (59). We now find here that adipogenesis genes are upregulated in both Bakay's EDMD data and our data (Fig. 2). This is especially prominent for the five patients in group 1 while group 3 showing strong downregulation of a subset of the same genes and group 2 broadly looking like an intermediate of the other two groups (Fig. 4L and Supplementary Material, Figs S11 and S12), and thus could also contribute to the metabolic defect differences between patients.

Splicing pathways uniformly altered in EDMD patients yield loss of muscle-specific splice variants

Among the downregulated functional categories, mRNA splicing stood out with many genes uniformly downregulated in all patient samples (Fig. 2B and C and Supplementary Material, Fig. S14). Because of that we decided to investigate various subcategories and we found that there was a striking and uniform upregulation of factors supporting alternative splicing (AS) while constitutive splicing factors involved in spliceosome assembly and *cis* splicing are downregulated (Fig. 5A and Supplementary Material, Fig. S15). Expression changes of as little as 10% ($\log_2FC > 0.1$) have been shown to result in biologically relevant changes for vital

proteins like kinases and splicing factors (60). We thus assume that upregulation and downregulation of whole spliceosome sub-complexes even in low \log_2FC ranges lead to significant splicing misregulation. Notably, snRNAs of the U1 spliceosomal sub-complex, responsible for 5' SS recognition, constitute as much as 20% of all downregulated splicing factors ($|\log_2FC| > 0.1$, RNU1s and RNVU1s). Interestingly, a similar sharp cut-off between alternative and constitutive splicing has been reported in myotonic dystrophy (DM1/DM2) with similar genes being affected, namely *CELFs*, *MBNLs*, *NOVA*, *SMN1/2* and *SF3A1*, among others (61). DM1 is one of the best studied splicing diseases and shares typical muscular dystrophy symptomatology with EDMD, namely progressive muscle weakness and wasting, cardiac arrhythmia and contractures. Moreover, a number of splicing changes in DM1 and DM2 also occur in other muscular dystrophies (62). Of note, *MBNL3* is 4-fold transcriptionally upregulated in EDMD compared with controls. Its protein product impairs muscle cell differentiation in healthy muscle and thus needs to be downregulated upon differentiation onset (63).

Next, we performed splicing analysis to determine whether mis-splicing could drive some of the pathway alterations observed in the EDMD samples. For this purpose, we used three different methods: DEXSeq analyses exon usage, rMATS provides information about the five most common AS events and isoform Switch Analyzer (ISA) indicates which splicing events lead to annotated isoform switches. This revealed varying amounts of alternatively spliced genes in all samples and the three subgroups (Fig. 5B). Since every method focuses on a different event type/aspect of splicing, a higher amount of unique than overlapping genes is to be expected. Accordingly, an overlap of all three methods indicates genes with exon skipping events that lead to annotated isoform switches. The number of mis-spliced genes overlapping between the three algorithms was only 1 gene, *ZNF880*, in G1. In contrast, when analyzing group 1, 2 and 3 separately, each patient grouping had many mis-spliced genes identified by all three algorithms with 18 mis-spliced genes in the intersect for group 1, the group including half of the patients, and as much as 95 in group 3 (Supplementary Material, Table S4). These genes include *Nesprin 3* (*SYNE3*), the splicing factor kinase *CLK1* and the chromatin regulator *HMG3*, all of which are potentially contributing to EDMD, given their functions. All results can be found in Supplementary Material, Table S5. The rMATS analysis includes five AS events: exon skipping (SE), intron retention (RI), mutually exclusive exons (MXE), alternative 3' splice site (A3SS) and alternative 5' splice site (A5SS) usage. Using this comprehensive dataset, we searched for AS events that are significantly differentially used ($|\text{percent-spliced-in}(\text{psi})\text{-value}| > 0.1$ and $P\text{-value} < 0.05$, Supplementary Material, Table S5). Comparing AS event inclusion between control and EDMD samples, we find thrice as much intron retention in EDMD, while all other events are similarly included as excluded. We hypothesize that this could be a result of downregulated U1 snRNAs which are necessary for proper spliceosome assembly. Supporting the likely importance of the splicing pathway to the EDMD pathomechanism, pathway analysis on these genes revealed a strong enrichment for pathways associated with metabolism, gene expression and the cytoskeleton (Fig. 5C). Moreover, group 1 and group 3 display an enrichment for myogenesis and muscle contraction. Using a custom-made set of genes either specific or relevant for muscle development and structure (see Materials and Methods section), we then scanned all significant and differential AS events (Fig. 5D and Supplementary Material, Fig. S16). Notably, mis-splicing led to the absence of many muscle-specific splice variants (Fig. 5E),

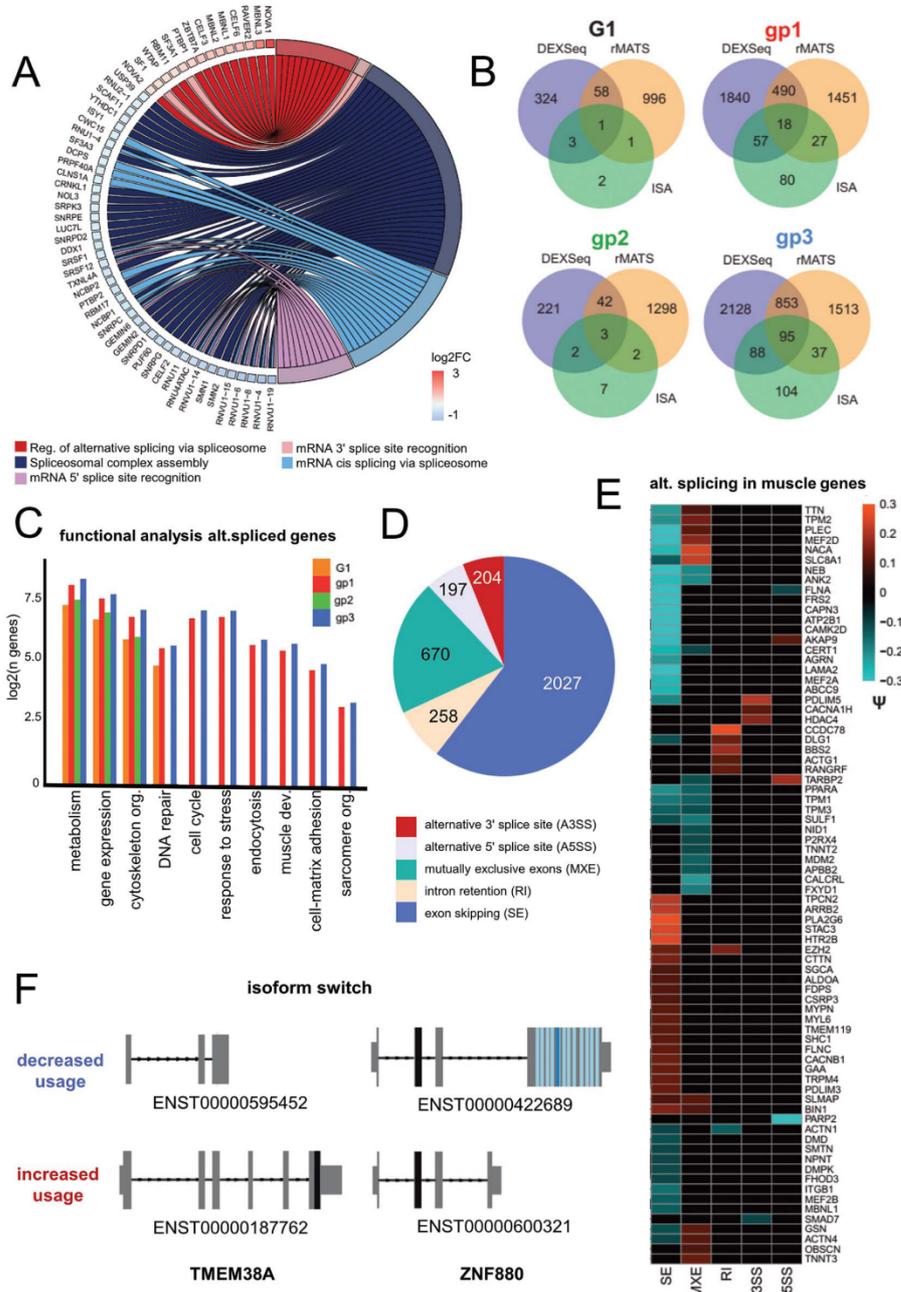


Figure 5. Analysis of splicing defects reveals loss of muscle-specific splice variants. **(A)** GOchord plot for misregulated splicing factors indicates the primary change is upregulation of alternative splicing and downregulation of constitutive splicing. Plot for gp1 is shown as a representative. DE genes $\log_2FC > |0.15|$ and $P < 0.05$. **(B)** Venn diagrams of alternatively spliced genes predicted by rMATS, DEXSeq and ISA for all samples (G1) and the three subgroups (gp1, gp2 and gp3). Overlap between all three methods indicates genes with exon skipping events that lead to annotated isoform switches. **(C)** Bar chart for functional pathways reveals an enrichment in altered splice variants for pathways associated with metabolism, gene expression, cytoskeleton organization, DNA repair, proliferation/differentiation, stress, ECM/fibrosis/and sarcomere structure. Number of genes detected displayed as \log_2 . **(E)** Pie chart of significant AS events with $\psi > |0.1|$ in gp1 detected with rMATS, which were used to scan for muscle specific-splice variants, shown as heatmap. AS events, alternative splicing events. SE = exon skipping, MXE = mutually exclusive exons, RI = intron retention, A3SS = alternative 3' splice site, A5SS = alternative 5' splice site. **(F)** Isoform switches as analyzed using isoformSwitchAnalyzer for ZNF880 and TMEM38A. ZNF880 shows the same isoform switch in all groups with preferential use of the shorter isoform that contains the KRAB domain (black), but not the zinc finger domain (blue). TMEM38A shows a clear switch from the muscle isoform to a shorter isoform.

among their vital muscle structural genes like *TTN*, *TNNT3*, *NEB*, *ACTA4* and *OBSCN* as well as developmental regulators of the *MEF2* family. Importantly, many of these genes mis-spliced in the EDMD patients are linked to a variety of other muscular dystrophies. For example, *TTN*, *CAPN3*, *PLEC* and *SGCA* are linked to Limb-Girdle muscular dystrophy (64–69), *DMD* is linked to Duchenne muscular dystrophy and Becker muscular dystrophy (70,71), and *BIN1*, *TNNT2/3* and *MBNL1* are mis-spliced in myotonic dystrophy (72) and all of these are mis-spliced and/or have missing muscle-specific splice variants in many of the patients in our cohort.

Intriguingly, one of the mis-spliced genes that also displays an isoform switch in group 3 is *TMEM38A* that has been linked to EDMD (10). The altered splicing map for *TMEM38A* reveals that not only is its expression highly elevated in EDMD patients ($\log_2FC = 2.9$) but also that the protein-coding isoform displays a higher usage relative to abundance compared with the non-coding isoform (Fig. 5F). Many other notable mis-spliced genes are involved in myotube fusion such as the previously mentioned *NID1* that is under spatial genome positioning control of *NET39*, another of the genome organizing NETs causative of EDMD. Most compellingly, three mis-spliced genes having to do with myogenesis/myotube fusion had muscle-specific splice variants absent in all patients (G1 rMATS). These were *CLCC1* whose loss yields muscle myotonia (73), *HLA-A/B* that disappears during myogenesis and is linked as a risk factor for idiopathic inflammatory myopathies (74,75), and *SMAD2* that shuts down myoblast fusion (76). The above examples were found in all patients within a particular group, but not always amongst all patients from the study or determined by all algorithms; however, there were also some mis-spliced genes that are potentially even more interesting because they were mis-spliced in all patients and with all three algorithms yielding the same results. One of these was *ZNF880*. While overall transcript numbers remained similar, the isoform predominantly expressed in control cells, *ENST00000422689*, is strongly downregulated in group 1 while the shorter isoform, *ENST00000600321*, is strongly upregulated (Fig. 5F). Interestingly, the dominant isoform in EDMD loses the zinc finger domain (light and dark blue) and is left with the repressive KRAB domain (black). Little is known about *ZNF880* except that it has an unclear role in breast and rectal cancer (77,78), and additional experiments are necessary to elucidate its role in EDMD.

miRNA-Seq analysis of EDMD patient cells

Changes in miRNA levels have been observed in a number of muscular dystrophies and are often used as biomarkers (79–81), but a comprehensive investigation of miRNA levels in EDMD has thus far not been engaged. Thus, the *in vitro* differentiated EDMD patient cells used for the preceding analysis were also analyzed by miRNA-Seq. We identified 28 differentially expressed miRNAs with some variation among patients (Fig. 6A). We extracted their putative targets from the miRDB database (<http://mirdb.org>) and selected those targets whose expression changed in the opposite direction of the miRNAs. Pathway analysis revealed misregulation of miRNAs largely associated with the same pathways that were misregulated from the RNA-Seq data, e.g. metabolism, ECM/fibrosis and signaling/differentiation (Fig. 6B and Supplementary Material, Table S6). More specifically, for metabolism 9 of the misregulated miRNA were linked to metabolic functions and with only partial overlap another 9 linked to mitochondria function, for ECM/fibrosis 19 of the misregulated miRNAs were linked to ECM and again with only partial overlap 10 to fibrosis and 13 to cytokines and inflammation. As noted before the ECM category in

addition to potentially contributing to fibrosis is also relevant for myotube fusion along with cell cycle regulation that was targeted by 13 misregulated miRNAs and myogenesis that was targeted by 5 miRNAs. Several misregulated miRNAs had functions relating to alternative differentiation pathways with 4 relating to adipogenesis, 12 to neurogenesis, 17 to angiogenesis and for signaling there were 9 misregulated miRNAs affecting MAPK pathways, 8 for Akt signaling, 1 for JAK–STAT signaling, 5 for TGF-beta signaling, 2 for Notch signaling and 3 for TLR signaling. Interestingly, some misregulated miRNAs were also reported as being linked to disease states such as miR-140-3p to dilated cardiomyopathy through its repressive effect on the integrin metalloproteinase gene *ADAM17* (82). As well as working within cells, miRNAs are often detected within a circulating exosomal microvesicle population that can be harvested from blood serum. This makes them especially attractive as potential biomarkers when compared with more invasive biopsies, but a much larger sample size together with more clinical information will be required to clarify these as biomarkers.

EDMD gene expression signature suggests relationships to other muscular dystrophies

The earlier Bakay study analyzed patient samples from other muscular dystrophies for comparison to EDMD. Several of the disorders show a high degree of pairwise similarity from a transcriptome point of view (Supplementary Material, Fig. S12A). In fact, many of the functional categories misregulated in EDMD are also altered in the same direction in several other muscular dystrophies, although with some differences (Supplementary Material, Fig. S12B). For example, collagens as a general category showed the highest correlations between EDMD and most other muscular disorders with the exception of hereditary spastic paraplegia, suggesting that the ECM is a major player in most muscular dystrophies. In contrast functional categories related to metabolism and mitochondrial function such as glycolysis, oxidative phosphorylation and mitophagy were more uniquely changed, i.e. weaker correlated between EDMD and other muscular disorders. Splicing also showed a weaker correlation with other muscular disorders compared with other muscle-specific functions (Supplementary Material, Fig. S12B). Because of the overall degree of similarity among neuromuscular diseases, and the fact that Bakay *et al.* used terminally differentiated muscle while our data come from a very early stage of differentiation *in vitro*, we reasoned that GSEA could be a sensitive method to compare the datasets. To this effect, we reanalyzed Bakay's data and extracted the DE genes for each of the neuromuscular diseases. We then checked our data for enrichment of those specific genesets. We then plotted the GSEA normalized enrichment score against the $-\log_{10}(P\text{-value})$. This way we can visualize positive or negative associations on the x-axis, and the higher on the y-axis the higher the confidence (lower *P*-values).

When we looked at all patients as a single group (G1), EDMD was the best match, with the highest score and lowest *P*-value (FDR 0.001), although unsurprisingly a few other diseases came very close, notably *LGMD2A* and facioscapulohumeral muscular dystrophy (FSHD) (Fig. 7A). This indicates that despite the differences in the individual DE genes between the mature muscle data from Bakay's EDMD geneset and our early *in vitro* differentiation geneset, a clear EDMD gene expression signature was displayed in our data. The next best match is Limb-Girdle muscular dystrophy 2A (*LGMD2A*), which shares some symptomatology with EDMD including muscle wasting, contractures and mild cardiac conduction defects although cardiac involvement is infrequent (83). The

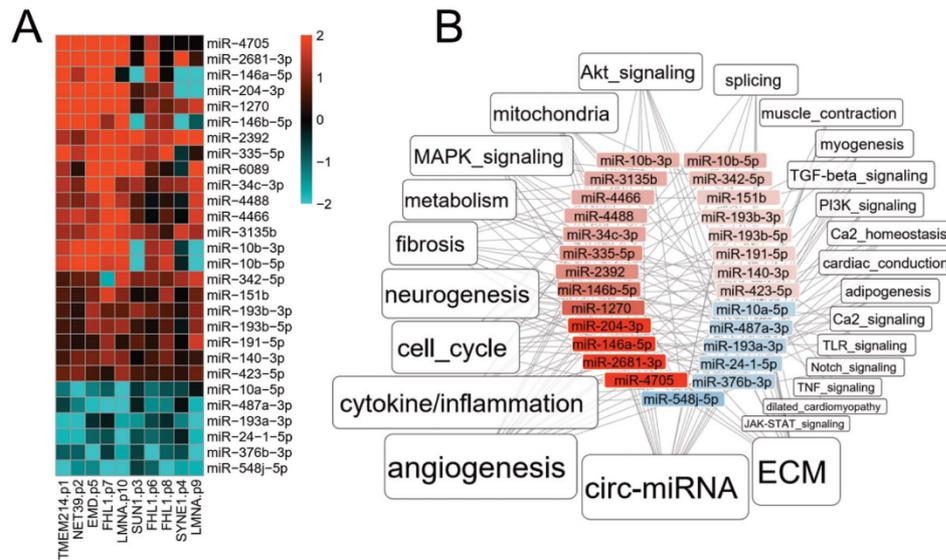


Figure 6. miRNA analysis. (A) Heatmap of miRNAs that had altered levels in EDMD patient cells compared with controls. Red indicates upregulated and blue downregulated with intensity according to the log₂FC values. (B) Overview of functional categories linked to differentially expressed miRNAs in EDMD. The label size is proportional to the number of DE miRNAs associated with each category.

differences in gene signatures that broke down the 10 patients into three EDMD patient subgroups could reflect an underlying cause of clinical disease spectrum or indicate that a group may not be adequately classified as EDMD. Therefore, we performed the same GSEA analysis on each subgroup separately. Group 1, which had both classic emerin and lamin A EDMD mutations, showed an even better match with the Bakay EDMD group which was again very close to LGMD2A but also to DMD, Becker muscular dystrophy (BMD), FSHD and Limb-Girdle muscular dystrophy 2I (LGMD2I) (Fig. 7B). LGMD2B was still separate and closer to juvenile dermatomyositis (JDM). For group 2, none of the diseases matched at 5% FDR, although the Bakay EDMD set remained the most like our set. Interestingly, two diseases exhibited an anti-correlation: DMD and BMD, which are both caused by mutations in the dystrophin gene *DMD*. In contrast, group 3 appeared to be the most distinct and in many ways opposite to group 1, which is a pattern that was often observed in the functional gene subsets analyzed (Supplementary Material, Figs S5–S11). Group 3 was anti-correlated with EDMD and most of the other muscular dystrophies, while the neurogenic amyotrophic lateral sclerosis appeared as the best match, possibly suggesting a neuronal bias in this group (Fig. 7B). This is further supported by the appearance of axonal neuropathy, ataxia, undergrowth and speech problems in one of the two patients from this group (patient 4, SYNE1; Table 1), while none of the others exhibited any signs of neuropathy.

The relationship of the patient groups segregated by gene signatures to potential differences in clinical presentation is underscored by the functional pathways enriched in each group over the others (Fig. 7C). Group 1 showed a strong enrichment of pathways associated with ECM and fibrosis, such as interferon signaling, TNF signaling, ECM organization, ECM proteoglycans, integrin cell surface interactions, collagen formation and signaling by PDGF all upregulated. Adipogenesis was also particularly promoted in group 1 compared with the others, and cardiac conduction defects were also highlighted. Group 2 was more uniquely associated with Hippo signaling and BMP2-WNT4-FOXO1 pathway and had

fewer links to ECM and fibrosis. Group 3 was more uniquely associated with metabolism, particularly upregulation of oxidative phosphorylation, mitochondrial biogenesis, glucagon signaling pathway, gluconeogenesis, glycolysis and gluconeogenesis, metabolism, TP53 regulates metabolic genes and thermogenesis pathways. This would suggest that group 1 pathophysiology may have more characteristics of fibrosis and altered myofibers, while group 2 may have more differentiation or mechanosignaling defects and group 3 more metabolic defects (Fig. 7D).

Discussion

Attempts to identify the EDMD pathomechanism or clinical biomarkers purely through gene expression signatures are limited because there is too little uniformity in differential gene expression between all patients, although there have been some promising reports using miRNA profiling or detection of cytokines in serum (81,84). We therefore engaged a functional pathway analysis using *in vitro* differentiated myotubes derived from 10 unrelated EDMD patients with known mutations in seven EDMD-linked genes. While it is difficult to detect many individual genes that were uniformly changed in all patients, we found many pathways that were affected in all patients. Thus, although different genes may have been targeted in different patients, the same functional pathway would be disrupted and thus yield a pathology with similar clinical features. Many pathways were disrupted when we re-analyzed data from the previously published Bakay study (20) and we postulated that, as they just analyzed mutations in two of the over two dozen genes linked to EDMD, analyzing a larger set of linked genes might narrow down the number of pathways to highlight those most relevant to EDMD pathophysiology. Indeed, when we considered a wider set of patients with mutations in seven different genes the set of affected pathways narrowed to the point that we could identify four likely candidate umbrella pathways.

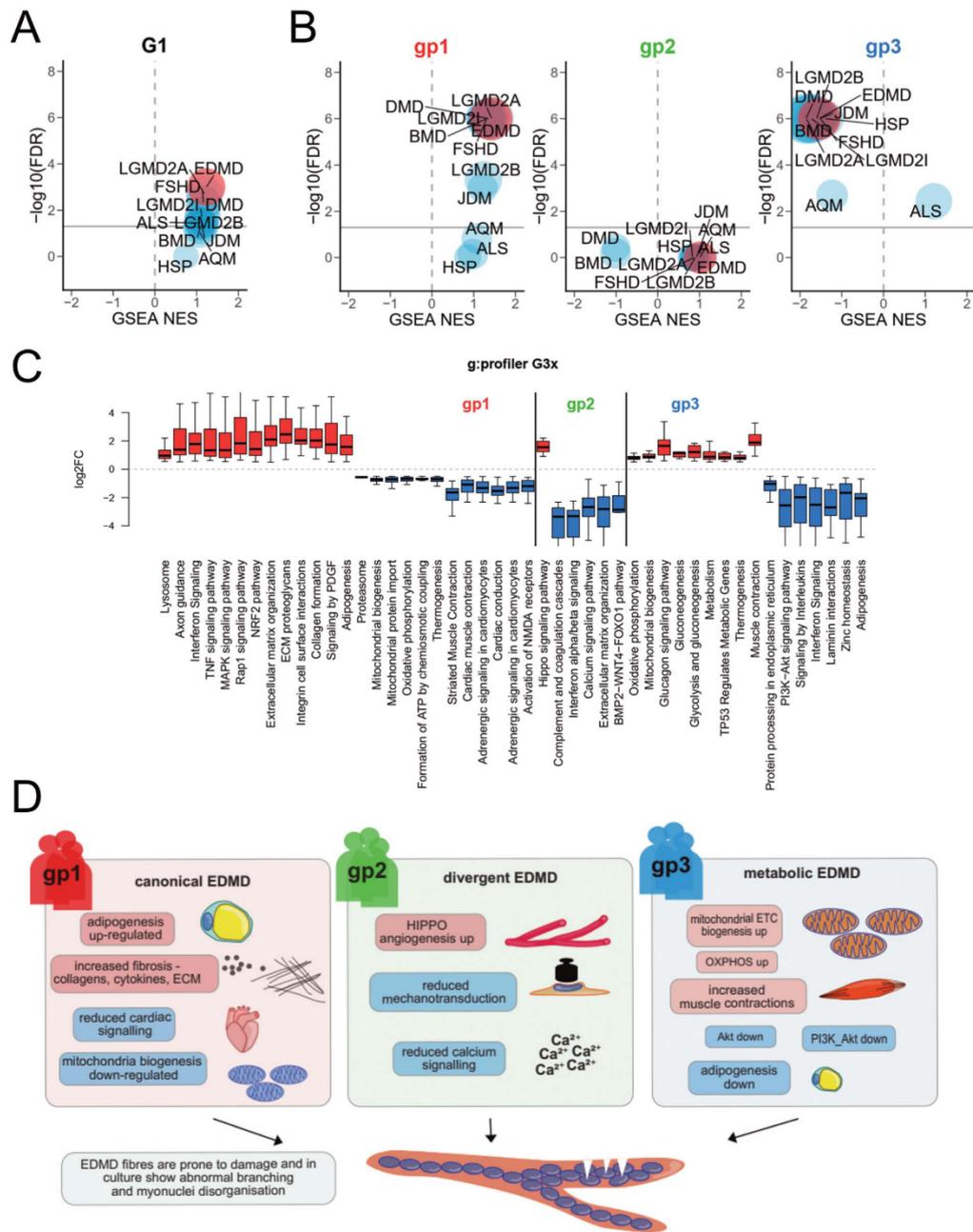


Figure 7. EDMD is distinct from other MDs. **(A)** Scatterplots from GSEA analysis comparing the new patient data from this study (all 10 patients as a single group=G1). On the x-axis, the normalized enriched score (NES) is a measurement of the enrichment of the DE geneset identified in our study compared with each of the diseases in the Bakay study. The y-axis shows the $-\log_{10}(\text{FDR})$, which is a measurement of statistical confidence. The gray horizontal line marks the 5% FDR threshold. Muscular dystrophies from the Bakay study are EDMD, Limb-Girdle muscular dystrophy 2A (LGMD2A), Limb-Girdle muscular dystrophy 2B (LGMD2B), Limb-Girdle muscular dystrophy 2I (LGMD2I), fascioscapulohumeral muscular dystrophy (FSHD), Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), juvenile dermatomyositis (JDM), acute quadriplegic myopathy (AQM), amyotrophic lateral sclerosis (ALS) and hereditary spastic paraplegia (HSP). **(B)** Same as (A), for each individual patient subgroup. **(C)** Box plot of $\log_2\text{FC}$ values for differentially expressed genes within significantly enriched functional categories for each patient subgroup compared with the rest, using g:Profiler. **(D)** Each patient subgroup was relatively more enriched for certain functional pathways than other subgroups, suggesting that treatments for example targeting different metabolic pathways for groups 1 and 3 might partially ameliorate some patient difficulties.

These four umbrella pathways all make sense for contributing to or even driving the EDMD pathomechanism (85). Disruption of metabolism pathways was consistent with the significantly reduced glycolysis and mitochondrial respiration output we showed in patient myoblasts compared with controls and it makes sense that this could lead to fatigue, weakness and muscle atrophy. ECM changes and fibrosis pathways are consistent with pathology observed in EDMD and similarly could drive some of the initial pathology and, as fibrosis accumulates, contribute to disease progression. De-repression of genes from alternate differentiation pathways and defects in myogenesis through disrupted signaling pathways and cell cycle regulation could generate aberrant myotubes to yield pathology. Finally, the last disrupted pathway of splicing yields a loss of muscle-specific splice variants that could impact on all three preceding pathways.

There is much scope for intersection between the four highlighted pathways altered in all sampled EDMD patient cells. For example, amongst the de-repressed differentiation pathways was adipogenesis that could also have impact on the metabolism pathway. Even amongst the few genes that were uniformly altered in all patients sampled, though not originally obvious, a more detailed reading of the literature leads to intersections with these pathways. For example, while the *MYH14* general upregulation did not make obvious sense for muscle defects since it is not part of the contractile machinery, it has been shown that a mutation in *MYH14* disrupts mitochondrial fission in peripheral neuropathy (31). Thus, *MYH14* could potentially feed into the mitochondrial deficits noted in the patient cells. Many of the miRNAs found to be altered in the patients feed into several of these pathways. For example, miR-2392 that is increased in all patients downregulates oxidative phosphorylation in mitochondria (86) but at the same time also is reported to promote inflammation (87). miR-140 that is up in all groups has roles in fibrosis through collagen regulation (88), is pro-adipogenic (89) and inhibits skeletal muscle glycolysis (90). miRNAs could also be used potentially prognostically between the different groups as for example miR-146a is upregulated in group 1, unchanged in group 2 and downregulated in group 3. This miRNA has a strong effect on inflammation and has been implicated in fibrosis in the heart (91). miRNAs show some promise as biomarkers, especially if they could be isolated from circulating exosome vesicles in serum. Several miRNAs have been proposed as markers for lamin A/C-associated muscular dystrophies, targeting functions such as muscle repair through TGF-beta and Wnt signaling. Some of those were also identified in our study, such as miR-335, which plays a role in muscle differentiation (81). Other miRNAs identified in that study appear to be misregulated only in one subgroup of patients but not others. For example, miR-100 and miR-127-3p are misregulated in group 1 only, while miR-136, miR-376c and miR-502-3p are only misregulated in group 3, and miR-148a is upregulated in group 1 but downregulated in group 3. Because there is so much functional overlap between miRNA targets and the pathways noted from the RNA-Seq analysis, it is unclear to what extent the gene expression changes observed could be indirect from the misregulated miRNAs. Nonetheless, there are four core functions targeted by multiple mechanisms that we argue are likely to be central to the core EDMD pathomechanism. Interestingly, the literature is filled with many examples of mutation or loss of different splicing factors causing muscle defects though no individual mis-spliced gene was identified as mediating these effects. Similarly, in myotonic dystrophy type 1 (DM1) there are many mis-spliced genes thought to contribute to the disease pathology (72). For example, the splicing factor SRSF1

that is down in most patients is important for neuromuscular junction formation in mice (92). It has to be noted that while the present study used *in-vitro* differentiated myotubes, these pathways may cross-talk not just at the level of gene expression in the myotubes themselves, but also through effects determined *in vivo* by the muscle microenvironment which may vary depending on the activation of inflammatory pathways, for example.

How so many genes become misregulated has not been experimentally proven, but for lamin A/C, emerin, Sun1, nesprin, TMEM214 and PLPP7/NET39, the fact that mutations to all individually yield many hundreds of gene expression changes with considerable overlap strongly suggests that they function in a complex at the nuclear envelope to direct genome organization. Knockdown of Tmem214 and NET39 as well as several other muscle-specific NETs each alters the position and expression of hundreds of genes (17). Separately it was found that lamins and NETs, including emerin, function together in distinct complexes involved in tethering genes to the nuclear envelope in fibroblasts (93) and in muscle cells (94). Thus, disruption of emerin, lamin A/C or any other component of these tethering complexes could yield sufficiently similar gene/pathway expression changes to yield the core clinical features of EDMD. We propose that the different muscle-specific NETs give specificity to a complex containing lamin A/C and emerin and that Sun1 and nesprin proteins can indirectly impact on these complexes through mediating mechanosignal transduction and FHL1 in interpreting such signals. Since 15% of all genes changing here were affected by at least one of the muscle-specific genome-organizing NETs that were tested by knockdown, this would provide a core set of genome organization and expression changes to cause the core EDMD pathology. Since the majority of genes affected by each NET tested were unique to that NET with the exception of Tmem214, this could account for other gene expression changes that drive the segregation into subgroups which could in turn contribute to clinical variation. This interpretation is consistent with the numbers of genes changing for mutations in different nuclear envelope proteins (Fig. 1B). That the genome organizing NET mutations yielded fewer genes changing than the lamin A/C mutations may be because lamin A/C mutations disrupt multiple genome tethering complexes and thus affect more genes. The segregation of the two LMNA mutation gene signatures into separate groups might reflect separate complexes for lamin C or each mutation disrupting different sets of complexes. In either case, the extreme differences in lamin mutations gene expression profiles is not entirely surprising as different lamin mutations also exhibited large differences in studies of nuclear mechanics (95); so this could also impact on mechanosignal transduction. The Sun1 mutation may have affected fewer genes because of redundancy with Sun2 in its mechanosignal transduction function while the Nesprin 1 mutation may have had more genes changing because it is more central to mechanosignal transduction. More work is needed to clarify on all these possibilities.

The FHL1 mutations add another level of complexity to EDMD as there are several splice variants of FHL1 and only the B variant (ENST00000394155) targets to the nuclear envelope (96). That EDMD is a nuclear envelope disorder is underscored by the fact that none of the FHL1 mutations occur in exons found in the much shorter C variant (ENST00000618438) and the patient 8 mutation p.V280M is in an exon unique to FHL1B. Thus, the nuclear envelope splice variant is the only one that could yield pathology in all patients, though some of the variation could come from one of the patients also expressing the mutant A splice variant (ENST00000543669).

While further work is needed to validate the correlations between the gene expression profile subgroupings and their clinical presentation and disease progression, our finding of such distinct gene expression profiles amongst clinically diagnosed EDMD patients argues that the currently used clinical phenotype spectrum umbrella of the EDMD classification may be too broad and it might be reclassified in more precise subtypes. What is clear is that the original classifications of EDMD subtypes based just on the mutated gene often allow for cases with very dissimilar gene profiles to be classified together, while similar gene signature cases are classified as separate classes. The two mutations in LMNA yielded changes in gene expression profiles that were far more different from one another than the group 1 lamin A mutation gene profile was from the TMEM214, NET39, emerin and FHL1 mutation gene signatures. Similarly, the FHL1 p.C224W mutation yielded greater gene expression differences from the other two FHL1 mutations than it did for the other proteins in group 1. Thus, EDMD might be better classified by similarities in gene expression profiles than by the particular gene mutated and our study shows proof of principle for this, even if the groupings may change slightly once a larger patient cohort can be established and examined. Regardless, these groups have distinctive gene expression and miRNA signatures that could be used as biomarkers both diagnostically and perhaps prognostically. To get to that point will require a more comprehensive modern description of clinical Gestalt phenotypes including *e.g.* imaging datasets and disease progression timelines deciphering unique groups. Importantly, and regardless of the disease nomenclature, the different pathways we found enriched for in each subgroup could be converted to clinical recommendations based on the much more conserved individual gene expression changes for each subgroup (Fig. 7C). For example, EDMD patients have been considered by some clinicians to be at risk for malignant hyperthermia (97), though a consensus was never achieved. Our data show that the three genes currently associated with malignant hyperthermia (RYR1, CACNA1S and STAC3) are all misregulated in groups 1 and 3, but not group 2 (Supplementary Material, Fig. S17). Thus, checking expression of these genes might indicate whether a patient is likely to be at risk or not. Finally, an additional new aspect coming up from our datasets is that it might be worth further investigating the role of splicing in muscle differentiation because it might be of wider relevance to muscular dystrophy beyond DM and EDMD.

Material and Methods

Patient materials

The sources of patient samples were the Muscle Tissue Culture Collection (MTCC) at the Friedrich-Baur-Institute (Department of Neurology, Ludwig-Maximilians-University, Munich, Germany) and the MRC Centre for Neuromuscular Disorders Biobank (CNDB) in London.

Ethical approval and consent to participate

All materials were obtained with written informed consent of the donor at the CNDB or the MTCC. Ethical approval of the rare diseases biological samples biobank for research to facilitate pharmacological, gene and cell therapy trials in neuromuscular disorders is covered by REC reference 06/Q0406/33 with MTA reference CNMDBL63 CT-2925/CT-1402, and for this particular study was obtained from the West of Scotland Research Ethics Service (WoSRES) with REC reference 15/WS/0069 and IRAS project ID 177946. The study conduct and design complied with the criteria set by the Declaration of Helsinki.

Myoblast culture and in vitro differentiation into myotubes

Myoblasts were grown in culture at 37°C and 5% CO₂ using a ready to use formulation for skeletal muscle (PELOBiotech #PB-MH-272-0090) and maintained in subconfluent conditions. In order to induce differentiation, the cells were grown to confluency and 24 h later the growth medium replaced with skeletal muscle differentiation medium (Cell Applications #151D-250). The differentiation medium was replaced every other day. Myotubes were selectively harvested after 6 days by partial trypsinization followed by gentle centrifugation (Supplementary Material, Fig. S1). Each differentiation experiment was performed in triplicate. To avoid batch effects, experiments were distributed over several weeks, so that no two samples or replicates were differentiated at the same time. Myotubes were stored in Trizol at -80°C until all samples were collected.

RNA extraction

Myotubes were stored in Trizol at -80°C until all samples were available. Subsequent steps were performed simultaneously for all samples. Total RNA was extracted from each sample and separated into a high molecular weight fraction (>200 nt, for mRNA-Seq) and a low molecular weight fraction (<200 nt, for miRNA-Seq) with the Qiagen RNeasy (#74134) and miRNeasy (#1038703) kits, according to the manufacturer's instructions. RNA quality was assessed with a Bioanalyzer (Agilent Technologies), and all samples had an RIN > 7, with an average of 9.4 (Supplementary Material, Table S1).

mRNA-Seq analysis

Between 3 and 5 µg of total RNA were sent to Admera Health LLC (NJ, USA) for sequencing in paired-end mode, 2x 150 nucleotides, using an Illumina HiSeq 2500 sequencer. The sequencing library was prepared with the NEBNext Ultra II kit, with RiboZero rRNA depletion (NEB #E7103, Illumina #20040526). Between 60 and 90 million paired end reads were obtained from each sample and mapped to the human genome (Hg38) with STAR v2.7.5a (98) using default parameters. Mapping quality was assessed with FastQC v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Sequencing adaptors were removed with trimmomatic v0.35 (99). Low-quality reads and mitochondrial contaminants were removed, leaving on average 70 million useful reads per sample (Supplementary Material, Table S1). Differential expression analysis was performed in R with DESeq2 v1.32.0 (100) after transcript quantitation with Salmon v1.4.0 (101). We used an FDR threshold of 5% for differential expression.

miRNA-Seq analysis

miRNA was sent to RealSeq Biosciences Inc. (SC, USA) for sequencing using an Illumina NextSeq 500 v2 sequencer in single end mode, 1 × 75 nucleotides. The sequencing library was prepared with Somagenics' Low-bias RealSeq-AC miRNA library kit (#500-00012) and quality assessed by TapeStation (Lab901/Agilent). On average, 5 million good quality reads were obtained per sample. Mapping and quality trimming was performed using the NextFlow nf-core/smrnaseq pipeline (<https://nf-co-re/smrnaseq>) with default parameters, which summarizes the reads per miRNA using the annotations from mirTop (<https://github.com/mirTop/mirtop>). Differential expression analysis was performed in R with DESeq2 v1.32.0. We used an FDR threshold of 0.2 for differential expression. Putative miRNA targets were extracted from miRDB (<https://mirdb.org/>) for each differentially expressed miRNA and

their expression compared against the miRNA. We kept as potential targets those genes whose expression changed in the opposite direction of the miRNA.

Bakay muscular dystrophy dataset analysis

Normalized (MAS5.0) microarray transcriptome data for a panel of 11 muscular dystrophies and healthy controls were downloaded from the Gene Expression Omnibus database (GEO), accession GSE3307 (<https://www.ncbi.nlm.nih.gov/geo/>). Differential expression analysis comparing each disease to the controls was performed using Limma 3.48.1 (102). We used an FDR threshold of 5% for differential expression.

Functional analyses

Functional analyses were performed with g:Profiler (103) and Gene Set Enrichment Analysis (GSEA v4.1.0) (35) tools. g:Profiler was used to determine enriched categories within a set of DE genes, with an FDR of 5% as threshold. GSEA was performed with default parameters, in particular using 'Signal2Noise' as ranking metric and 'meandiv' normalization mode. Redundancy in category lists was reduced by comparing the similarity between each pair of enriched categories using Jaccard similarity coefficients. Hierarchical clustering (k-means) was then applied to the resulting matrix in order to identify groups of similar functional categories, and a representative from each group chosen. Full unfiltered results are shown in [Supplementary Material, Table S3](#). Tissue-specific gene enrichment analysis was evaluated with TissueEnrich (104).

For the miRNA-Seq experiments, functional analysis was first performed using g:profiler on the set of DE miRNA genes. Then, putative targets for each miRNA were extracted and their expression compared with the relevant miRNA. Putative targets whose expression was not altered in the opposite direction as the miRNA were removed from the list. Significant functions were displayed using Cytoscape v3.8.2 (105), with the size of the functional labels proportional to the number of miRNAs assigned to each function.

Real-time metabolic measurements

Metabolic measurements on primary human myoblast cultures were performed using a Seahorse XFp Extracellular Flux Analyzer (Agilent Technologies). For this, myoblasts of matched passage number were seeded in XFp Cell Culture Miniplates (103025-100, Agilent Technologies) at a density of 1.5×10^4 cells per well. Cell density was assessed using an automated cell counter (TC20, BioRad). Oxygen consumption rates and extracellular acidification rates were measured using the Mito Stress Test Kit and the Glycolysis Stress Test Kit (Agilent Technologies #103020-100), respectively, according to the manufacturer's instructions. Samples were measured in triplicates and each measurement was repeated between two and four times. Data were normalized to the number of cells and analyzed for each well.

Fuel dependency tests

Glucose dependency and fatty acid dependency were determined according the instruction of Agilent Seahorse XF Mito Fuel Flex Test kit (Agilent Technologies #103260-100). The glutamine dependency was determined from the glucose and fatty acid measurements.

Mitochondrial gene quantification

Reverse transcription of RNA was performed using the QuantiTect Reverse Transcription Kit (Qiagen #205311) following the manufacturer's instructions. For the reaction we used the SYBR®

Green Master Mix (Bio-Rad #1725150) and samples were run and measured on CFX Connect™ (Bio-Rad). As genome reference gene B2M (FP: 5'-TGCTGTCTCCATGTTTGATGTATCT-3'; RP: 5'-TCTCTGCTCCCCACCTCTAAGT-3') (106). Primer sequences for the mitochondrial genome were: FP: 5'-TTAACTCCACCATTAGCACC-3'; RP: 5'-GAGGATGGTGGTCAAGGGA-3' (107). Samples were analyzed using the delta Ct method.

Splice site prediction analysis

Raw data was mapped to the human genome assembly GRCh38 (hg38) and sorted by coordinate using STAR 2.7.9a (98) for analysis in DESeq2 (100) and DEXSeq (108), trimmed using an in-built trimming function for rMATS (109) or counted using Kallisto 0.48.0 (110) for isoformSwitchAnalyzer (ISA) (111). All analyses were performed for G1, gp1, gp2 and gp3 separately. Visualizations were conducted in R version 4.1.2.

DESeq2: Mapped reads were counted using FeatureCounts, then analyzed using DESeq2 (100). The R package fgsea was used for GSEA, and genes were assigned to biological pathways retrieved from MSigDB v7.5.1 (c5.go.bp.v7.5.1.symbols.gmt, 7658 gene sets, (35)). Splicing pathways and their genes were plotted using GOPlot (112).

DEXseq: Mapped reads were counted using the in-built DEXseq counting function in Python 3.9. Standard DEXseq workflow was followed. Exons with $|\logFCs| > 1$ and P-values < 0.05 were set to be significantly different.

rMATS: Standard workflow was followed, and code was executed in Python 2.7. Results were analyzed in R and set to be significantly differentially spliced with $|\psi\text{-values}| > 0.1$ and P-values < 0.05 . Pie charts displaying the distribution of event usage were generated for all groups (Supps. something). GO term enrichment analysis was performed using g:profiler2 (113). All events were searched for muscle-specific genes using a set of 867 genes relevant for muscle system process, development, structure and contraction, combined from GO terms (GO:0003012, GO:0006936, GO:0055001 and GO:0061061).

ISA: Kallisto counts were read into R and standard ISA procedure was followed, including splicing analysis using DEXseq, coding potential using CPC 2.0 (114), domain annotation using HmmerWeb Pfam 35.0 (115), signal peptides using SignalP 5.0 (116) and prediction of intrinsically unstructured proteins using IUPred2A (117). In-built visualization tools were used for splicing maps.

Data availability

Bakay et al. muscular dystrophy dataset is available at NCBI GEO with accession GSE3307.

RNA-Seq and miRNA-Seq datasets have been deposited at NCBI GEO with accession GSE204804 and GSE204826, respectively.

Supplementary Material

[Supplementary Material](#) is available at HMG online.

Conflict of Interest statement: The authors have no conflicts of interest to declare.

Funding

This work was supported by Muscular Dystrophy UK (grant number 18GRO-PG24-0248); Medical Research Council (grant number MR/R018073 to E.C.S.); Deutsche Forschungsgemeinschaft (grant number 470092532 to P.M.).

Authors' contributions

J.I.H. processed patient samples for RNA- and miRNA-Seq and analyzed the data with assistance from SW. V.T. performed splicing analysis. L.K.-B., S.H. and P.M. performed various metabolic analyses. R.C. helped with generation of figures and critical discussion. B.S. provided patient samples and inspiration. E.C.S. designed the study and wrote the manuscript.

References

- Norwood, F.L.M., Harling, C., Chinnery, P.F., Eagle, M., Bushby, K. and Straub, V. (2009) Prevalence of genetic muscle disease in Northern England: in-depth analysis of a muscle clinic population. *Brain*, **132**, 3175–3186.
- Bonne, G., Mercuri, E., Muchir, A., Urtizberea, A., Becane, H.M., Recan, D., Merlini, L., Wehnert, M., Boor, R., Reuner, U. et al. (2000) Clinical and molecular genetic spectrum of autosomal dominant Emery-Dreifuss muscular dystrophy due to mutations of the lamin A/C gene. *Ann. Neurol.*, **48**, 170–180.
- Mercuri, E., Poppe, M., Quinlivan, R., Messina, S., Kinali, M., Demay, L., Bourke, J., Richard, P., Sewry, C., Pike, M. et al. (2004) Extreme variability of phenotype in patients with an identical missense mutation in the lamin A/C gene: from congenital onset with severe phenotype to milder classic Emery-Dreifuss variant. *Arch. Neurol.*, **61**, 690–694.
- Rankin, J., Auer-Grumbach, M., Bagg, W., Colclough, K., Nguyen, T.D., Fenton-May, J., Hattersley, A., Hudson, J., Jardine, P., Josifova, D. et al. (2008) Extreme phenotypic diversity and nonpenetrance in families with the LMNA gene mutation R644C. *Am. J. Med. Genet. A*, **146A**, 1530–1542.
- Hoeltzenbein, M., Karow, T., Zeller, J.A., Warzok, R., Wulff, K., Zschesche, M., Herrmann, F.H., Grosse-Heitmeyer, W. and Wehnert, M.S. (1999) Severe clinical expression in X-linked Emery-Dreifuss muscular dystrophy. *Neuromuscul. Disord.*, **9**, 166–170.
- Emery, A.E. and Dreifuss, F.E. (1966) Unusual type of benign x-linked muscular dystrophy. *J. Neurol. Neurosurg. Psychiatry*, **29**, 338–342.
- Emery, A.E. (1989) Emery-Dreifuss syndrome. *J. Med. Genet.*, **26**, 637–641.
- Knoblauch, H., Geier, C., Adams, S., Budde, B., Rudolph, A., Zacharias, U., Schulz-Menger, J., Spuler, A., Yaou, R.B., Nurnberg, P. et al. (2010) Contractures and hypertrophic cardiomyopathy in a novel FHL1 mutation. *Ann. Neurol.*, **67**, 136–140.
- Bonne, G. and Quijano-Roy, S. (2013) Emery-Dreifuss muscular dystrophy, laminopathies, and other nuclear envelopathies. *Handb. Clin. Neurol.*, **113**, 1367–1376.
- Meinke, P., Kerr, A.R.W., Czapiewski, R., de Las Heras, J.I., Dixon, C.R., Harris, E., Kolbel, H., Muntoni, F., Schara, U., Straub, V. et al. (2020) A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism. *EBioMedicine*, **51**, 102587. <https://doi.org/10.1016>.
- Meinke, P. (2018) Molecular genetics of Emery-Dreifuss muscular dystrophy. *eLS*. <https://doi.org/10.1002/9780470015902.a0022438.pub2>.
- Bridger, J.M., Foeger, N., Kill, I.R. and Herrmann, H. (2007) The nuclear lamina. Both a structural framework and a platform for genome organization. *FEBS J.*, **274**, 1354–1361.
- Worman, H.J. and Bonne, G. (2007) "Laminopathies": a wide spectrum of human diseases. *Exp. Cell Res.*, **313**, 2121–2133.
- Meinke, P., Mattioli, E., Haque, F., Antoku, S., Columbaro, M., Straatman, K.R., Worman, H.J., Gundersen, G.G., Lattanzi, G., Wehnert, M. and Shackleton, S. (2014) Muscular dystrophy-associated SUN1 and SUN2 variants disrupt nuclear-cytoskeletal connections and myonuclear organization. *PLoS Genet.*, **10**, e1004605.
- Zhang, Q., Bethmann, C., Worth, N.F., Davies, J.D., Wasner, C., Feuer, A., Ragnauth, C.D., Yi, Q., Mellad, J.A., Warren, D.T. et al. (2007) Nesprin-1 and -2 are involved in the pathogenesis of Emery-Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. *Hum. Mol. Genet.*, **16**, 2816–2833.
- Piccus, R. and Brayson, D. (2020) The nuclear envelope: LINCing tissue mechanics to genome regulation in cardiac and skeletal muscle. *Biol. Lett.*, **16**, 20200302.
- Robson, M.I., de Las Heras, J.I., Czapiewski, R., Le Thanh, P., Booth, D.G., Kelly, D.A., Webb, S., Kerr, A.R. and Schirmer, E.C. (2016) Tissue-specific gene repositioning by muscle nuclear membrane proteins enhances repression of critical developmental genes during myogenesis. *Mol. Cell*, **62**, 834–847.
- Wilkie, G.S., Korfali, N., Swanson, S.K., Malik, P., Srsen, V., Batrakou, D.G., de las Heras, J., Zuleger, N., Kerr, A.R., Florens, L. and Schirmer, E.C. (2011) Several novel nuclear envelope transmembrane proteins identified in skeletal muscle have cytoskeletal associations. *Mol. Cell. Proteomics*, **10**, M110.003129.
- Zuleger, N., Boyle, S., Kelly, D.A., de Las Heras, J.I., Lazou, V., Korfali, N., Batrakou, D.G., Randles, K.N., Morris, G.E., Harrison, D.J. et al. (2013) Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. *Genome Biol.*, **14**, R14.
- Bakay, M., Wang, Z., Melcon, G., Schiltz, L., Xuan, J., Zhao, P., Sartorelli, V., Seo, J., Pegoraro, E., Angelini, C. et al. (2006) Nuclear envelope dystrophies show a transcriptional fingerprint suggesting disruption of Rb-MyoD pathways in muscle regeneration. *Brain*, **129**, 996–1013.
- Melcon, G., Kozlov, S., Cutler, D.A., Sullivan, T., Hernandez, L., Zhao, P., Mitchell, S., Nader, G., Bakay, M., Rottman, J.N., Hoffman, E.P. and Stewart, C.L. (2006) Loss of emerin at the nuclear envelope disrupts the Rb1/E2F and MyoD pathways during muscle regeneration. *Hum. Mol. Genet.*, **15**, 637–651.
- Bione, S., Maestrini, E., Rivella, S., Mancini, M., Regis, S., Romeo, G. and Toniolo, D. (1994) Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat. Genet.*, **8**, 323–327.
- Bonne, G., Di Barletta, M.R., Varnous, S., Becane, H.M., Hammouda, E.H., Merlini, L., Muntoni, F., Greenberg, C.R., Gary, F., Urtizberea, J.A. et al. (1999) Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nat. Genet.*, **21**, 285–288.
- Gueneau, L., Bertrand, A.T., Jais, J.P., Salih, M.A., Stojkovic, T., Wehnert, M., Hoeltzenbein, M., Spuler, S., Saitoh, S., Verschuere, A. et al. (2009) Mutations of the FHL1 gene cause Emery-Dreifuss muscular dystrophy. *Am. J. Hum. Genet.*, **85**, 338–353.
- Ruiz, E., Penrose, H.M., Heller, S., Nakhoul, H., Baddoo, M., Flemington, E.F., Kandil, E. and Savkovic, S.D. (2020) Bacterial TLR4 and NOD2 signaling linked to reduced mitochondrial energy function in active inflammatory bowel disease. *Gut Microbes*, **11**, 350–363.
- Campbell, I.L., Bizilj, K., Colman, P.G., Tuck, B.E. and Harrison, L.C. (1986) Interferon-gamma induces the expression of HLA-A, B, C but not HLA-DR on human pancreatic beta-cells. *J. Clin. Endocrinol. Metab.*, **62**, 1101–1109.

27. Browne, S.K., Roesser, J.R., Zhu, S.Z. and Ginder, G.D. (2006) Differential IFN-gamma stimulation of HLA-A gene expression through CRM-1-dependent nuclear RNA export. *J. Immunol.*, **177**, 8612–8619.
28. Behrends, C., Sowa, M.E., Gygi, S.P. and Harper, J.W. (2010) Network organization of the human autophagy system. *Nature*, **466**, 68–76.
29. Choi, B.-O., Kang, S.H., Hyun, Y.S., Kanwal, S., Park, S.W., Koo, H., Kim, S.-B., Choi, Y.-C., Yoo, J.H., Kim, J.-W. et al. (2011) A complex phenotype of peripheral neuropathy, myopathy, hoarseness, and hearing loss is linked to an autosomal dominant mutation in MYH14. *Hum. Mutat.*, **32**, 669–677.
30. Iyadurai, S., Arnold, D.A., Kissel, J.T., Ruhno, C., McGovern, V.L., Snyder, P.J., Prior, T.W., Roggenbuck, J., Burghes, A.H. and Kolb, S.J. (2017) Variable phenotypic expression and onset in MYH14 distal hereditary motor neuropathy phenotype in a large, multigenerational North American family. *Muscle Nerve*, **56**, 341–345.
31. Almutawa, W., Smith, C., Sabouny, R., Smit, R.B., Zhao, T., Wong, R., Lee-Glover, L., Desrochers-Goyette, J., Ilamathi, H.S., Consortium, C.R.C et al. (2019) The R941L mutation in MYH14 disrupts mitochondrial fission and associates with peripheral neuropathy. *EBioMedicine*, **45**, 379–392.
32. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. and Vilo, J. (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.*, **47**, W191–W198.
33. Supek, F., Bosnjak, M., Skunca, N. and Smuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, **6**, e21800.
34. Fisk, B., Ioannides, C.G., Aggarwal, S., Wharton, J.T., O'Brian, C.A., Restifo, N. and Glisson, B.S. (1994) Enhanced expression of HLA-A,B,C and inducibility of TAP-1, TAP-2, and HLA-A,B,C by interferon-gamma in a multidrug-resistant small cell lung cancer line. *Lymphokine Cytokine Res.*, **13**, 125–131.
35. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 15545–15550.
36. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. and Mesirov, J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.
37. Miranda, S., Foncea, R., Guerrero, J. and Leighton, F. (1999) Oxidative stress and upregulation of mitochondrial biogenesis genes in mitochondrial DNA-depleted HeLa cells. *Biochem. Biophys. Res. Commun.*, **258**, 44–49.
38. L'honore, A., Commere, P.-H., Negroni, E., Pallafacchina, G., Friguet, B., Drouin, J., Buckingham, M. and Montarras, D. (2018) The role of Pitx2 and Pitx3 in muscle stem cells gives new insights into P38a MAP kinase and redox regulation of muscle regeneration. *elife*, **7**, e32991.
39. Chang, N.C., Nguyen, M. and Shore, G.C. (2012) BCL2-CISD2: an ER complex at the nexus of autophagy and calcium homeostasis? *Autophagy*, **8**, 856–857.
40. Yeh, C.-H., Chou, Y.-J., Kao, C.-H. and Tsai, T.-F. (2020) Mitochondria and calcium homeostasis: Cisd2 as a big player in cardiac ageing. *Int. J. Mol. Sci.*, **21**, 9238.
41. Meinke, P., Schneiderat, P., Srsen, V., Korfali, N., Le Thanh, P., Cowan, G.J.M., Cavanagh, D.R., Wehnert, M., Schirmer, E.C. and Walter, M.C. (2015) Abnormal proliferation and spontaneous differentiation of myoblasts from a symptomatic female carrier of X-linked Emery-Dreifuss muscular dystrophy. *Neuromuscul. Disord.*, **25**, 127–136.
42. Jöbsis, G.J., Keizers, H., Vreijling, J.P., de Visser, M., Speer, M.C., Wolterman, R.A., Baas, F. and Bolhuis, P.A. (1996) Type VI collagen mutations in Bethlem myopathy, an autosomal dominant myopathy with contractures. *Nat. Genet.*, **14**, 113–115.
43. Lampe, A.K., Dunn, D.M., von Niederhausern, A.C., Hamil, C., Aoyagi, A., Laval, S.H., Marie, S.K., Chu, M.-L., Swoboda, K., Muntoni, F. et al. (2005) Automated genomic sequence analysis of the three collagen VI genes: applications to Ullrich congenital muscular dystrophy and Bethlem myopathy. *J. Med. Genet.*, **42**, 108–120.
44. Baker, N.L., Morgelin, M., Pace, R.A., Peat, R.A., Adams, N.E., McKinlay Gardner, R.J., Rowland, L.P., Miller, G., De Jonghe, P., Ceulemans, B. et al. (2007) Molecular consequences of dominant Bethlem myopathy collagen VI mutations. *Ann. Neurol.*, **62**, 390–405.
45. Zou, Y., Zwolaneck, D., Izu, Y., Gandhi, S., Schreiber, G., Brockmann, K., Devoto, M., Tian, Z., Hu, Y., Veit, G. et al. (2014) Recessive and dominant mutations in COL12A1 cause a novel EDS/myopathy overlap syndrome in humans and mice. *Hum. Mol. Genet.*, **23**, 2339–2352.
46. Kaar, J.L., Li, Y., Blair, H.C., Asche, G., Koepsel, R.R., Huard, J. and Russell, A.J. (2008) Matrix metalloproteinase-1 treatment of muscle fibrosis. *Acta Biomater.*, **4**, 1141–1120.
47. Zhu, X., Yao, T., Wang, R., Guo, S., Wang, X., Zhou, Z., Zhang, Y., Zhuo, X., Wang, R., Li, J.Z., Liu, T. and Kong, X. (2020) IRF4 in skeletal muscle regulates exercise capacity via PTG/glycogen pathway. *Adv. Sci. (Weinh)*, **7**, 2001502.
48. Chiba, K., Tsuchiya, M., Koide, M., Hagiwara, Y., Sasaki, K., Hattori, Y., Watanabe, M., Sugawara, S., Kanzaki, M. and Endo, Y. (2015) Involvement of IL-1 in the maintenance of masseter muscle activity and glucose homeostasis. *PLoS One*, **10**, e0143635.
49. Fairley, E., Riddell, A., Ellis, J. and Kendrick-Jones, J. (2002) The cell cycle dependent mislocalisation of emerin may contribute to the Emery-Dreifuss muscular dystrophy phenotype. *J. Cell Sci.*, **115**, 341–354.
50. Mademtoglou, D., Asakura, Y., Borok, M.J., Alonso-Martin, S., Mourikis, P., Kodaka, Y., Mohan, A., Asakura, A. and Relaix, F. (2018) Cellular localization of the cell cycle inhibitor Cdkn1c controls growth arrest of adult skeletal muscle stem cells. *elife*, **7**, e33337.
51. Caruelle, D., Mazouzi, Z., Husmann, I., Delbe, J., Duchesnay, A., Gautron, J., Martelly, I. and Courty, J. (2004) Upregulation of HARP during in vitro myogenesis and rat soleus muscle regeneration. *J. Muscle Res. Cell Motil.*, **25**, 45–53.
52. Dietrich, S., Abou-Rebyeh, F., Brohmann, H., Bladt, F., Sonnenberg-Riethmacher, E., Yamaai, T., Lumsden, A., Brand-Saberi, B. and Birchmeier, C. (1999) The role of SF/HGF and c-Met in the development of skeletal muscle. *Development*, **126**, 1621–1629.
53. Mousavi, K. and Jasmin, B.J. (2006) BDNF is expressed in skeletal muscle satellite cells and inhibits myogenic differentiation. *J. Neurosci.*, **26**, 5739–5749.
54. Peng, H.B., Ali, A.A., Dai, Z., Daggett, D.F., Raulo, E. and Raulava, H. (1995) The role of heparin-binding growth-associated molecule (HB-GAM) in the postsynaptic induction in cultured muscle cells. *J. Neurosci.*, **15**, 3027–3038.
55. Stark, D.A., Karvas, R.M., Siegel, A.L. and Cornelison, D.D. (2011) Eph/ephrin interactions modulate muscle satellite cell motility and patterning. *Development*, **138**, 5279–5289.

56. Singh, R., Artaza, J.N., Taylor, W.E., Gonzalez-Cadavid, N.F. and Bhasin, S. (2003) Androgens stimulate myogenic differentiation and inhibit adipogenesis in C3H 10T1/2 pluripotent cells through an androgen receptor-mediated pathway. *Endocrinology*, **144**, 5081–5088.
57. Bryan, B.A., Mitchell, D.C., Zhao, L., Ma, W., Stafford, L.J., Teng, B.-B. and Liu, M. (2005) Modulation of muscle regeneration, myogenesis, and adipogenesis by the Rho family guanine nucleotide exchange factor GEFT. *Mol. Cell. Biol.*, **25**, 11089–11101.
58. Czapiewski, R., Batrakou, D.G., de Las Heras, J.I., Carter, R., Sivakumar, A., Sliwinska, M., Dixon, C.R., Webb, S., Lattanzi, G., Morton, N. et al. (2022) Genomic loci mispositioning in Tmem120a knockout mice yields latent lipodystrophy. *Nat. Commun.*, **13**, 321. <https://doi.org/10.1038/s41467-41021-27869-41462>.
59. Oldenburg, A.R., Delbarre, E., Thiede, B., Vigouroux, C. and Collas, P. (2014) Dereglulation of Fragile X-related protein 1 by the lipodystrophic lamin A p.R482W mutation elicits a myogenic gene expression program in preadipocytes. *Hum. Mol. Genet.*, **23**, 1151–1162.
60. Anczuków, O., Rosenberg, A.Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S.K. and Krainer, A.R. (2012) The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.*, **19**, 220–228.
61. Todorow, V., Hintze, S., Kerr, A.R.W., Hehr, A., Schoser, B. and Meinke, P. (2021) Transcriptome analysis in a primary human muscle cell differentiation model for myotonic dystrophy type 1. *Int. J. Mol. Sci.*, **22**, 8607.
62. Bachinski, L.L., Baggerly, K.A., Neubauer, V.L., Nixon, T.J., Raheem, O., Siritto, M., Unruh, A.K., Zhang, J., Nagarajan, L., Timchenko, L.T. et al. (2014) Most expression and splicing changes in myotonic dystrophy type 1 and type 2 skeletal muscle are shared with other muscular dystrophies. *Neuromuscul. Disord.*, **24**, 227–240.
63. Lee, K.-S., Cao, Y., Witwicka, H.E., Tom, S., Tapscott, S.J. and Wang, E.H. (2010) RNA-binding protein Muscleblind-like 3 (MBNL3) disrupts myocyte enhancer factor 2 (Mef2) (beta)-exon splicing. *J. Biol. Chem.*, **285**, 33779–33787.
64. Hackman, P., Vihola, A., Haravuori, H., Marchand, S., Sarparanta, J., De Seze, J., Labeit, S., Witt, C., Peltonen, L., Richard, I. et al. (2002) Tibial muscular dystrophy is a titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. *Am. J. Hum. Genet.*, **71**, 492–500.
65. Penisson-Besnier, I., Hackman, P., Suominen, T., Sarparanta, J., Huovinen, S., Richard-Cremieux, I. and Udd, B. (2010) Myopathies caused by homozygous titin mutations: limb-girdle muscular dystrophy 2J and variations of phenotype. *J. Neurol. Neurosurg. Psychiatry*, **81**, 1200–1202.
66. Gundersli, H., Talim, B., Korkusuz, P., Balci-Hayta, B., Cirak, S., Akarsu, N.A., Topaloglu, H. and Dincer, P. (2010) Mutation in exon 1f of PLEC, leading to disruption of plectin isoform 1f, causes autosomal-recessive limb-girdle muscular dystrophy. *Am. J. Hum. Genet.*, **87**, 834–841.
67. Richard, I., Broux, O., Allamand, V., Fougereuse, F., Chian-nikulchai, N., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C. et al. (1995) Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell*, **81**, 27–40.
68. Vissing, J., Barresi, R., Witting, N., Van Ghelue, M., Gammelgaard, L., Bindoff, L.A., Straub, V., Lochmuller, H., Hudson, J., Wahl, C.M. et al. (2016) A heterozygous 21-bp deletion in CAPN3 causes dominantly inherited limb girdle muscular dystrophy. *Brain*, **139**, 2154–2163.
69. Roberds, S.L., Leturcq, F., Allamand, V., Piccolo, F., Jeanpierre, M., Anderson, R.D., Lim, L.E., Lee, J.C., Tome, F.M., Romero, N.B. et al. (1994) Missense mutations in the adhalin gene linked to autosomal recessive muscular dystrophy. *Cell*, **78**, 625–633.
70. Hoffman, E.P., Brown, R.H., Jr. and Kunkel, L.M. (1987) Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell*, **51**, 919–928.
71. Kunkel, L.M., Hejtmancik, J.F., Caskey, C.T., Speer, A., Monaco, A.P., Middlesworth, W., Colletti, C.A., Bertelson, C., Muller, U., Bresnan, M. et al. (1986) Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. *Nature*, **322**, 73–77.
72. López-Martínez, A., Soblechero-Martín, P., de-la-Puente-Ovejero, L., Nogales-Gadea, G. and Arechavala-Gomez, V. (2020) An overview of alternative splicing defects implicated in myotonic dystrophy type I. *Genes*, **11**, 1109.
73. Jia, Y., Jucius, T.J., Cook, S.A. and Ackerman, S.L. (2015) Loss of Clcc1 results in ER stress, misfolded protein accumulation, and neurodegeneration. *J. Neurosci.*, **35**, 3001–3009.
74. Isa, A., Nehlin, J.O., Sabir, H.J., Andersen, T.E., Gaster, M., Kassem, M. and Barington, T. (2010) Impaired cell surface expression of HLA-B antigens on mesenchymal stem cells and muscle cell progenitors. *PLoS One*, **5**, e10900.
75. O'Hanlon, T.P., Carrick, D.M., Targoff, I.N., Arnett, F.C., Reveille, J.D., Carrington, M., Gao, X., Oddis, C.V., Morel, P.A., Malley, J.D. et al. (2006) Immunogenetic risk and protective factors for the idiopathic inflammatory myopathies: distinct HLA-A, -B, -Cw, -DRB1, and -DQA1 allelic profiles distinguish European American patients with different myositis autoantibodies. *Medicine*, **85**, 111–127.
76. Melendez, J., Sieiro, D., Salgado, D., Morin, V., Dejardin, M.-J., Zhou, C., Mullen, A.C. and Marcelle, C. (2021) TGFβ signalling acts as a molecular brake of myoblast fusion. *Nat. Commun.*, **12**, 749.
77. Yue, L., Wentao, L., Xin, Z., Jingjing, H., Xiaoyan, Z., Na, F., Tonghui, M. and Dalin, L. (2020) Human epidermal growth factor receptor 2-positive metastatic breast cancer with novel epidermal growth factor receptor -ZNF880 fusion and epidermal growth factor receptor E114K mutations effectively treated with pyrotinib: a case report. *Medicine (Baltimore)*, **99**, e23406.
78. Bala, P., Singh, A.K., Kavadiyala, P., Kotapalli, V., Sabarinathan, R. and Bashyam, M.D. (2021) Exome sequencing identifies ARID2 as a novel tumor suppressor in early-onset sporadic rectal cancer. *Oncogene*, **40**, 863–874.
79. Cacchiarelli, D., Legnini, I., Martone, J., Cazzella, V., D'Amico, A., Bertini, E. and Bozzoni, I. (2011) miRNAs as serum biomarkers for Duchenne muscular dystrophy. *EMBO Mol. Med.*, **3**, 258–265.
80. Israeli, D., Poupiot, J., Amor, F., Charton, K., Lostal, W., Jeanson-Leh, L. and Richard, I. (2016) Circulating miRNAs are generic and versatile therapeutic monitoring biomarkers in muscular dystrophies. *Sci. Rep.*, **6**, 28097.
81. Sylvius, N., Bonne, G., Straatman, K., Reddy, T., Gant, T.W. and Shackleton, S. (2011) MicroRNA expression profiling in patients with lamin A/C-associated muscular dystrophy. *FASEB J.*, **25**, 3966–3978.
82. Fedak, P.W.M., Moravec, C.S., McCarthy, P.M., Altamentova, S.M., Wong, A.P., Skrtic, M., Verma, S., Weisel, R.D. and Li, R.-K. (2006) Altered expression of disintegrin metalloproteinases and their inhibitor in human dilated cardiomyopathy. *Circulation*, **113**, 238–245.

83. Winckler, P.B., da Silva, A.M.S., Coimbra-Neto, A.R., Carvalho, E., Cavalcanti, E.B.U., Sobreira, C.F.R., Marrone, C.D., Machado-Costa, M.C., Carvalho, A.A.S., Feio, R.H.F. et al. (2019) Clinicogenetic lessons from 370 patients with autosomal recessive limb-girdle muscular dystrophy. *Clin. Genet.*, **96**, 341–353.
84. Bernasconi, P., Carboni, N., Ricci, G., Siciliano, G., Politano, L., Maggi, L., Mongini, T., Vercelli, L., Rodolico, C., Biagini, E. et al. (2018) Elevated TGF β 2 serum levels in Emery-Dreifuss muscular dystrophy: implications for myocyte and tenocyte differentiation and fibrogenic processes. *Nucleus*, **9**, 292–304.
85. Lehka, L. and Redowicz, M.J. (2020) Mechanisms regulating myoblast fusion: a multilevel interplay. *Semin. Cell Dev. Biol.*, **104**, 81–92.
86. Fan, S., Tian, T., Chen, W., Lv, X., Lei, X., Zhang, H., Sun, S., Cai, L., Pan, G., He, L. et al. (2019) Mitochondrial miRNA determines chemoresistance by reprogramming metabolism and regulating mitochondrial transcription. *Cancer Res.*, **79**, 1069–1084.
87. Narozna, M. and Rubis, B. (2021) Anti-SARS-CoV-2 strategies and the potential role of miRNA in the assessment of COVID-19 morbidity, recurrence, and therapy. *Int. J. Mol. Sci.*, **22**, 8663.
88. Liao, W., Liang, P., Liu, B., Xu, Z., Zhang, L., Feng, M., Tang, Y. and Xu, A. (2020) MicroRNA-140-5p mediates renal fibrosis through TGF- β 1/Smad signaling pathway by directly targeting TGFBR1. *Front. Physiol.*, **11**, 1093.
89. Zhang, X., Chang, A., Li, Y., Gao, Y., Wang, H., Ma, Z., Li, X. and Wang, B. (2015) miR-140-5p regulates adipocyte differentiation by targeting transforming growth factor- β signaling. *Sci. Rep.*, **5**, 18118.
90. Liu, L., Li, T.-M., Liu, X.-R., Bai, Y.-P., Li, J., Tang, N. and Wang, X.-B. (2019) MicroRNA-140 inhibits skeletal muscle glycolysis and atrophy in endotoxin-induced sepsis in mice via the WNT signaling pathway. *Am. J. Physiol. Cell Physiol.*, **317**, C189–C199.
91. Feng, B., Chen, S., Gordon, A.D. and Chakrabarti, S. (2017) miR-146a mediates inflammatory changes and fibrosis in the heart in diabetes. *J. Mol. Cell. Cardiol.*, **105**, 70–76.
92. Liu, Y., Luo, Y., Shen, L., Guo, R., Zhan, Z., Yuan, N., Sha, R., Qian, W., Wang, Z., Xie, Z., Wu, W. and Feng, Y. (2020) Splicing factor SRSF1 is essential for satellite cell proliferation and postnatal maturation of neuromuscular junctions in mice. *Stem Cell Rep.*, **15**, 941–954.
93. Zullo, J.M., Demarco, I.A., Pique-Regi, R., Gaffney, D.J., Epstein, C.B., Spooner, C.J., Luperchio, T.R., Bernstein, B.E., Pritchard, J.K., Reddy, K.L. et al. (2012) DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell*, **149**, 1474–1487.
94. Demmerle, J., Koch, A.J. and Holaska, J.M. (2013) Emerin and histone deacetylase 3 (HDAC3) cooperatively regulate expression and nuclear positions of MyoD, Myf5, and Pax7 genes during myogenesis. *Chromosom. Res.*, **21**, 765–779.
95. Zwerger, M., Jaalouk, D.E., Lombardi, M.L., Isermann, P., Mauerer, M., Dialynas, G., Herrmann, H., Wallrath, L.L. and Lammerding, J. (2013) Myopathic lamin mutations impair nuclear stability in cells and tissue and disrupt nucleo-cytoskeletal coupling. *Hum. Mol. Genet.*, **22**, 2335–2349.
96. Ziat, E., Mamchaoui, K., Beuvin, M., Nelson, I., Azibani, F., Spuler, S., Bonne, G. and Bertrand, A.T. (2016) FHL1B interacts with Lamin A/C and Emerin at the nuclear lamina and is Mis-regulated in Emery-Dreifuss muscular dystrophy. *Neuromuscul. Disord.*, **3**, 497–510.
97. Bonne, G., Leturcq, F. and Yaou, R.B. (2004) Emery-Dreifuss muscular dystrophy. *GeneReviews*, PMID:20301609. <https://www.ncbi.nlm.nih.gov/books/NBK1436/>.
98. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
99. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
100. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
101. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods*, **14**, 417–419.
102. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
103. Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H. and Vilo, J. (2016) g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.*, **44**, W83–W89.
104. Jain, A. and Tuteja, G. (2019) TissueEnrich: tissue-specific gene enrichment analysis. *Bioinformatics*, **35**, 1966–1967.
105. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
106. Rooney, J.P., Ryde, I.T., Sanders, L.H., Howlett, E.H., Colton, M.D., Germ, K.E., Mayer, G.D., Greenamyre, J.T. and Meyer, J.N. (2015) PCR based determination of mitochondrial DNA copy number in multiple species. *Methods Mol. Biol.*, **1241**, 23–38.
107. Nakhle, J., Ozkan, T., Lněničková, K., Briolotti, P. and Vignais, M.-L. (2020) Methods for simultaneous and quantitative isolation of mitochondrial DNA, nuclear DNA and RNA from mammalian cells. *BioTechniques*, **69**, 436–442.
108. Anders, S., Reyes, A. and Huber, W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res.*, **22**, 2008–2017.
109. Shen, S., Park, J.W., Lu, Z.-X. and Xing, Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, E5593–E5601.
110. Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, **34**, 525–527.
111. Vitting-Seerup, K. and Sandelin, A. (2019) IsoformSwitchAnalyzer: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics*, **35**, 4469–4471.
112. Walter, W., Sánchez-Cabo, F. and Ricote, M. (2015) GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics*, **31**, 2912–2914.
113. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. and Peterson, H. (2020) gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res*, **9**, ELIXIR-709.
114. Kang, Y.-J., Yang, D.-C., Kong, L., Hou, M., Meng, Y.-Q., Wei, L. and Gao, G. (2017) CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.*, **45**, W12–W16.
115. Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.*, **39**, W29–W37.

116. Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int. J. Neural Syst.*, **8**, 581–599.
117. Meszaros, B., Erdos, G. and Dosztányi, Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.

5.4 Manuscript I

Novel insights in the FSHD pathology: meta-analysis reveals misregulation of the neuromuscular junction, nuclear envelope, and spliceosome

Teresa Schätzl, **Vanessa Todorow**, Lars Kaiser, Helga Weinschrott, Benedikt Schoser, Hans-Peter Deigner, Peter Meinke, Matthias Kohl
Submitted to Communications Biology

Contributions

I contributed to the following manuscript by conducting a sub-analysis of an MRI-informed RNAseq data set from FSHD patients, including raw data processing and quality control, expression and splicing analyses and design and assembly of the figures. My findings are shown in figures 4 and 5 and described in the respective sections of the manuscript. I further participated in the supervision and consulting of the first author in both interpretation of the findings and writing of the manuscript. Manuscript I is in submission.

Novel insights in the FSHD pathology: meta-analysis reveals misregulation of the neuromuscular junction, nuclear envelope, and spliceosome

Teresa Schätzl¹, Vanessa Todorow², Lars Kaiser¹, Helga Weinschrott¹, Benedikt Schoser², Hans-Peter Deigner¹, Peter Meinke², Matthias Kohl¹

¹ Institute of Precision Medicine, Furtwangen University, Germany

² Friedrich-Baur-Institute at the Department of Neurology, LMU University Hospital, Ludwig Maximilians University, Munich, Germany

Abstract

Facioscapulohumeral muscular dystrophy (FSHD) is one of the most common autosomal dominant muscle disorders, yet no cure or amelioration exists. The clinical presentation is diverse, making it difficult to identify the actual driving pathomechanism among many downstream events. To unravel this complexity, we performed a meta-analysis of 13 original omics datasets (in total 167 FSHD and 129 control samples). Our approach confirmed previous findings about the disease pathology and specified them further. We confirmed increased expression of former proposed *DUX4* biomarkers, and furthermore impairment of the respiratory chain. Notably, the meta-analysis provides new insights about so far not reported pathways, including misregulation of neuromuscular junction protein encoding genes, downregulation of the spliceosome, and extensive alterations of nuclear envelope protein expression. Finally, we developed a publicly available shiny app, called “*meta-FSHD*” to provide a platform for researchers who want to search our analysis for genes of interest in the future.

Keywords: Facioscapulohumeral muscular dystrophy (FSHD); Meta-analysis; Transcriptomics; Neuromuscular junction; Signal transduction; Nuclear envelope; Splicing

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant inherited muscle disorder characterized by weakness and atrophy. It is one of the most common muscular dystrophies. According to the latest European epidemiological study (published in 2014), the prevalence of FSHD is 5-12 affected individuals per 100,000 population¹. The pathomechanism of FSHD has not been fully elucidated yet, and no drug cures the disease or slows its progression.

A milestone in FSHD research was the discovery of the abnormal activity of a gene called *DUX4*, which is thought to be involved in regulating the cleavage phase of embryonic development², but is otherwise silenced throughout life, except for low-level expression in the thymus and testes³. In healthy individuals, it has been reported to be hypermethylated at its locus on chromosome 4q35, a macrosatellite consisting of 11 to 100 or more D4Z4 repeats, each containing a *DUX4* gene. In FSHD, this chromosomal segment is shortened to 10 or fewer repeats, with concomitant detection of hypomethylation allowing reading of the most distant *DUX4* gene. In combination with the 4qA haplotype, which contains a polyadenylation signal, *DUX4* can be expressed due to this contraction⁴. The encoded double homeobox protein 4 (DUX4) is a transcription factor that ultimately triggers signaling cascades by activating other transcription factors and hundreds of genes, resulting in cell death⁵. However, the exact mechanisms of the signaling pathways prevalent in FSHD are still unclear. Further decelerating research progression and insight, FSHD is characterized by extreme variability in phenotype, even compared with other muscular dystrophies. This variability can be observed between affected family members and in a frequent body asymmetry, with individual muscle groups having different degrees of damage⁶. Moreover, although several studies have reported an inverse correlation between the number of D4Z4 repeats and disease severity^{7,8}, other studies examining few⁹ or comparatively many repeats¹⁰ have detected that D4Z4 allele size is not always related to clinical severity. In addition, the size of repeats in the upper range was reported to have lower methylation than predicted for a comparatively milder phenotype¹¹. It suggests that unknown additional factors are involved, which may ultimately mitigate or exacerbate disease progression. This is further supported by the fact that pathological changes in muscle morphology and integrity generally start in the second decade of life⁴ with varying muscle groups affected only

slightly or not at all, hinting at rescue or compensation mechanisms preventing *DUX4* expression and toxicity.

Given the unknown factors of FSHD pathology, advances in transcriptomics technologies over the past decades harbor great potential in elucidating molecular mechanisms and pathogenic signalling pathways using bioinformatics approaches and statistical methods. In this context, to better understand the complexity of FSHD, we performed a meta-analysis (PROSPERO ID: CRD42022330489¹²) to verify existing knowledge and to identify additional characteristics that could advance FSHD research.

Results

Our search identified a total of 11 studies, summarized in 13 datasets, that met the predefined criteria based on data from databases, related publications, and information obtained during our literature search through email contact with the corresponding authors (**Figure 1A**, detailed description of the individual phases of the meta-analysis: *Supplementary Information S1, S2, S3* and *Supplementary Table S1*; the PRISMA 2020 Checklist¹³, and Checklist for Conducting Meta-Analysis of Microarray Datasets¹⁴: *Supplementary Information S4*). The stages of unified preprocessing and statistical analysis (**Figure 1B**) culminated in the summary results of a *random effects* model¹⁵, which were confirmed in all subsequent areas by a secondary analysis approach, the *vote counting*¹⁴ (*Online Methods* and *Supplementary Information S12*). The *random effects* meta-analysis yielded 1935 significant results (adj. $p < 0.05$). These results are decreasingly sorted by SMDH ("*standardized mean difference with heteroscedastic population variances in the two groups*")¹⁶ as shown in the heatmap in **Figure 1C**, which is linked to data summarized in *Supplementary Table S2*. The results are verified by sensitivity analyses (*Supplementary Information S5* and *S6*), enrichment analyses (*random effects* model and *vote-counting* approach; *Supplementary Tables S3* and *S5-S6*), clustering of Gene Ontology (GO) terms using the Bioconductor package *simplifyEnrichment* (*Supplementary Information S7-S9*), and the results of the *meta-FSHD* app we developed (*Online Methods*).

Meta-FSHD app

We have developed a publicly available shiny app called *meta-FSHD* to provide a tool for researchers to quickly and easily get the estimated meta-analytic effect for any gene of potential interest (**Figure 1D**). The app considers all genes measured in at least 3 datasets, corresponding to 26858 unique ENSEMBL-IDs, 21080 unique ENTREZ-IDs, and 22791 unique gene names. It gives detailed parameters regarding significance, effect size, confidence interval (CI), and degree of heterogeneity. Using *meta-FSHD*, all results of the meta-analysis on significant genes can be confirmed and are traceable easily and quickly for any person (*Online Methods*).

Confirmation of previous knowledge of FSHD pathology

Several studies have already demonstrated that the clinical picture of FSHD is associated with the highly toxic expression of the *DUX4* transcription factor¹⁷⁻¹⁹. Since the gene is expressed both sporadically and in only a few myonuclei (1 in 200 - 1000 cells), detection is difficult⁵, which is why *DUX4* target genes are investigated²⁰. In the meta-analysis, *DUX4* biomarker genes are the most upregulated (**Figure 2A; Supplementary Table S2**). The highest expression is found in *H3Y1* (STD log₂-FC (95% CI) = +2.89); while histone variant *H3* was previously linked to *DUX4*²¹, it was later stated that *DUX4 induces H3Y and H3X* and mark *DUX4* target genes for expression²². Besides, as shown by our GO clusters (*Supplementary Information S7*), there are enriched Biological Process (BP) sections of up- and downregulated genes related to development, differentiation, and morphogenesis. *DUX4* generally has been described as a disruptor of muscle myogenesis that, when present at high levels, leads to apoptosis and, when present at lower levels, inhibits myogenesis²³. In this context, the BP GO clusters and the gene list (*Supplementary Table S2*) further show strong inflammasome activation and upregulation of apoptotic processes (**Figure 2B**).

In addition, the meta-analysis shows that metabolic genes are misregulated in FSHD (**Figure 2C** and BP clusters in *Supplementary Information S7*). Several studies have already pointed to mitochondrial abnormalities in FSHD^{24,25}. A recent study was able to identify mitochondria as a source of excessive reactive oxygen species (ROS) due to impairment of mitochondrial oxidative phosphorylation (OXPHOS), particularly in complex I, as an early event of *DUX4*-induced toxicity²⁶. This is of great interest because in proliferating healthy myoblasts, approximately 30% of the ATP consumed by the cells is generated by OXPHOS. In contrast, in terminally differentiated myotubes,

mitochondrial respiration is the major source of ATP (approximately 60%)²⁷. The impairment of the respiratory chain was described to lead to an immediate decrease in metabolic activity followed by a gradual increase in mitochondrial membrane potential ($\Delta\Psi_m$). The general consequences observed were apoptosis by mitochondrial ROS and impairment of mitochondrial health by lipid peroxidation²⁶. Furthermore, it was reported that impaired metabolic adaptation would lead to misdirected increase in hypoxia signaling^{26,28}. About these findings, we noticed a significant upregulation of *Hypoxia Inducible Factor 1 subunit α* (*HIF1 α*) in 10 of the 13 datasets (STD log₂-FC (95% CI) = +0.619). The meta-analysis also shows the downregulation of respiratory chain genes. In complex IV, *COX2* (which is among the approximately 5% of the most downregulated genes; STD log₂-FC (95% CI) = -0.759) and *COX3* (STD log₂-FC (95% CI) = -0.622), two of the three mitochondrial DNA-derived genes involved in cytochrome c oxidase activity²⁹, were strongly downregulated. In complex I the downregulation refers to *ND4L* (STD log₂-FC (95% CI) = -0.625), *ND5* (STD log₂-FC (95% CI) = -0.736) and *ND1* (STD log₂-FC (95% CI) = -1.07). The latter constitute three of the seven subunits of NADH dehydrogenase that originate in mitochondrial DNA and catalyze the electron transfer of NADH through the respiratory chain²⁹. *ND1* is among the 10 most downregulated genes of all 1935 significant genes (**Figure 2C**; *Supplementary Table S2*). The GO terms in the enrichment analysis highlight the relevance of the findings on dysfunctional OXPHOS in FSHD. Of the first 25 downregulated Cellular Component (CC) categories with comparatively smallest p-values, 10 refer to the mitochondrial respirasome, either complex I, IV, or both (*Supplementary Table S3*).

Another area that emerges from the meta-analysis is the impact of the nuclear lamina (NL) on FSHD pathology, as already reported in several studies^{30,31}. This is highly interesting in terms of genome regulation, especially regarding why skeletal muscle is almost exclusively affected in FSHD. Our list of significant genes (*Supplementary Table S2*) supports the findings on long-distance interactions between D4Z4, the NL, and the telomere³¹, showing altered expression of *FAT1* (STD log₂-FC (95% CI) = +0.576) and *SORBS2* (STD log₂-FC (95% CI) = +0.583). Interestingly, *FAT1*, reported previously to be lower in FSHD muscles compared to control muscles³², is downregulated in the *DUX4* model but upregulated in almost all patient datasets. As a general trend, we find genes associated with the NL expressed in opposite directions when comparing the *DUX4* model with the patient datasets. This not only refers to long-distance interactions but also to genes directly involved in the scaffold of the NL, like *LMNA* (STD log₂-FC (95% CI)

= +0.698), which has already been associated with several muscle diseases³³. The differential expression of NL-associated genes in the *DUX4* model and the patient datasets suggests mechanisms independent of *DUX4*, as previously shown concerning *FAT1*³². These findings show altered genome organization, evident from our enrichment analysis results regarding the first entry of upregulated BP pathways “*supramolecular fiber organization*” (GO:0097435; p-value: 8.592309e⁻⁰⁹; *Supplementary Information S7*). It encompasses a total of 102 genes (including *FAT1* and *SORBS2*), which are differently regulated between FSHD and controls in the meta-analysis.

New insights into FSHD

Using the *simplifyEnrichment* package for GO clustering³⁴, we became aware of genes within the CC category (in both up- and downregulated pathways) that are involved in pre- and postsynaptic processes, membranes and transitions, and neuronal processes of nerve projection (*Supplementary Information S8*). Interestingly, these neuronal aspects have not been described in FSHD, yet. To test the relevance of these findings, we examined our enrichment analysis results (*Supplementary Table S3*) and detected that upregulated signaling pathways with small p-values refer to processes within the extracellular matrix (ECM; **Figure 3D**). Since muscle fibers are located within the ECM in a three-dimensional scaffold composed of various collagens, glycoproteins, proteoglycans, and elastin, the ECM is vital for muscle contraction, integrity, and elasticity³⁵. Notably, overgrowth of the ECM, also referred to as fibrosis, is well described in FSHD and results in the hardening of the interconnective tissue, which leads to impaired muscle contraction and stiffness³⁶. Consistently, our data also show many genes upregulated at the deepest and smallest component of the ECM, the basal lamina (BL), which is adjacent to the sarcolemma of the myofibers (such as type IV collagen genes, which are reported to dominate the BL (upregulation of *COL4A1* and *COL4A2*), specific laminins (*LAMA2*, *LAMB1*, *LAMA5-AS1*), several integrins and elastin³⁷). Due to this general upregulation of ECM-related genes, we were surprised to find one very specific set of collagens downregulated, which is reported to form a functional trimer at the NMJ (**Figure 3D**). Intriguingly, the BL differs in morphology depending on whether it is synaptic or extra-synaptic. The synaptic BL is composed mainly of type IV collagen $\alpha3$ -, $\alpha4$ -, and $\alpha5$ -chains (**Figure 3A**), which, together with certain laminins (laminin $\beta2$, $\alpha4$

and $\alpha 5$) serve transmission of signals and mechanical forces that perform muscle innervation^{37,38,39}. Notably, all three synaptic type IV collagen isoforms, $\alpha 3$ - (*COL4A3*), $\alpha 4$ - (*COL4A4*), and $\alpha 5$ (*COL4A5*), are strongly downregulated in FSHD patients compared with controls throughout all datasets (*Supplementary Table S2*). *COL4A3* is even the most downregulated gene across the entire meta-analysis (STD log₂-FC (95% CI) = -1.52 (-2.38 to -0.659), **Figure 3C**). We thus hypothesised that NMJ architecture may be altered in FSHD patients probably affecting muscle innervation and searched for other factors that have been associated with impaired signal transduction at the NMJ. We discovered more than 60 genes that are misregulated on the transcriptional level (including *AGRN*, *MACF1*, *DOK7*, *WNT4*, *DVL1*, etc.; **Figure 3B**). The data show altered processes regarding ion channels and ion pumps, NMJ maintenance and formation, acetylcholine (ACh) receptor clustering, and vesicle transfer (*Supplementary Information S11*).

Nuclear envelope (NE) protein-encoding genes are misregulated in FSHD

The role of nuclear lamina-associated genes (*Supplementary Table S2*), the alteration of supramolecular fiber organization (upregulation; *Supplementary Information S7*), and downregulation of signaling pathways associated with the nuclear body and nuclear speckles (*Supplementary Information S8*) led us to examine the results of the meta-analysis in more detail. In this context, the meta-analysis demonstrated the variability of the FSHD phenotype, not only between patients but also between different muscle types, left and right muscles, or even different loci in the same muscle (*Supplementary Information S1*); some FSHD samples show gene signatures like controls depending on the extraction of muscle tissue⁴⁰ (*Supplementary Information S13*). Hence, providing additional information about the degree to which the investigated muscle is affected is necessary to differentiate between actual contributors to the phenotype and noise⁴¹. Additionally, RNA sequencing (RNA-Seq) has a significantly higher detection rate of differentially expressed genes (DEG) than microarrays and also allows the analysis of splicing^{42,43}. Therefore, we used the RNA-Seq dataset generated by Wang et al. (2019)²⁰ using muscle biopsies of 36 FSHD patients for a deeper analysis of pathways we found altered in the meta-analysis. For the said study, Wang et al. correlated the expression of four *DUX4*-regulated biomarker genes with MRI data of the lower extremities and

histopathological changes. Based on this, they divided the patients into four groups, with group 1 being similar to the controls and group 4 displaying the strongest pathology²⁰.

While the number of DEGs was already high when analyzing all samples together (1879), we found even more genes to be significantly misregulated when comparing the groups to the controls separately, with group 4 having the highest amount (8400). The reason for this could be the reduced variability within the groups, but also the increased disease pathology. Since there are so many DEGs in severely affected individuals, we wondered whether alterations in genome organization induce these changes. Several nuclear envelope transmembrane proteins (NETs), including muscle-specific ones, have been proven to be involved in genome organization and gene expression regulation⁴⁴. Notably, a NET has been described previously to be misregulated in FSHD⁴⁵, and long-distance interactions between the D4Z4 locus and the nuclear envelope have been reported³¹. Further, there are muscular dystrophies with similar symptoms to FSHD, that are linked with genes of the nuclear envelope: striated muscle laminopathies (e.g. Emery-Dreifuss-Muscular-Dystrophy (EDMD) and Limb-Girdle-Muscular-Dystrophy 1B (LGMD1B)) are caused by mutations in *EMD*, *LMNA* or *SYNE1*, among others. We thus first screened the data for a list of 386 NE associated genes that are known to be expressed and relevant in muscle^{46,47}. **Figure 4A** shows the expression of these genes in strongly affected individuals (group 4) compared to controls. This revealed that many NE protein-encoding genes were significantly differentially expressed. Noteworthy, the majority is downregulated (152 down vs. 83 up with $\log_2FC > 0.5$ and < -0.5), while the majority of all DEGs in group 4 is upregulated (2032 down vs. 6368 up). Many of these genes play a role in positioning specific genes to the NE and thereby repressing their expression^{48,49}. Thus, it is conceivable that downregulation of these NE genes might contribute to the upregulation of many DEGs found in FSHD. Importantly, we found genes associated with EDMD altered in FSHD patients, including *LMNA* (2-fold upregulated) and *EMD* (1.5-fold down **Figure 4A**, red). We then checked whether the expression of these genes correlates positively with disease severity and indeed saw a clear correlation for many of them, some of which we present in **Figure 4B**. Mutations in *LMNA*, *FHL1*, *PLPP7* and *TMEM38A* have been linked to EDMD^{48,50,51}. Notably, *PLPP7* and *TMEM38A* were shown to be important for muscle regeneration as their knockdown leads to inefficient differentiation in C2C12 myoblasts. Since they are both significantly downregulated in FSHD (*TMEM38A* in all groups, *PLPP7* in groups 3 and 4), they might contribute to

muscle weakness and wasting through impaired muscle regeneration. Therefore, we conjectured that an actual contribution of TMEM38A and PLPP7 to the FSHD phenotype would lead to expression changes of genes regulated by them. Employing a list of target genes of these two NETs generated in C2C12 mouse myoblasts⁵², we found 610 genes misregulated in FSHD that are potentially regulated by TMEM38A and PLPP7 (**Figure 4C**). When analyzing these 610 genes for a GO term analysis among others metabolism, signaling, and differentiation are enriched in FSHD (**Figure 4D**). It is noteworthy that TMEM38A and PLPP7 are only two of several genome organizing NETs being misregulated.

Components of the splicing machinery are downregulated in FSHD and result in the mis-splicing of muscle genes

Looking at the results of the random effects analysis, "spliceosomal complex" is among the top 15 results for downregulated CC clusters (*Supplemental Table S3*), whereas downregulated "mRNA splicing, via spliceosome" affects almost all datasets with significant results in the vote-counting approach (*Supplemental Table S6*).

We previously found components of alternative splicing upregulated, while constitutive splicing was downregulated in two distinct muscular dystrophies, myotonic dystrophy type I (DM1) and EDMD^{53,54}. Since we observed more and more similarities between muscular dystrophies on the molecular level, we analyzed the dataset generated by Wang et al. (2019)²⁰ about the expression of splicing components using a gene set enrichment analysis (GSEA) for splicing associated terms. While splicing factors are misregulated in all samples, groups 2 and 3 (*Supplementary Information S14* and *S15*) are divided into upregulation of alternative splicing and downregulation of constitutive splicing. As in DM1 and EDMD, there is a general downregulation of alternative *and* constitutive splicing in group 4 (**Figure 5A**). This suggests a major disruption of the splicing machinery in strongly affected FSHD patients. We expect fewer spliceosomes to assemble at splice sites, leading to many (constitutive and alternative) splice sites being unused. To investigate if this holds true, we next looked into splicing variations, primarily in group 4, using *Modeling Alternative Junction Inclusion Quantification* (MAJIQ)⁵⁵ to identify differentially used splicing events. We set the default of 10% (percent spliced in, psi/ Ψ -value ≥ 0.1) and *False Discovery Rate* (FDR) of 10% for local splicing variations (LSVs) to be significantly different. We found 730 events differentially used between

control and FSHD (*Supplementary Table S7*). Consistent with a general downregulation of splicing components in group 4, we found many exon skipping events in FSHD patients (**Figure 5B**). However, we were startled to detect a similarly large number of introns retained in controls but spliced out in FSHD patients. As a downregulation of many splicing factors could lead to fewer spliceosomes assembling to functional units, we expected, instead of an increased amount of intron skipping, less splicing in general. We thus checked these intron exclusion events separately. We found that only around 25% of these events are intron splice events, but rather an exon skipping events opposed to intron retention in controls. MAJIQ correctly identifies these events as introns retained in controls with a Ψ -value ≥ 0.1 .

To investigate the potential effect on the muscle phenotype, we next looked into the genes mis-spliced in FSHD. A GO term enrichment analysis revealed that these genes are highly relevant for muscle-specific signaling, development, and enervation (neuron projection morphogenesis; **Figure 5C**). The heatmap in **Figure 5D** shows a selection of genes differentially spliced in FSHD, many of which are involved in pathways we found misregulated in our meta-analysis, e.g. in muscle structure, mitochondria and metabolism, signaling, and splicing. Interestingly, microtubule-associated factor 1 (MACF1), which is a top hit in the meta-analysis (*Supplementary Tables S2 and S4*) and regulates myonuclear positioning at the NMJ⁵⁶, shows a preference for skipping of exon 116 in FSHD (Ψ -value 0.325). MAJIQ further detects an exon skipping event in *ATP1B3* (Ψ -value 0.359), which encodes a subunit of the Na⁺/K⁺-ATPase and is involved in the electrical excitability of muscle and nerves. We further identified several splicing factors alternatively spliced. Interestingly, *MBNL1*, a main contributor to myotonic dystrophy, shows the same double exon skipping event of exons 5 and 6 as in DM1. There are intron exclusion events in β -tubulin *TUBB6* and *FHL1*, of which the latter is a component of the nuclear envelope and differentially expressed in FSHD, as described above. *TUBB6* was proposed to act in muscle regeneration and be upregulated in dystrophy muscle⁵⁷. We also show two examples of intron retention events in controls instead of exon skipping in FSHD patients in two highly important sarcomeric genes, skeletal muscle troponin 3 (*TNNT3*) and tropomyosin 1 (*TPM1*).

Discussion

This meta-analysis confirms *DUX4* as a main driver of FSHD pathology, as shown by the upregulation of *DUX4* biomarker genes. However, many genes behave seemingly independently of *DUX4* expression. Since disease severity is sometimes comparatively mild or severe regardless of the number of D4Z4 repeats, these genes may be disease modulators under a *DUX4*-independent mechanism, as suggested by other research groups highlighting the need for new biomarkers to track disease progression and stratify patients⁵⁸. Based on the meta-analysis results, we hypothesize that gene expression alterations affecting the NMJ, NE, and spliceosome are significant factors in FSHD disease progression.

We identified downregulation of genes essential for the NMJ architecture. A consequence is very likely impairment of muscle innervation in FSHD. This might be the key to a highly interesting area of research that addresses specific strength in FSHD, as intrinsic force production capacity was found to be decreased in patients with both mild and severe FSHD, regardless of disease severity and even before the onset of fat infiltration or lower limb weakness⁵⁹. While possible reasons for these observations have been associated with early myopathic changes and non-muscular factors such as fatigue or musculoskeletal pain, our results showing that impairment of muscle innervation may play an essential role in FSHD pathology.

Moreover, researchers have suggested that the epigenetic landscape is the missing link between the FSHD phenotype and underlying genetic parameters⁶⁰. Interestingly, the significant effect of DNA methylation on 3D genome structure has been described previously⁶¹, which may also apply to FSHD pathology according to the meta-analysis results. The altered expression of NE proteins likely contributes to the high amount of differentially expressed genes in FSHD. NE proteins are involved in genome organization and subsequent gene expression control⁴⁴. Therefore, the consequence of the misregulation of NE proteins would be a secondary effect on many other genes. While this is difficult to filter out in multifactorial disease, there is clear evidence of this being true as genes that have been shown under the control of the muscle-specific NETs *PLPP7* and *TMEM38a* (in mouse myotubes⁵²) are also misregulated in FSHD, where the expression of *PLPP7* and *TMEM38A* appears to be correlated to disease severity. Similar effects have been observed in DM1⁵³. Thus, it may reflect a more general pathomechanism in muscular dystrophies. However, due to the correlation with disease

severity, the expression of NE genes may be predictive, which should be further investigated.

As an additional contributing factor, we identified altered expression of genes encoding spliceosome proteins and, subsequently, a different splicing profile affecting a bulk of genes with essential roles in muscle. In the most severely affected patients, we hypothesize that there are fewer functional spliceosomes: it seems they assemble only at every second or third splice site compared to controls, as we also observe many double exon skipping events.

Another, although already described factor, are mitochondrial impairments in FSHD²⁶. Given the confirmation of these defects by this meta-study, studies specifically testing mitochondria-targeted agents, NAD⁺ precursors, or OXPHOS modulators to slow disease progression should be endorsed.

Although we lack the biopsy material to validate all these aspects of the FSHD pathology, we could identify secondary effects on general gene expression and splicing, most likely caused by the NE and spliceosome alterations. Thus, we rate them and the NMJ alterations, as potential candidates involved in the FSHD phenotype and strongly endorse a deeper investigation.

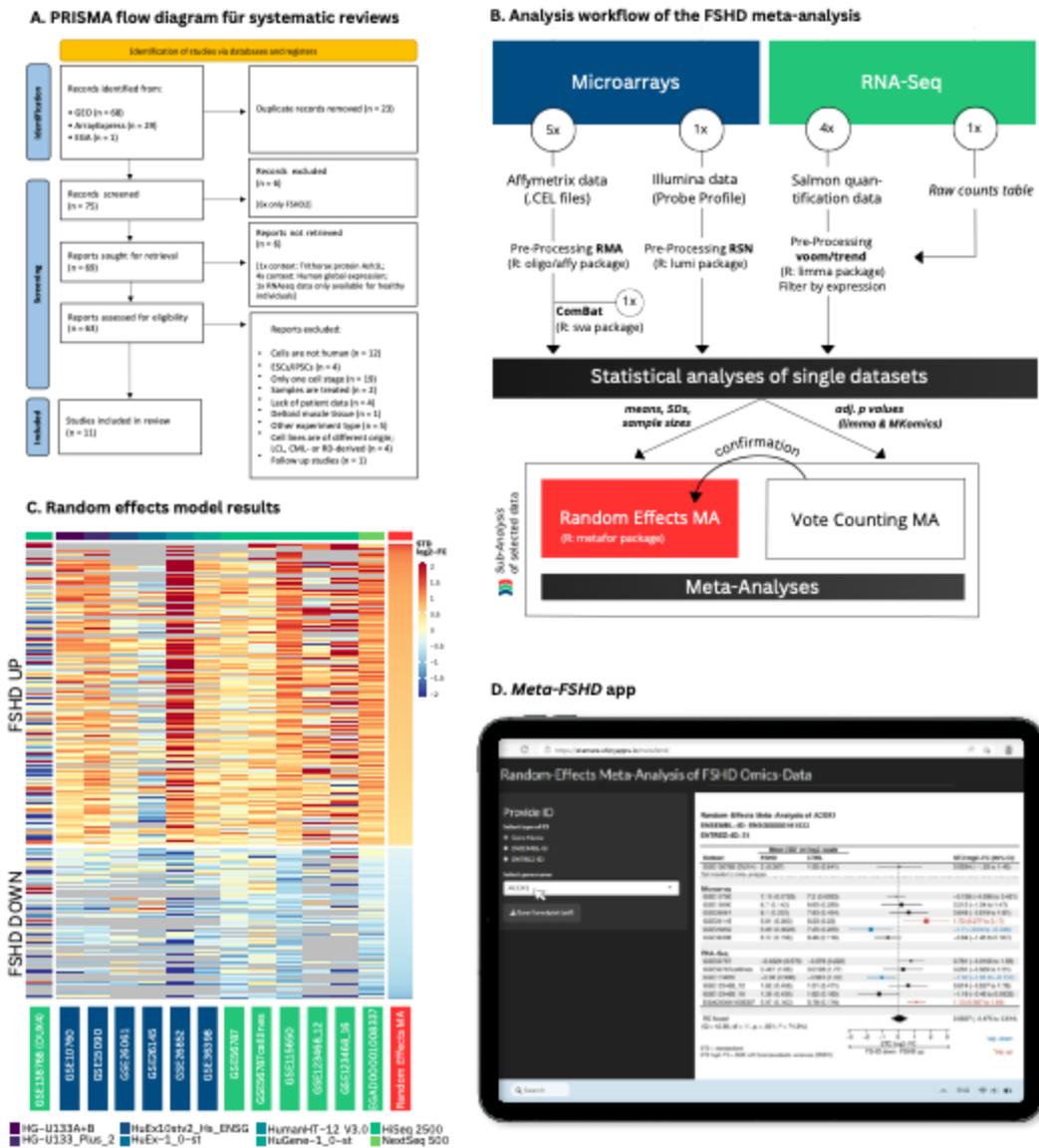


Figure 1. A) PRISMA 2020 flow diagram for new systematic reviews [65]. LCL = Lymphoblastoid Cell Line; CML = Chronic Myeloid Leukemia Cell Line; RD = Rhabdomyosarcoma. **B) Transcriptome data analysis workflow.** The meta-analysis consists of both microarray (blue) and RNA-Seq (green) datasets; it contains 5 Affymetrix (as.CEL files), 1 Illumina microarray (as bead summary data), 4 RNA-Seq (quantified via Salmon), and 1 RNA-Seq dataset for which fastq data could not be obtained for privacy reasons, but externally generated raw data (EGAD00001008337 [66], *shown in italics*). Different pre-processing methods were chosen depending on the technology; In case of 1 Affymetrix dataset (see *Supplementary Material S2*) batch effects were corrected using ComBat, an empirical Bayes approach [67] implemented in

Bioconductor package *sva* (v3.44.0) [68]. Limma *voom/trend* was used for pre-processing the RNA-Seq datasets (see *Supplementary Material S1*). The *vote-counting* approach [16] (white box) was done separately to validate the results of the *random effects* model [15] (red box), which was calculated using means, standard deviations (SDs) and sample sizes by using the package *metafor* (v1.4-0) [69] (see detailed information in *Supplementary Material S1*). The data also provided the basis for a sub-analysis, which is described in “*Genes of the nuclear envelope are mis-regulated in FSHD*” **C) Random effects model results.** The heatmap contains 1935 significant results (adj. p-value < 0.05), the exact expression of which is shown in *Supplementary Material S7*. One RNA-Seq dataset contains data from an artificial *DUX4* model and was therefore excluded from the overall calculation, but was contrasted for comparison since *DUX4*-induced gene expression has been reported to be the major molecular signature of FSHD skeletal muscle [70]. The heatmap shows a clear separation between up- and down-regulated genes. **D) Meta-FSHD app.** The gene *ACOX1* was chosen as an example for clarity, since it appears significantly upregulated (red) or downregulated (blue) in addition to normal expression (black), depending on the individual datasets. In case of *ACOX1*, the diamond crosses the vertical line of no effect. Thus, the expression of *ACOX1* (STD log₂-FC (95% CI) = +0.0697 (-0.475 to 0.614)) is not significantly different between FSHD samples and controls in the meta-analysis, with significant Cochran’s Q test (degrees of freedom = df = 11, p < 0.001) and substantial heterogeneity (I² = 74.9%) between studies (bottom left; see *Supplementary Material S5*).

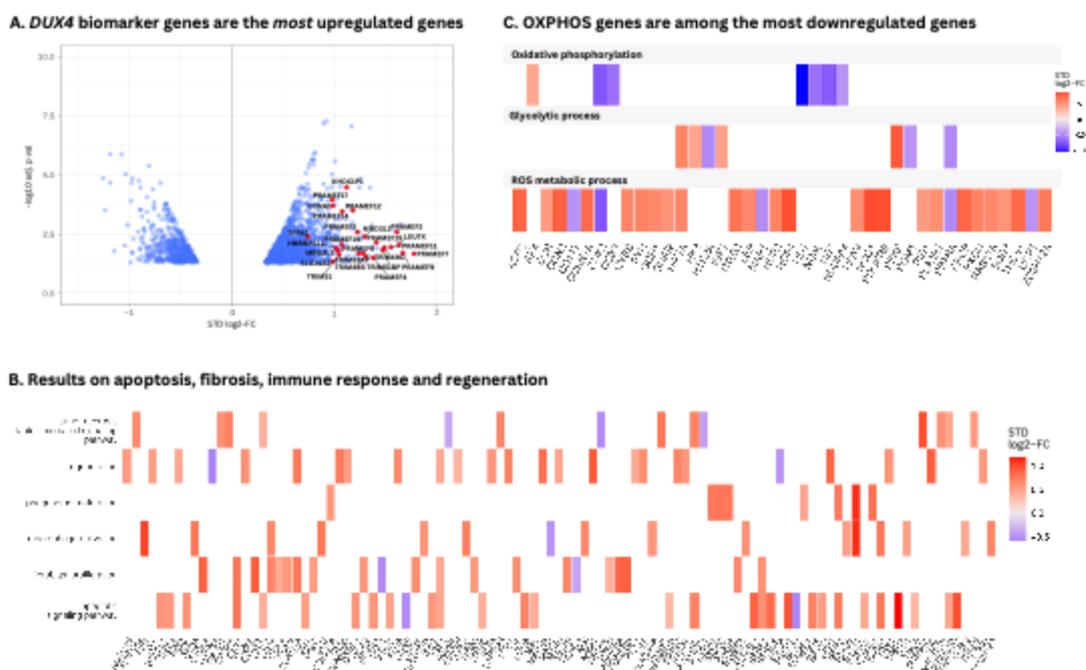


Figure 2. A) *DUX4* biomarker genes are the most upregulated genes. *DUX4* biomarker genes, as e.g. listed at Wang et al. (2019) [24] are the most upregulated genes in the entire meta-analysis (see *Supplementary Material S7*). **B) Results on apoptosis, fibrosis, immune response and regeneration.** The meta-analysis data show strong upregulation of genes associated with inflammasome, fibrosis and apoptotic processes; genes linked to regeneration are also upregulated. **C) OXPHOS genes are among the most downregulated genes.** Looking at the gene list of significantly differently expressed genes between patients and controls (see *Supplementary Material S7*), genes involved in metabolism are dysregulated. While genes involved in ROS and glycolytic processes are generally upregulated, OXPHOS genes at complex I and IV are downregulated. This refers to genes regulated by the mitochondrial genome (*COX2* and *COX3* in complex IV [34] and *ND1*, *ND4L* and *ND5* in complex I [35]), but also to other genes, such as *NDUFA4*, associated with both complex I and complex IV [71][72].

associated with the NMJ, are differentially expressed between FSHD patients and controls (such as *AGRN*, *DVLI*, *MACF1* etc.; see *Supplementary Material SX*). Interestingly, most of the upregulated genes in the *DUX4* model are expressed in the opposite manner, indicating either a mechanism independent of *DUX4* or a compensatory mechanism. As for the downregulated genes, the direction of expression is relatively similar, possibly indicating a direct consequence or concomitant effects. **C) Meta-FSHD app forest plot of *COL4A3*.** *COL4A3* is the most downregulated significantly differentially expressed gene in the entire meta-analysis with STD log₂-FC (95% CI) = -1.52 (-2.38 to -0.659). **D) Top 5 Enrichment analyses results on CC pathways.** While ECM genes are upregulated, downregulated pathways (consistent with information on downregulation of type IV collagen isoforms $\alpha 3$, $\alpha 4$, and $\alpha 5$; see *Supplementary Material S7*) indicate problems at the NMJ.

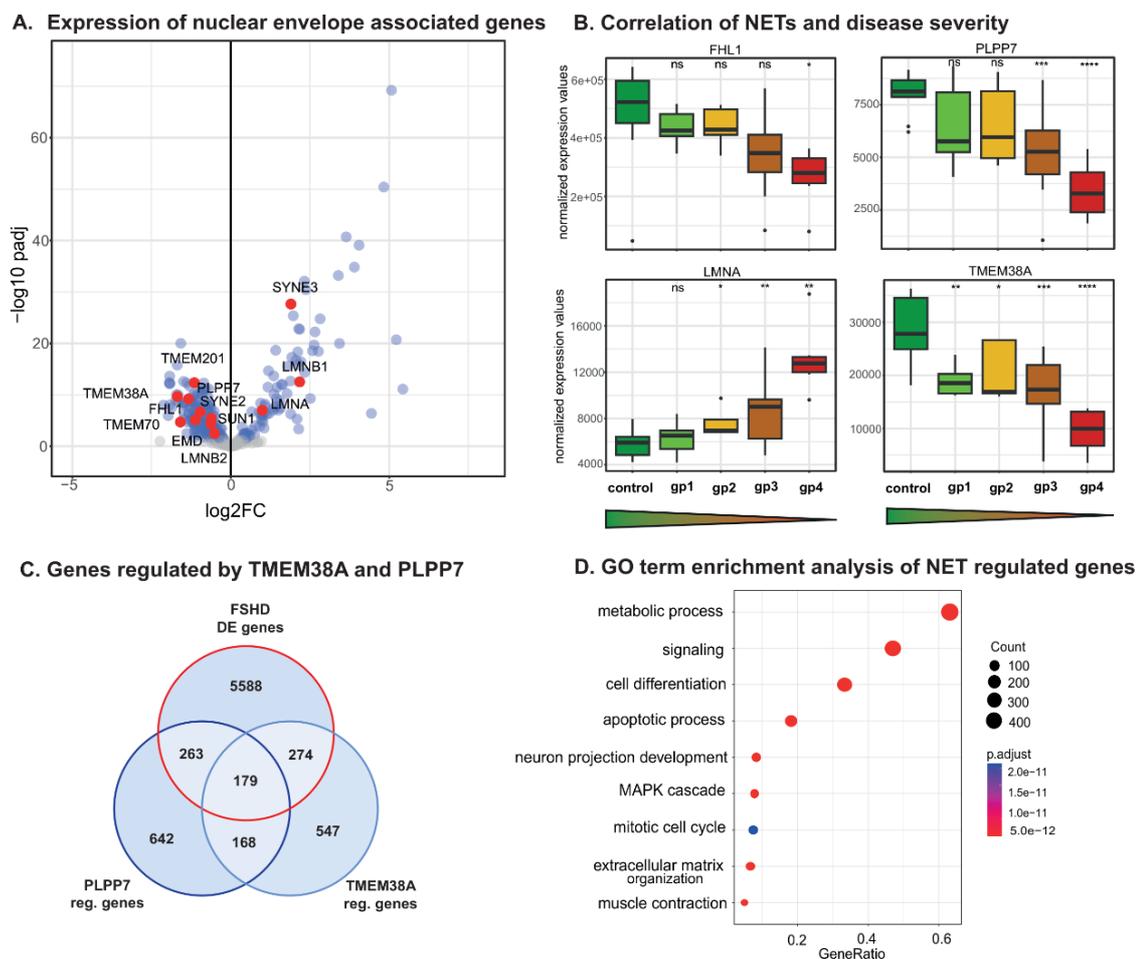


Figure 4: Genes encoding nuclear envelope proteins are misregulated in FSHD and contribute to the phenotype. A) Expression of nuclear envelope associated genes. Expression changes of nuclear envelope genes in strongly affected FSHD patients (group

4) compared to controls. No change in grey, change in blue ($\log_2FC > 0.5$), and muscle disease-related genes in red. The majority of nuclear envelope-associated genes is downregulated. **B) Correlation of NETs and disease severity.** Expression of four disease-related nuclear envelope-associated genes in the five groups (patients vs. controls) sorted after disease severity (controls in green, group 1 light green, group 2 orange, group 3 dark orange, group 4 red). Significance was measured using a t-test with controls as the reference group and is indicated by asterisks. FHL1, TMEM38A, and PLPP7 expression are inversely correlated to disease severity, while LMNA expression increases with severity. **C) Genes regulated by TMEM38A and PLPP7.** Differentially expressed genes in FSHD (group 4) were compared with genes regulated by TMEM38A and PLPP7, with 610 genes potentially under their control. **D) GO term enrichment analysis of NET regulated genes.** These 610 genes fall in GO terms relevant to the FSHD pathology.

FSHD patients. This leads to increased Exon skipping (ES) events in FSHD. In controls, many intron retentions (IR) events are found that are not present in FSHD patients. However, ~75% of these intron exclusion events in FSHD have actually skipped exons, which naturally include adjacent introns. Around 25% of introns are spliced out although retained in controls, while there are also intron retention events in FSHD that are not present in controls (pie chart). The amount of alternative splice site usage (3' Alt3, 5' Alt5) is slightly reduced in FSHD patients. **C) GO terms enriched in differentially spliced genes.** All genes with differentially used splicing events were used for a GO term enrichment analysis, showing that these genes contribute to the known FSHD phenotype. **D) Genes differentially spliced in FSHD.** Heatmap of selected genes with alternative splice site usage, colour indicates Ψ -values (percent spliced in), and all events have at least an absolute $\Delta\Psi$ -value of $|0.1|$ between controls and FSHD patients. **E) Exon skipping (ES) and intron retention (IR) events in selected genes.** ES and IR events are differentially used in FSHD patients. Reference exons are indicated in yellow and colored lines show the possible local splicing variations (LSVs). Bars show proportional usage of event usage in controls (left) and FSHD (right), while violin plots display Ψ -values of the respective LSV (middle).

Methods (online)

Meta-Analysis

The key advantage of our meta-analysis and the foundation of its statistical power is using the original omics data from the included studies instead of summarized data. All datasets, which were generated with different approaches and at different time points, could thus be standardized in terms of data retrieval from databases and uniform analysis procedures to provide optimal conditions for directly comparing significantly differentially expressed genes and molecular signalling pathways between FSHD patients and controls.

The meta-analysis follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement⁶⁵ and the guidelines described by Ramasamy et al. (2008)¹⁴. Timing and decision-making within the meta-analysis was divided into the work areas of *literature search, inclusion and exclusion criteria, data extraction, quality assessment, data preprocessing, statistical analysis, and data synthesis* (Supplementary Information S1, S2, S3 and Supplementary Table S1). In May 2022, the meta-analysis was registered in the International Prospective Register of Systematic

Reviews (PROSPERO; ID: CRD42022330489¹²). After the meta-analysis, all codes for preprocessing and reproducible analysis were published on GitHub (<https://github.com/FSHDresearch/Meta-Analysis-of-FSHD>) to make them traceable and available for further experiments. As shown in **Figure 1A**, our search identified a total of 11 studies that met the predefined criteria based on data from the databases^{62,66-75}, related publications and information obtained during our literature search through email contact with the corresponding authors (detailed description in *Supplementary Information S1* and *Supplementary Table S1*).

The meta-analysis encompasses data from five Affymetrix GeneChips⁷⁶, one Illumina BeadArray⁷⁷ and five Illumina RNA-Seq studies⁷⁸. For statistical analysis, the Benjamini-Hochberg correction was used to adjust for multiple testing⁷⁹ and adjusted (adj.) p-values of < 0.05 were considered significant⁸⁰. In total, the meta-analysis comprises data from 296 samples, including 167 FSHD samples and 129 controls (*Supplementary Information S3*). 13 datasets were generated from the 11 studies, as one dataset included both cell lines and biopsies (GSE56787⁷²). One dataset included two families (two sisters each as patient and control), which resulted in family-related batch effects in our statistical analysis due to strong genetic similarity in the families (GSE123468⁷³). The individual steps of pre-processing and statistical analysis are shown in **Figure 1B**. A random effects model was selected to identify significantly differentially expressed genes between patients and controls¹⁵. In this context, we used SMDH ("*standardized mean difference with heteroscedastic population variances in the two groups*") as an effect measure, as suggested by Bonett (2009)¹⁶. The model was chosen because, besides the biological heterogeneity between study participants, a relevant degree of technical heterogeneity could be assumed due to the use of different technologies (microarray vs. RNA-Seq; *Supplementary Information S5* and corresponding sensitivity analyses in *Supplementary Information S6*). The *DUX4* model found in our literature search was not included in the overall calculation due to its artificial nature, but was contrasted for comparison since *DUX4*-induced gene expression has been reported to be the major molecular signature in FSHD skeletal muscle⁶³. Without the *DUX4* dataset, we identified a total of 53113 unique Ensembl IDs (Ensembl database version 108 from December 2022). We decided to use only unique IDs measured in at least three datasets to get reliable results from the meta-analyses, which gave us 26858 unique IDs. We filtered our data to increase the power of the analysis^{81,82}, whereupon

13274 unique IDs remained. Using the *random effects* model, a p-value was assigned to each individual unique Ensemble ID. In this way, 13274 associated meta-analyses were theoretically performed for each dataset within the 288 samples (without *DUX4* model samples), ultimately leading to 1935 significant results (adj. p-value < 0.05). These results are decreasingly sorted by SMDH, as shown in the heatmap in **Figure 1C**, which is linked to data depicted in *Supplementary Table S2*.

To roughly divide the functional terms related to the genes into clusters, we used the Bioconductor *simplifyEnrichment* package to cluster and visualize the functional enrichment results³⁴. In this context, genes were divided into the three *Gene Ontology* (GO) domains *Cellular Component* (CC), *Molecular Function* (MF), and *Biological Process* (BP), each distinguishing between up- and down-regulated genes (*Supplementary Information S7-S9*). These clusters were used in combination with the gene list of our heatmap shown in **Figure 1C** (*Supplementary Table S2*), corresponding enrichment analyses with Bioconductor packages⁸³ (*Supplementary Table S3*) and the forest plots obtained with our shiny app meta-FSHD (**Figure 1D**) to analyse the results of our meta-analysis in terms of existing expertise and potential new findings in the context of FSHD.

Meta-FSHD App

The goal of *meta-FSHD* is to support future research in FSHD. For its implementation, we used the R packages shiny⁸⁴ and shinythemes⁸⁵ in combination with the R packages metafor⁸⁶, grid⁸⁷, forestploter⁸⁸ and ggplot2⁸⁹. The entire R code as well as the data for the app are publicly available on GitHub (<https://github.com/stamats/metaFSHD>). The app can easily be used by anyone without installing R and RStudio at the website hosted by Posit Software, PBC (<https://stamats.shinyapps.io/metafshd>). The corresponding background data for each study, including all samples, are provided in *Supplementary Information S3*.

In total, *meta-FSHD* considers all genes measured in at least 3 datasets, corresponding to 26858 unique ENSEMBLE-IDs, 21080 unique ENTREZ-IDs and 22791 unique gene names. In this regard, the user can choose between *Gene Name*, *ENSEMBL-ID*, and *ENTREZ-ID* to enter the corresponding gene of interest. This has the decisive advantage that, in addition to well-described genes, novel transcripts, e.g.

previously described only by their ENSEMBL-ID, can be investigated. (For *DUX4* and biomarker gene searches, additional information in *Supplementary Information S10*).

Once the gene of interest is entered, a forest plot appears, encompassing all 13 datasets. Although the *DUX4* model is not considered in the *random effects* calculation, it is presented in addition to the patient datasets for comparison purposes (*Meta-Analysis*). There is an additional option to save each forest plot in PDF format. The datasets are further divided into microarray and RNA-Seq datasets for clarity. The statistical calculation is based on the standardized log₂ fold change (STD log₂-FC) between FSHD and controls due to the different scales for microarrays and sequencing data, where STD log₂-FC corresponds to SMDH (standardized mean difference with heteroscedastic variances)¹⁶. The mean (SD) on the log₂-scale per group (FSHD or control) is given next to the name of each study, followed by a graphical representation incorporating the studies' impact (size of squares proportional to weight (inverse of standard error) of the single dataset within the meta-analysis) with the 95% confidence intervals (CI; horizontal lines) and the corresponding numbers on the right. If a gene is significantly upregulated in FSHD compared to control samples, the app displays it in red; if it is significantly downregulated, it is shown in blue. If a gene is insignificant, the CI-line crosses the vertical line of no effect (STD log₂-FC=0). The overall result of the meta-analysis is represented by the diamond (at the bottom). Furthermore, additional parameters such as the Cochran's Q test for heterogeneity⁹⁰ and I², a measure of heterogeneity classified in *Supplementary Information S5*.

Sub-analysis of an MRI-informed RNA-Seq dataset

Raw data from Wang et al. (2019)²⁰ was mapped to the human genome assembly GRCh38 (hg38) and sorted by coordinate using STAR 2.7.9a⁹¹ for analysis in DESeq2⁹² and MAJIQ⁵⁵.

A gene count matrix was generated using featureCounts⁹³ and standard DESeq2 workflow was followed, inbuilt lfcShrink function was used with apeglm⁹⁴. Patients were grouped according to the author's assessment: biomarker expression (*LEUTX*, *KHDCLI*, *TRIM43* and *PRAMEF2*), which correlated with pathology (section stainings and MRI). Thus, five groups were formed: controls and FSHD groups 1, 2, 3, and 4, with the latter

having the strongest phenotype and biomarker expression. All DESeq2 results are found in in *Supplementary Table S8*.

A comprehensive list of 386 genes that are associated with the nuclear envelope (either transmembrane proteins or interacting with them on the nucleoplasmic or cytoplasmic side) and preferentially expressed in muscle^{46,47} (*Supplementary Table S9*) was used to screen for nuclear envelope genes misregulated in FSHD. This was done for either group 4 or all groups and differences were evaluated by calculating p-values with *ggpmisc*⁹⁵. Genes regulated by TMEM38A and PLPP7 were extracted from Robson et al. (2016)⁵², and the venn diagram was generated with *ggVennDiagram*⁹⁶. GO term enrichment analysis (for NET regulated genes and mis-spliced genes) was conducted using *gprofiler2*⁹⁷ and visualized with *enrichplot*⁹⁸. Next to classic GO term analysis, gene set enrichment analysis (GSEA) was used for finding splicing-related pathways and their genes with the Bioconductor package *fgsea*⁹⁹, which was then visualized using *GOplot*¹⁰⁰. Splicing analysis was done with MAJIQ in python (v3.9) for group 4 compared to controls and visualized using Voila (<https://majiq.biociphers.org/>). All other plots were generated with *ggplot2*⁸⁹.

1. Deenen, J. C. W. *et al.* Population-based incidence and prevalence of facioscapulohumeral dystrophy From the Department of Neurology. *Neurology* **83**, 1056–1059 (2014).
2. Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* **49**, 925–934 (2017).
3. Das, S. & Chadwick, B. P. Influence of repressive histone and DNA methylation upon D4Z4 transcription in non-myogenic cells. *PLoS One* **11**, 1–26 (2016).
4. Tawil, R. *et al.* Evidence-based guideline summary: Evaluation, diagnosis, and management of facioscapulohumeral muscular dystrophy. *Neurology* **85**, 357–364 (2015).
5. Tassin, A. *et al.* DUX4 expression in FSHD muscle cells: How could such a rare protein cause a myopathy? *J Cell Mol Med* **17**, 76–89 (2013).
6. Tawil, R. & Van Der Maarel, S. M. Facioscapulohumeral muscular dystrophy. *Muscle Nerve* **34**, 1–15 (2006).
7. Ricci, E. *et al.* Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of KpnI repeats at the 4q35 locus and clinical phenotype. *Ann Neurol* **45**, 751–757 (1999).
8. Lunt, P. W. *et al.* Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). *Hum Mol Genet* **4**, 951–958 (1995).

9. Nikolic, A. *et al.* Clinical expression of facioscapulohumeral muscular dystrophy in carriers of 1-3 D4Z4 reduced alleles: Experience of the FSHD Italian National Registry. *BMJ Open* **6**, (2016).
10. Butz, M. *et al.* Facioscapulohumeral muscular dystrophy: Phenotype-genotype correlation in patients with borderline D4Z4 repeat numbers. *J Neurol* **250**, 932–937 (2003).
11. Statland, J. M. *et al.* Milder phenotype in facioscapulohumeral dystrophy with 7-10 residual D4Z4 repeats. *Neurology* **85**, 2147–2150 (2015).
12. Schätzl, T. *et al.* Meta-analysis of datasets in facioscapulohumeral muscular dystrophy using the original omics data for up to date comparability. Available from: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=330489 (2022).
13. Page, M. J. *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ* vol. 372 Available from: <https://doi.org/10.1136/bmj.n71> (2021).
14. Ramasamy, A., Mondry, A., Holmes, C. C. & Altman, D. G. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med* **5**, 1320–1332 (2008).
15. Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* **1**, 97–111 (2010).
16. Bonett, D. G. Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychol Methods* **14**, 225–238 (2009).
17. Kowaljew, V. *et al.* The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscular Disorders* **17**, 611–623 (2007).
18. Bosnakovski, D. *et al.* An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO Journal* **27**, 2766–2779 (2008).
19. Corona, E. D., Jacquelin, D., Gatica, L. & Rosa, A. L. Multiple Protein Domains Contribute to Nuclear Import and Cell Toxicity of DUX4, a Candidate Pathogenic Protein for Facioscapulohumeral Muscular Dystrophy. *PLoS One* **8**, (2013).
20. Wang, L. H. *et al.* MRI-informed muscle biopsies correlate MRI with pathology and DUX4 target gene expression in FSHD. *Hum Mol Genet* **28**, 476–486 (2019).
21. Choi, S. H. *et al.* DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res* **44**, 5161–5173 (2016).
22. Resnick, R. *et al.* DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. *Cell Rep* **29**, 1812-1820.e5 (2019).
23. Bosnakovski, D. *et al.* Low level DUX4 expression disrupts myogenesis through deregulation of myogenic gene expression. *Sci Rep* **8**, (2018).
24. Turki, A. *et al.* Functional muscle impairment in facioscapulohumeral muscular dystrophy is correlated with oxidative stress and mitochondrial dysfunction. *Free Radic Biol Med* **53**, 1068–1079 (2012).
25. Denny, A. P. & Heather, A. K. Are Antioxidants a Potential Therapy for FSHD? A Review of the Literature. *Oxid Med Cell Longev* **2017**, (2017).
26. Heher, P. *et al.* Interplay between mitochondrial reactive oxygen species, oxidative stress and hypoxic adaptation in facioscapulohumeral muscular dystrophy: Metabolic stress as potential therapeutic target. *Redox Biol* **51**, (2022).
27. Leary, S. C., Battersby, B. J., Hansford, R. G. & Moyes, C. D. Interactions between bioenergetics and mitochondrial biogenesis. *Biochim Biophys Acta* **1365**, 522–530 (1998).
28. Lek, A. *et al.* Applying genome-wide CRISPR-Cas9 screens for therapeutic discovery in facioscapulohumeral muscular dystrophy. *Sci Transl Med* **12**, 9–11 (2020).

29. Vercellino, I. & Sazanov, L. A. The assembly, regulation and function of the mitochondrial respiratory chain. *Nat Rev Mol Cell Biol* **23**, 141–161 (2022).
30. Masny, P. S. *et al.* Localization of 4q35.2 to the nuclear periphery: Is FSHD a nuclear envelope disease? *Hum Mol Genet* **13**, 1857–1871 (2004).
31. Gaillard, M. C. *et al.* Analysis of the 4q35 chromatin organization reveals distinct long-range interactions in patients affected with Facio-Scapulo-Humeral Dystrophy. *Sci Rep* **9**, (2019).
32. Mariot, V. *et al.* Correlation between low FAT1 expression and early affected muscle in facioscapulohumeral muscular dystrophy. *Ann Neurol* **78**, 387–400 (2015).
33. Maggi, L. *et al.* LMNA-associated myopathies: The Italian experience in a large cohort of patients. *Neurology* **83**, 1634–1644 (2014).
34. Gu, Z. & Hübschmann, D. Simplify enrichment: A bioconductor package for clustering and visualizing functional enrichment results. *Genomics Proteomics Bioinformatics* (2022) doi:<https://doi.org/10.1016/j.gpb.2022.04.008>.
35. Csapo, R., Gumpenberger, M. & Wessner, B. Skeletal Muscle Extracellular Matrix – What Do We Know About Its Composition, Regulation, and Physiological Roles? A Narrative Review. *Front Physiol* **11**, (2020).
36. Lieber, R. L. & Ward, S. R. Cellular Mechanisms of Tissue Fibrosis. 4. Structural and functional consequences of skeletal muscle fibrosis. *American Journal of Physiology-Cell Physiology* **305**, C241–C252 (2013).
37. Patton, B. L. Basal lamina and the organization of neuromuscular synapses. *J Neurocytol* **32**, 883–903 (2003).
38. Sanes, J. Laminin, Fibronectin, and Collagen in Synaptic and Extrasynaptic Portions of Muscle Fiber Basement Membrane. *J Cell Biol* **93**, 442–451 (1982).
39. Zhang, W., Liu, Y. & Zhang, H. Extracellular matrix: an important regulator of cell functions and skeletal muscle development. *Cell Biosci* **11**, (2021).
40. Tasca, G. *et al.* Different molecular signatures in magnetic resonance imaging-staged facioscapulohumeral muscular dystrophy muscles. *PLoS One* **7**, (2012).
41. Tasca, G. *et al.* Magnetic resonance imaging in a large cohort of facioscapulohumeral muscular dystrophy patients: Pattern refinement and implications for clinical trials. *Ann Neurol* **79**, 854–864 (2016).
42. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).
43. Wang, C. *et al.* The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol* **32**, 926–932 (2014).
44. de las Heras, J. I. *et al.* Tissue specificity in the nuclear envelope supports its functional complexity. *Nucleus* **4**, 460–477 (2013).
45. Östlund, C. *et al.* Reduction of a 4q35-encoded nuclear envelope protein in muscle differentiation. *Biochem Biophys Res Commun* **389**, 279–283 (2009).
46. Korfali, N. *et al.* The nuclear envelope proteome differs notably between tissues. *Nucleus* **3**, 552–564 (2012).
47. Wilkie, G. S. *et al.* Several novel nuclear envelope transmembrane proteins identified in skeletal muscle have cytoskeletal associations. *Mol Cell Proteomics* **10**, M110.003129 (2011).
48. Meinke, P. *et al.* A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism. *EBioMedicine* **51**, (2020).
49. Dong, C. H. *et al.* LMNB2 promotes the progression of colorectal cancer by silencing p21 expression. *Cell Death Dis* **12**, (2021).

50. Bonne, G. *et al.* Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nat Genet* **21**, 285–288 (1999).
51. Gueneau, L. *et al.* Mutations of the FHL1 Gene Cause Emery-Dreifuss Muscular Dystrophy. *The American Journal of Human Genetics* **85**, 338–353 (2009).
52. Robson, M. I. *et al.* Tissue-Specific Gene Repositioning by Muscle Nuclear Membrane Proteins Enhances Repression of Critical Developmental Genes during Myogenesis. *Mol Cell* **62**, 834–847 (2016).
53. Todorow, V. *et al.* Transcriptome analysis in a primary human muscle cell differentiation model for myotonic dystrophy type 1. *Int J Mol Sci* **22**, (2021).
54. de las Heras, J. I. *et al.* Metabolic, fibrotic and splicing pathways are all altered in Emery-Dreifuss muscular dystrophy spectrum patients to differing degrees. *Hum Mol Genet* **32**, 1010–1031 (2023).
55. Vaquero-Garcia, J. *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. (2016) doi:10.7554/eLife.11752.001.
56. Ghasemizadeh, A. *et al.* Macf1 controls skeletal muscle function through the microtubule-dependent localization of extra-synaptic myonuclei and mitochondria biogenesis. *Elife* **10**, (2021).
57. Randazzo, D. *et al.* Persistent upregulation of the β -tubulin tubb6, linked to muscle regeneration, is a source of microtubule disorganization in dystrophic muscle. *Hum Mol Genet* **28**, 1117–1135 (2019).
58. Salsia, V., Vattei, G. N. A. & Tupler, R. G. The FSHD jigsaw: are we placing the tiles in the right position? *Curr Opin Neurol* (2023) doi:10.1097/WCO.0000000000001176.
59. Lassche, S. *et al.* Reduced specific force in patients with mild and severe facioscapulohumeral muscular dystrophy. *Muscle Nerve* **63**, 60–67 (2021).
60. Erdmann, H. *et al.* Reply: An epigenetic basis for genetic anticipation in facioscapulohumeral muscular dystrophy type 1. *Brain* (2023) doi:10.1093/brain/awad216.
61. Buitrago, D. *et al.* Impact of DNA methylation on 3D genome structure. *Nat Commun* **12**, (2021).
62. FSHD-group Leiden University Medical Center (LUMC). Dataset. RNA-sequencing data from human FSHD and control skeletal muscle biopsies. <https://ega-archive.org/datasets/EGAD00001008337> (2022).
63. Yao, Z. *et al.* DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum Mol Genet* **23**, 5342–5352 (2014).
64. Tryggvason, K. & Patrakka, J. Chapter 4 - Alport's Disease and Thin Basement Membrane Nephropathy. in *Genetic Diseases of the Kidney* (eds. Lifton, R. P., Somlo, S., Giebisch, G. H. & Seldin, D. W.) 77–96 (Academic Press, 2009). doi:<https://doi.org/10.1016/B978-0-12-449851-8.00004-8>.
65. Moher, D. *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Revista Espanola de Nutricion Humana y Dietetica* **20**, 148–160 (2016).
66. Cheli, S. & Meneveri, R. Series GSE26061. Expression profiling of 4q-linked and phenotypic FSHD in different steps of myogenic differentiation. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26061> (2011).
67. Ehrlich, M. & Tsumagari, K. Series GSE26145. Expression profiling FSHD vs. control myoblasts and myotubes. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26145> (2011).
68. Welle, S. Series GSE10760. Effect of facioscapulohumeral dystrophy (FSHD) on skeletal muscle gene expression. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10760> (2008).

69. Kho, A., Arashiro, P., Kunkel, L. & Zatz, M. Series GSE15090. Gene expression profiles in muscle tissue from FSHD patients. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15090> (2009).
70. Rahimov, F. Series GSE36398. Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. *2012* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36398>.
71. Tasca, G., Pescatori, M., Cubeddu, T. & Ricci, E. Series GSE26852. Gene expression analysis of FSHD muscle with different MRI pattern. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26852> (2012).
72. Yao, Z. *et al.* Series GSE56787. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56787> (2014).
73. Banerji, C. & Zammit, P. Series GSE123468. RNA-seq of FSHD and control immortalised myoblasts II. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123468> (2018).
74. Freidman, S. *et al.* Series GSE115650. MRI-informed muscle biopsies correlate MRI with pathology and DUX4 target gene expression in FSHD. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115650> (2018).
75. Watt, K. *et al.* Series GSE138768. DUX4 promotes mitochondrial impairment in skeletal muscle. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138768> (2021).
76. Dalma-Weiszhausz, D. D., Warrington, J., Tanimoto, E. Y. & Miyada, C. G. DNA Microarrays, Part A: Array Platforms and Wet-Bench Protocols. [1] The Affymetrix GeneChip® Platform: An Overview. in *Methods in Enzymology* vol. 410 3–28 (Academic Press, 2006).
77. Fan, J. *et al.* DNA Microarrays, Part A: Array Platforms and Wet-Bench Protocols. [3] Illumina Universal Bead Arrays. in *Methods in Enzymology* vol. 410 57–73 (Academic Press, 2006).
78. Dillies, M.-A. *et al.* A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* **14**, 671–683 (2013).
79. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society series b-methodological* **57**, 289–300 (1995).
80. Krzywinski, M. & Altman, N. Points of significance: Significance, P values and t-tests. *Nat Methods* **10**, 1041–1042 (2013).
81. Bourgon, R., Gentleman, R. & Huber, W. Independent filtering increases detection power for high-throughput experiments. *PNAS* **107**, (2010).
82. Hackstadt, A. J. & Hess, A. M. Filtering for increased power for microarray data analysis. *BMC Bioinformatics* **10**, (2009).
83. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* **12**, 115–121 (2015).
84. Chang, W. *et al.* shiny: Web Application Framework for R. R package version 1.7.4. <https://CRAN.R-project.org/package=shiny> (2022).
85. Chang, W. shinythemes: Themes for Shiny. R package version 1.2.0. <https://CRAN.R-project.org/package=shinythemes> (2021).
86. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw* **36**, 1–48 (2010).
87. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available from: <https://www.R-project.org/> (2023).

88. Dayimu, A. forestploter: Create Flexible Forest Plot. R package version 1.1.0. <https://CRAN.R-project.org/package=forestploter> (2023).
89. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016). doi:<https://doi.org/10.1007/978-3-319-24277-4>.
90. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539–1558 (2002).
91. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
92. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
93. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
94. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).
95. Aphalo, P. ggpmisc: Miscellaneous Extensions to ‘ggplot2’. R package version 0.5.3. <https://CRAN.R-project.org/package=ggpmisc> (2023).
96. Gao, C. H., Yu, G. & Cai, P. ggVennDiagram: An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram. *Front Genet* **12**, (2021).
97. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res* **9**, 709 (2020).
98. Yu, G. enrichplot: Visualization of Functional Enrichment Result. R package version 1.20.0. <https://yulab-smu.top/biomedical-knowledge-mining-book/> (2023).
99. Korotkevich, G., Sukhov, V. & Sergushichev, A. A. Fast gene set enrichment analysis. *bioRxiv* (2019) doi:<https://doi.org/10.1101/060012>.
100. Walter, W., Sánchez-Cabo, F. & Ricote, M. GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912–2914 (2015).

5.5 Manuscript II

Splicing mediated by U2-associated Scaf6/CHERP is necessary for myogenesis in
Drosophila and vertebrates

Shao-Yen Kao, **Vanessa Todorow**, Pablo Monteagudo Mesas, Linus Hoelzel, Paul Walper, Keshika Ravichandran, Oliver Weinert, Alexander Grauer, Israa Alqassem, Luka Borozan, Mara Barone, Andrea David Re Cecconi, Ignasi Forne, Michael Menden, Tobias Straub, Rosanna Piccirillo, Peter Meinke, Benedikt Schoser, Rippei Hayashi, Stefan Canzar, Maria L. Spletter

Manuscript in preparation

Contributions

In the following manuscript, I conducted the experiments in mouse and human cell culture and analyzed and interpreted their results. Methods included Western Blotting, qPCR, silver staining, immunoprecipitation and immunofluorescence, siRNA mediated knockdown as well as sample preparation for mass spectrometry and RNAseq. I further analyzed the mass spec and RNAseq data and designed figures 7 and 8 and described them in the respective sections. Please notice that manuscript II is not yet complete nor ready for submission as the discussion and references are missing, however, the experimental part is complete.

Splicing mediated by U2-associated Scaf6/CHERP is necessary for myogenesis in *Drosophila* and vertebrates

Shao-Yen Kao¹, Vanessa Todorow², Pablo Monteagudo Mesas³, Linus Hoelzel⁴, Paul Walper¹, Keshika Ravichandran¹, Oliver Weinert¹, Alexander Grauer¹, Israa Alqassem⁴, Luka Borozan⁴, Mara Barone⁵, Andrea David Re Cecconi⁵, Ignasi Forne⁶, Michael Menden³, Tobias Straub⁷, Rosanna Piccirillo⁵, Peter Meinke², Benedikt Schoser², Rippei Hayashi⁸, Stefan Canzar⁴, Maria L. Spletter^{1,9,#}

1. Biomedical Center, Department of Physiological Chemistry, Ludwig-Maximilians-Universität München, Großhaderner Str. 9, 82152 Martinsried-Planegg Germany

2. Friedrich-Baur-Institute at the Department of Neurology, University Hospital, Ludwig-Maximilians-Universität München, Munich, Germany

3. Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany

4. Institute of Computational Biology, Helmholtz Zentrum München – German Research Center for Environmental Health, 85764 Munich, Germany

5. Unit of Neuromuscular Dysfunctions, Neuroscience Department, Istituto di Ricerche Farmacologiche “Mario Negri”, via Mario Negri, 2 20156 Milan, Italy

6. Biomedical Center, Zentrallabor für Proteinanalytik, Ludwig-Maximilians-Universität München, Großhaderner Str. 9, 82152 Martinsried-Planegg Germany

7. Biomedical Center, Bioinformatics Core Facility, Ludwig-Maximilians-Universität München, Großhaderner Str. 9, 82152 Martinsried-Planegg Germany

8. The John Curtin School of Medical Research, Australian National University College of Health and Medicine, Canberra, Australia

9. School of Science and Engineering, Division of Biological and Biomedical Systems, 5009 Rockhill Road, Kansas City, MO, 64110 USA

#Corresponding author:

Dr. Maria L. Spletter, School of Science and Engineering, Division of Biological and Biomedical Systems, 5009 Rockhill Road, Kansas City, MO, 64110 USA.

E-mail address: maria.spletter@umkc.edu, Telephone: +1-816-235-6379, Fax: +1-816-235-6911

Key words: Scaf6, CHERP, *Drosophila*, myogenesis, flight muscle, alternative splicing

Running title: Scaf6/CHERP regulates splicing in muscle and neurons

Abstract

Tissue-specific patterns of alternative splicing confer muscle fibers with distinct morphological and contractile properties. Mis-regulation of splicing alters the balance of gene isoform expression, impacting steps in muscle development from myoblast division and differentiation to sarcomere assembly and contractile function. Altered isoform expression dynamics further contribute to muscle aging, atrophy and disease. RNA binding proteins regulate RNA processing, but the majority of these proteins have never been studied in muscle. Here we identify spliceosome associated factor Scaf6/CHERP to play an essential and conserved role in muscle development. Whole animal Scaf6 mutant flies display myofibril defects and are flightless due to detachment and atrophy of the flight muscles. mRNA-seq data demonstrate that Scaf6 regulates alternative splicing and suppresses cryptic splicing, and we show that these molecular defects lead to the loss of select sarcomere proteins, notably Tm1, Sallimus, Unc-89, Zasp-52 and Zasp-66. The altered balance in sarcomere protein expression induces aberrant myosin contractility and myofiber loss. Interestingly, Scaf6 is required cell-intrinsically for splicing in both muscle and neurons, and neuronal-specific loss of Scaf6 results in behavioral defects and decreased motor neuron axon branching. We further demonstrate that Scaf6 has a conserved function in regulating myoblast proliferation, and knockdown of the vertebrate homologue CHERP in mouse C2C12 cells results in decreased proliferation and premature differentiation. From pull-down mass-spectrometry experiments in mouse and human muscle cells, we demonstrate that CHERP is tightly associated with the Sf3b1 complex, which is a core unit of the U2 spliceosomal complex. We also find that CHERP is misexpressed in myotonic dystrophy patient cells and a cancer cachexia model. Our results thus establish a novel, disease-relevant function for Scaf6/CHERP in myogenesis in both flies and vertebrates, and provide insight into how the physiological roles of splicing factors are defined by temporal and tissue-specific requirements.

Introduction

A wide diversity in muscle form and function is observed throughout the animal kingdom. The functional and morphological divergence among different muscle types arises during

development (Spletter and Schnorrer, 2014). The molecular basis of developmental heterogeneity among distinct muscle groups lies in differences in both transcription and alternative splicing. Alternative splicing is a regulatory mechanism that generates more than one mRNA from a single gene (Kim et al., 2014). Alternative mRNAs can differ for example in their coding sequences, level of translation or subcellular localization. Resulting protein isoforms can have varied physiological functions and stability (Nilsen and Graveley, 2010), for example driving cell differentiation or leading to the acquisition of tissue identity and muscle-type specific properties (Wang et al., 2008). Studies suggest that 90-95% of genes in the human body undergo some form of alternative splicing (Pan et al., 2008), and muscle in particular has high levels of alternative splicing. Misregulation of alternative splicing leads to muscle and heart diseases such as myotonic dystrophy and dilated cardiomyopathy, and is associated with loss of muscle mass and the normal ageing process. It is thus critical that muscle cells regulate alternative splicing patterns in a spatially and temporally restricted manner (Nikonova et al., 2020). Therefore, further insight into RNA processing might provide a path towards the development of therapeutic approaches for muscle disorders.

Splicing is carried out by an extensively regulated RNA-protein complex called the spliceosome (Wahl et al., 2009). In association with a number of regulatory RNA binding proteins, this enzyme complex processes pre-mRNA transcripts to produce different mRNA isoforms. RNA-binding proteins (RBPs) are the *trans*-acting factors that recognise and bind to distinct sequences across the transcriptome (Baralle and Giudice, 2017). Multiple RBPs regulate alternative splicing cooperatively or competitively by either promoting or inhibiting the recognition of splice sites leading to varied splice products (Fu and Ares, 2014). However, our current understanding is informed by only 4% of total RBPs (Nikonova et al., 2019), leaving fundamental questions open about the regulatory mechanism of alternative splicing as well as physiological and functional relevance of splice isoforms. Further work is therefore required for the identification of detailed molecular mechanisms governing splice site selection and the functional impact of splice events on cell fate decisions.

Scaf6 is the fly homolog of Calcium Homeostasis Endoplasmic Reticulum Protein (CHERP), a conserved SR-like protein with a basic domain structure consisting of a SWAP, RS, RPR and G-patch domain (Laplante et al., 2000). The primary characterized

role of CHERP in vertebrate cells is to regulate cell proliferation and viability, and recent studies in vertebrate Hek293T and cancer cell lines demonstrated that CHERP is a nuclear protein involved in the regulation of alternative splicing (Crisci et al., 2015; Sasaki-Osugi et al., 2013). As a splicing factor, CHERP is associated with the U2 complex of the spliceosome and co-localizes with SC35 in nuclear speckles (De Maio et al., 2018; Saitoh et al., 2004). Together with splicing factors like SF1, Rbm17 and U2SURP, CHERP is responsible for suppressing cryptic splicing and regulating alternative splice events (De Maio et al., 2018; Tan et al., 2016; Crisci et al., 2015). CHERP has also been shown to interact with Apoptosis-linked Gene 2 (ALG-2) to regulate alternative splicing of IP3R1 pre-mRNA (Sasaki-Osugi et al., 2013).

The animal kingdom, including species as diverse as flies and humans, possesses numerous distinct muscles that exhibit a wide range of morphological and contractile properties (Hill and Olsen, 2012). One major myth in the field of developmental biology is how this diversity arises, particularly regarding which cytoskeletal adaptations underlie the broad range of contractile abilities, which is composed of sarcomere as the basic unit. Comprising over one hundred proteins, sarcomeres link end-to-end to form long myofibrils that span the length of a muscle (Nakka et al., 2018). Even slight changes of modifications or isoforms in the structure and proportions of these sarcomeric proteins, via changes in gene expression, alternative splicing, or regulatory dynamics of messenger RNA (mRNA), can rise a huge impact on the property of muscle contraction, and thereby serve as a mechanism for fine-tuning contractile properties (Armstrong & Phelps, 1984; Bottinelli & Reggiani, 2000; Bottinelli, 2001; Schiaffino & Reggiani, 2011; Schiaffino et al, 2020). Therefore, comprehending the normal dynamics of RNA regulation during muscle development is critical for understanding both normal muscle physiology and disease.

RNA regulation involves multi-steps, such as alternative splicing and post-transcriptional modification. Within the DNA genome, protein-coding regions are interspersed with non-coding sequences called introns, which need to be removed for the production of functional mRNA. This process is known as RNA splicing. Splice sites at the boundaries of exons, typically marked by conserved nucleotide sequences GU and AG, are essential for splicing. However, alternative splice sites such as GT-AG, GC-AG or AT-AC can also be used within introns (Brackenridge, Wilkie, and Sreaton 2003; Burge et al. 1999; Pollard et al. 2002; Quan and Forte 1990; Ruskin and Green 1985;

Sheth et al. 2006; Szafranski et al. 2007; Twigg et al. 1998; Will and Lührmann 2011). The intron also contains a branch point sequence located 18-40 nucleotides upstream of the 3' splice site (Ruskin and Green 1985), which is necessary for spliceosomal recognition. In higher eukaryotes, a polypyrimidine tract (PPT) can be found downstream of the branch point. The splicing process consists of two transesterification reactions (Moore, Query, and Sharp 1993). In the first step, the 2'OH group at the branch point attacks the 5' splice site, resulting in the formation of a lariat structure. In the second step, the free 3'-OH upstream exon attacks the 3' splice site, joining the two exons and removing the intron as the final step of RNA splicing to become mature mRNA (Ooi et al. 2001).

To achieve of RNA splicing, it requires the integration of a large number of multi-subunit spliceosome complex and RNA binding proteins (RBPs) to achieve proper RNA regulation. RBPs play a vital role in regulating alternative splicing by generating splicing profiles in different cell types, and in controlling translation levels by binding to 3'-UTR elements and associating with other factors (Hentze et al. 2018; Ho et al. 2021). This makes RBPs crucial for eukaryotic genome information during development, as they are responsible for establishing, refining, and maintaining specific properties of tissues and fiber types, especially in the muscle system (Lunde BM et al. 2007; Kao et al. 2021; Shi and Grifone 2021). Disruptions in alternative splicing and protein isoform expression patterns are commonly observed in aging and diseases such as cancer and cardiomyopathy (Van Pelt 2019). Loss of RBP function can lead to severe neuromuscular disorders such as myotonic dystrophy (Klinck et al. 2014; André et al. 2018), amyotrophic lateral sclerosis (Wood et al. 2021), and spinal motor atrophy. However, despite of the feat that a substantial amount of RBPs have been identified expressed in muscle across species, only around 4% of RBPs are well-studied (Nikonova et al. 2019). We still only have limited knowledge to understand how RBPs function in muscle development. Therefore, a detailed understanding of RBP function is crucial.

CHERP (Calcium Homeostasis Endoplasmic Reticulum Protein), also known as SCAF6 (SR-related CTD-associated factor 6) in *Drosophila*, is an RNA-binding protein that has been shown to play a role in alternative splicing regulation. CHERP is a highly evolutionarily conserved protein that encodes SWAP/SURP, CID and G-patch domain (Lin-Moshier et al. 2013) across the species. The SWAP/SURP domain is a NH2 terminal domain (Kuwasako et al. 2006), which is reported to interact with spliceosome factors such as SF1. Proteins containing SWAP/SURP domains are predicted to regulate pre-mRNA

splicing, for example like splicing proteins such as members of the Prp21 family that interact with the U2 snRNP (Spikes et al. 1994). The CID domain is involved in pre-mRNA processing and contains a region shown to interact with the RNA polymerase II (RNA Pol II) C-terminal domain (Meinhart and Cramer 2004; Noble et al. 2005).

Originally, CHERP was thought to be an Endoplasmic reticulum (ER) membrane protein that regulates intracellular Ca^{2+} release (Ryan et al. 2011). However, subsequent studies showed that CHERP is localized in the nucleus and interacts with spliceosomal machinery and PolIII to regulate mRNA alternative splicing, particularly of calcium regulatory machinery in a Ca^{2+} -dependent manner (Lin-Moshier et al. 2013). CHERP is identified as one of accessory components of the U2 complex in both vertebrates and flies, and forms a subcomplex with RBM17/SPF45 and U2SURP/SR140 (De Maio et al. 2018). These proteins demonstrate reciprocal regulation of complex and expression stability, where loss of any of the three proteins leads to down-regulation of the other two. Knockdown of CHERP induces defects in alternative splicing, notably increased rates of cryptic splicing, which affects the cell proliferation program through alternative splicing. The role of CHERP in splicing was further confirmed in a study in HeLa cells, which showed that these three components form a tightly-associated complex in both a physical and functional perspectives (Cvitkovic and Jurica 2013; Yamanaka et al. 2022). Depletion of CHERP leads to intron retention in target mRNA and accumulated poly(A) RNAs in U2OS nucleus (Yamanaka et al. 2022), which results in splicing dynamic changes and delays in the M phase progression. Thus, current studies support the functional role of Scaf6/CHERP in primary/direct mRNA-splicing, mRNA metabolism regulation associated with U2 spliceosome, and potential indirect regulation of intracellular Ca^{2+} homeostasis. In addition, important insights into CHERP's function in regulating cell proliferation have been identified in various cancer cell models (Zhang et al. 2017, Takayama et al. 2017; Wang et al. 2019)). However, the understanding of CHERP *in vivo* animal systems is still limited.

Given the high degree of conservation in CHERP between flies and vertebrates, the *Drosophila* ortholog is also presumed to play a role in splicing. Indeed, high throughput studies indicate that SCAF6 is also an accessory component of the U2 complex in flies and binds to the tail of PolIII (Spliceosome Database, <http://spliceosomedb.ucsc.edu/proteins/12989>). Based on experimental protein interactions available in the String database, SCAF6 interacts with U2-associated protein SR140 (CG9346), Sf3b2 (CG3605), and Sf3b3 (CG13900). To date, there is a single

study hinting at Scaf6 function *in vivo*. In fly embryos, an EMS-screen identified a mutation of *scaf6* that disrupts the trafficking of *gurken* mRNA leading to mislocation, as well as impairs the transposon piRNA silencing (Hayashi et al. 2014). This suggests that Scaf6 might play additional roles in RNA regulation, although could reflect an indirect function through the regulation of a yet-unidentified intermediate target. Thus, the potential functional and regulatory role of SCAF6 in an *in vivo* muscle system remains unknown.

In this paper, we investigate the functional and molecular mechanism of SCAF6 *in vivo* *Drosophila* system and CHERP in mammalian cell line system. We demonstrate that SCAF6 plays a critical role during fly muscle development (IFM). When we scope into the molecular regulatory level, the disruption of SCAF6 causes severe misregulation of mRNA targets, including fundamental sarcomeric genes, and leads to intron-retention, indicating its importance of recognition of the splice site. In addition, we also found the potential role of myoblast proliferation regulation of SCAF6 during development, which indicates the conservational function between the muscle system and the cancer models.

Here we identify a novel, physiological function for Scaf6/CHERP in the development of muscle and neurons. Scaf6 mutant flies are flightless phenotype, as well as a multitude of other behavioral deficits. IFM myofibers detach early in development and have problems in climbing due to defects in muscle growth and development (Figure 1A,B). Scaf6 mutants show muscle and neuronal specific defects, and mRNA-Seq of flight muscle reveals defects in alternative splicing including intron retention and increased use of cryptic splice sites. These splicing defects, for example in key structural components of the sarcomere such as Zasp66 or Zasp52 (Figure 1C-G), lead to near complete loss of sarcomere proteins and likely underlie the observed defects in muscle and sarcomere development.

Results

We identified *scaf6* as one of hundreds of RNA-binding proteins that is expressed in mRNA-Seq data from *Drosophila* muscle (Kao et al., 2021), but which has not been previously characterized *in vivo* or in muscle. *scaf6* encodes a conserved RS-domain protein with SWAP/U2-Surp, CID and G-patch domains (Fig. 1 A, Fig. S1 A), which suggest a splicing-associated function. Although *scaf6* is expressed ubiquitously, *scaf6* mRNA expression significantly decreases across indirect flight muscle (IFM)

development, with the highest expression levels coinciding with early steps in myogenesis including myoblast migration, fusion and myofibrillogenesis (Fig. 1 B, Fig. S1 A, B). *scaf6* encodes three mRNA isoforms, and the full-length isoform *scaf6-RA* is predominantly expressed in IFM, leg and brain (Fig. 1 A, Fig. S1 I). Increased expression of *scaf6-RC* in brain and *scaf6-RE* in developing IFM likely reflect spatial and temporal regulatory dynamics (Fig. S1 I). We confirmed that Scaf6-PA protein is expressed and nuclear localized in IFM and neurons using an endogenous, C-terminal FLAG tag line (Fig. S1 C-E). This nuclear localization is maintained even with expression of UAS-Myc-Scaf6-RA with either the muscle-specific driver Mef2-Gal4 or the neuronal-specific driver Elav-Gal4 (Fig. 1 C-D, Fig. S1 F-H). The expression dynamics and subcellular localization of Scaf6 are thus consistent with a function in pupal myogenesis, which we test below.

Scaf6 is necessary for muscle development

To evaluate a possible function for Scaf6 in myogenesis, we first characterized two mutant alleles, *scaf6*^{12M9} and *scaf6*^{M2M1}. *scaf6*^{12M9} is an EMS allele that introduces a stop codon at position 560, resulting in early truncation of the Scaf6-PA protein before the G-patch domain (Hayashi et al., 2014). This allele contains a linked background mutation (Hayashi et al., 2014) and is pupal lethal when homozygous as well as over *Df(3L)ED4674*, a deficiency which covers the *scaf6* locus (Fig. S1 M). We therefore generated a new CRISPR deletion allele, *scaf6*^{M2M1}, that results in a frame-shift in the N-terminal region of the protein just after the SWAP domain (Fig. 1 A). Although this *scaf6*^{M2M1} allele is late pupal lethal when homozygous or combined with *Df(3L)ED4674*, transheterozygous *scaf6*^{M2M1/12M9} mutants (referred to as *scaf6*^{-/-}) are pharate lethal (Fig. S1 M) and produce adult escapers that survive 1-2 days, allowing us to assay adult phenotypes. On the RNA level, expression of the full-length *scaf6-RA* isoform is significantly decreased (Fig. S1 J, K, L). We additionally generated a short-hairpin RNAi (*scaf6-IR*) line that results in a significant decrease in *scaf6-RA* expression (Fig. 1 A, Fig. S1 K, L). These data indicate that our various genetic tools produce strong Scaf6 hypomorph conditions, either through a decrease in *scaf6-RA* expression or through production of a protein lacking the C-terminal G-patch domain.

Pupal and pharate lethality, such as we observed in *scaf6* mutant flies, can be caused by defects in muscle structure or function; thus, we next tested if Scaf6 regulates adult muscle development. To assay muscle function, we performed a flight test and found

that surviving adult *scaf6*^{-/-} flies are flightless, as are flies with muscle-specific *scaf6-IR* driven by Mef2-Gal4 (Fig. 1 I). When we examined the IFMs in thorax hemisections from 1 d adult flies, we observed a strong detachment defect in *scaf6*^{-/-} as compared to *w*¹¹¹⁸ control myofibers (Fig. 1 E, F, J). Myofibril structure in *scaf6*^{-/-} IFM was severely disrupted, and is characterized by dense actin accumulations, a loss of sarcomere structure, and frayed myofibrils (Fig. 1 F). *scaf6-IR* IFM showed similar defects including myofiber detachment, myofibril tearing and fraying, and indistinct sarcomeres (Fig. 1 G, H, J, Fig. S1 Q). This data shows that Scaf6 is necessary for IFM development or maintenance, and demonstrates that the observed phenotype is specifically caused by loss of Scaf6.

To determine when these defects arise during muscle development, we evaluated IFM morphology at 48, 72 and 88 hours (h) after puparium formation (APF). In *scaf6*^{-/-} flies, IFM myofibers are already detached at 48 h APF (Fig. 1 F). Although sarcomere architecture is still present at 48 h APF, *scaf6*^{-/-} sarcomeres are significantly shorter than wildtype (Fig. 1 K) and myofibrils are significantly thicker, display tearing and splitting, and contain prominent actin bodies at the Z-discs (zebra bodies) (Fig. 1 E, F, L). Sarcomere and myofibril structure progressively degenerate, leading to a loss of sarcomere architecture by 90 h APF (Fig. 1 E, F). We observed a similar phenotype in all mutant combinations we were able to examine, including *scaf6*^{M2M1/Df(3L)ED4674}, *scaf6*^{I2M9/Df(3L)ED4674}, and *scaf6*^{M2M1/M2M1} (Fig. S1 N, O, P), as well as in *scaf6-IR* IFM (Fig. 1 G, H, Fig. S1 Q). Leg muscle was structurally intact, indicating that fibrillar IFMs are more strongly affected than tubular muscles (Fig. S1 O, P, Q). Taken together, these data show that the Scaf6 phenotype is pronounced in fibrillar muscles and arises at an early stage of muscle development.

Gene expression is altered at early stages of myogenesis in *scaf6*^{-/-} IFM

To identify the molecular defects that contribute to the behavioral and cellular phenotypes we observed in *scaf6*^{-/-} animals, we performed mRNA-Seq on dissected IFM at 30 h and 72 h APF. Using DESeq2 to quantify differential expression (DE) between *scaf6*^{-/-} and *w*¹¹¹⁸ IFM, we observed significant changes in gene expression at both timepoints (Fig. S2 A, B, Supplemental Table X). We noted that more genes are differentially expressed at 30 h than at 72 h APF (Fig. 2 A), and although gene expression changes are moderately correlated between the two timepoints (Pearson's correlation coefficient = 0.47) (Fig. 2 B), there is little overlap between significantly DE genes (adjusted p-value ≤ 0.05) at 30 h and 72 h (Fig. 2 C). To understand which biological

processes are affected by changes in gene expression, we performed gene ontology (GO) term enrichments. At 30 h APF, significantly upregulated genes are enriched for terms such as “signal transduction,” “myoblast fusion,” “muscle structure development” and “axon guidance,” and significantly downregulated genes are notably enriched in mitochondrial-related GO terms (Fig. 2 E). At 72 h APF, significantly upregulated genes are enriched for terms such as “actin-mediated cell contraction,” “stabilization of membrane potential” and “regulation of cell shape” (Fig. 2 E, Supplemental Table X). In a complementary approach, we examined categories of genes important for muscle function, and found that while mitochondrial-related genes are significantly downregulated at 30 h but not 72 h APF, synapse genes are preferentially upregulated at both timepoints, and genes that have an RNAi phenotype in muscle display a bimodal regulatory distribution (Fig. S2 C). We wondered if this pattern might indicate a delay in early myogenesis, so we then examined how genes that are temporally regulated during IFM development are affected in *scaf6*^{-/-} muscle. Genes that are normally upregulated from 24-30 h APF, for example *Act88F*, *fln* and *TpnC4*, are downregulated in *scaf6*^{-/-} IFM at 30 h APF (Fig. 2 D). By contrast, myoblast fusion-associated genes such as *lmd*, *mbc* and *sns*, which are normally downregulated from 24-30 h APF, are upregulated in *scaf6*^{-/-} samples at 30 h APF (Fig. 2 D). We did not observe this trend in 72 h *scaf6*^{-/-} IFM, nor with genes that are temporally regulated from 30-72 h APF (Fig. S2 D). We also did not detect strong impairment in the temporal switch in gene expression from 30-72 h in *scaf6*^{-/-} IFM (Pearson’s correlation coefficient = 0.92) (Fig. S2 E). We interpret these findings to reflect a delay in IFM development at 30 h APF, possibly due to myoblast-related defects which we characterize below, that is resolved by 72 h APF. Taken together, our gene-expression level analysis points to several developmental mechanisms that contribute to the *scaf6*^{-/-} phenotype, including defects in myoblast fusion, cytoskeletal organization, contractility and neuronal development.

Scaf6 regulates exon use and alternative splicing of muscle genes

Scaf6 as well as its vertebrate homologue CHERP are reported to be accessory components of the U2 spliceosome complex, thus we next evaluated if defects in alternative splicing are present in *scaf6*^{-/-} IFM. We used DEXSeq to identify significant differences in exon use in our mRNA-Seq data, which reflects changes in alternative splicing (Anders et al., 2012a). We identified significantly DE exons (p-value \leq 0.05) at both 30 h and 72 h APF (Fig. 2 F, Fig. S2 G, H, Supplemental Table X). At both timepoints,

genes with significantly DE exons are enriched for cytoskeletal terms such as “cytoskeletal organization” and “actomyosin structure organization” (Fig. 2 E). At 72 h APF, genes with DE exons are also enriched for terms such as “calcium ion transport,” “sarcomere organization,” “muscle contraction,” “synapse organization,” and “neuromuscular junction development” (Fig. 2 G). Strikingly, although few sarcomere proteins (SPs) are misregulated on the gene level (Fig. 2 B, Fig. S2 A, B, C), a large number of SPs have DE exons (Fig. 2 G, Fig. S2 G, H, I). We noted that many more exons are misregulated at 72 h than at 30 h APF (Fig. 2 F, Fig. S2 I), the opposite trend to what we observed with our gene level analysis (Fig. 2 A, Fig. S2 C), and changes in exon use at the two timepoints are weakly but positively correlated (Pearson’s coefficient = 0.32) (Fig. 2 G). Interestingly, there is little overlap between significantly DE genes as detected by DESeq2 and genes with significantly DE exons at either 30 h or 72 h APF (Fig. S2 F), and there is little overlap in the DE exons between 30 h and 72 h APF (Fig. 2 C), indicating that genes that are misspliced are largely distinct at early and late stages of muscle development. This data demonstrates that loss of *Scaf6* leads to splicing defects, notably in cytoskeletal, mitochondrial and synaptic genes. Further, while early stages of the *scaf6*^{-/-} phenotype during muscle differentiation are characterized by changes in gene expression, later phases during muscle growth and maturation are characterized predominantly by splicing defects and changes in isoform expression.

We next evaluated the presence of novel splicing events in our mRNA-Seq data. We used fortuna to identify novel alternative splice (AS) events and a separate approach to identify intron retention (IR) events, as our analysis with DEXSeq was limited to annotated exons. We observed a dramatic increase in novel AS events as well as IR events in *scaf6*^{-/-} IFM at both 30 h and 72 h APF (Fig. 2 H, Fig. S2 J, K). If we consider all detected events unique to either *scaf6*^{-/-} or *w¹¹¹⁸* samples, 73-75% of *scaf6* and 87-88% of wild-type novel events are supported by only 1 or 2 reads (Fig. S2 J), likely reflecting noise in the splicing process. For further analysis we therefore considered only novel events with 5 or more supporting reads. The increase in novel AS events in *scaf6*^{-/-} samples is consistent across individual categories of genes important for muscle development and function, including mitochondrial, sarcomere and synaptic genes, as well genes with an RNAi phenotype in muscle (Fig. S2 L). Genes containing novel AS and IR events at 30 h APF are enriched for GO terms such as “cytoskeleton organization” and “muscle structure development,” and at both 30 h and 72 h genes with novel events are enriched for “axon guidance,” “regulation of mRNA processing,” “signal transduction,” and “synapse part”

(Fig. 2 G). The novel events identified by fortuna in *scaf6*^{-/-} but not in *w¹¹¹⁸* IFM, in total 4105 events at 30 h APF and 3814 events at 72 h APF, were predominantly exon skip (ES) events (72-78%), but also included use of alternative acceptors (AA, 6-8%), alternative donors (AD, 4-5%), and intron in exon events (IE, 3-5%) (Fig. 2 I). Taken together, our analysis demonstrates that loss of Scaf6 leads to alternative splicing defects accompanied by an increase in intron retention and novel AS events at both 30 h and 72h APF, notably in cytoskeletal, mitochondrial and synaptic genes.

To verify the novel AS events and changes in exon use in our mRNA-Seq data, we performed RT-PCR on IFMs dissected at 72 h APF from *w¹¹¹⁸* or *scaf6*^{-/-} flies. We were able to confirm three types of aberrant events, including intron retentions, novel junction use, and altered use of alternative exons. For example, in the z-disc protein *Zasp66*, we confirmed a switch in splicing towards inclusion of alternative cassette exon 10, as well as use of an alternative acceptor for exon 10 and an intron retention event (Fig. 2 J). In the M-line protein *Unc-89*, also known as *obscurin*, we confirmed increased use of exon 22 in IFM (Fig. 2 M). Interestingly, exon 10 in *Zasp66* and exon 22 in *Unc-89*, which are aberrantly spliced in *scaf6*^{-/-} IFM, are normally used more often in brain samples from wildtype flies, suggesting that loss of *scaf6* increases the use of splice events typically suppressed in flight muscle. In *Tropomyosin 1 (Tm1)*, we confirmed increased use of an alternative termination as well as intron retention events (Fig. 2 K). Finally, in the titin-like protein *sallimus (sls)*, we confirmed use of a cryptic exon between exons 17 and 18 (Fig. 2 L). The intron retention event in *Zasp66* as well as the cryptic exon in *sls* were observed prominently in IFM but only weakly in brain, indicating that splicing may be differentially affected in the two tissues. Additionally, we noted that in all cases, we could detect aberrant events together with the expected splice event. This suggests a lack of fidelity in the splicing process, but may also reflect a decreased stability of aberrantly spliced transcripts. Our RT-PCR data thus validate our mRNA-Seq analysis and confirm multiple aberrant splicing events in *scaf6*^{-/-} IFM.

Sarcomere protein expression is altered after loss of Scaf6

Since changes in splicing and mRNA expression levels do not always correlate with altered protein expression (Liu et al., 2016; Salovska et al., 2020), we performed whole proteome mass spectrometry to evaluate if the widespread splicing defects in *scaf6*^{-/-} IFM impact protein expression. We compared dissected IFM from *scaf6*^{-/-} mutant and wildtype *w¹¹¹⁸* flies at 72 h APF, when mutant myofibers are detached, but myofibril

structure is still intact. We identified significant changes in expression of a large number of proteins (1,110 proteins, FDR < 0.05), and notably evidence for up- and downregulation of different sets of proteins (Fig. 3 A, Fig. S3 A). Sarcomere and mitochondrial proteins were significantly enriched in the downregulated proteins, as reflected by gene ontology (GO) terms such as “contractile fiber part,” “respiratory chain complex,” “tricarboxylic acid cycle,” and “mitochondrion organization” (Fig. 3 A, B). Upregulated proteins were significantly enriched in terms related to splicing and RNA regulation, including “U2 snRNP,” “spliceosomal complex,” and “RNA splicing” (Fig. 3 B). Fold change values between the mRNA-Seq and mass spectrometry data were positively correlated (Fig. S3 B), and the Pearson’s correlation coefficient of $R^2 = 0.36$ (Spearman $R^2 = 0.44$) is consistent with previous estimates from various model systems (Becker et al., 2018; Brion et al., 2020; Koussounadis et al., 2015; Liu et al., 2016). Of the significantly differentially expressed proteins, while only 28 (2.5%) were significantly regulated on the gene level, 383 (34.5%) had significantly differentially regulated exons or contained novel intron retention or splicing events (Fig. S3 A, B). We conclude that loss of Scaf6 results in defects in mRNA splicing that lead to significant shifts in protein expression of affected transcripts.

To independently verify these protein-level expression changes, we employed a panel of GFP-tagged reporters expressed under native regulatory elements (Orfanos and Sparrow, 2013; Sarov et al., 2016; Spletter et al., 2015). We assayed GFP expression as well as protein localization in *scaf6-IR* IFM. Act88F, which is not significantly misregulated at the RNA-level in *scaf6^{-/-}* flies, does not show a change in GFP expression in *scaf6-IR* IFM (Fig. 3 D, Fig. S3 D-D’). By contrast, Mhc, Zasp66, Zasp52, Unc-89, Sls and Strn-Mlck, all of which are downregulated in the mass spec data and display changes in exon use, show a significant decrease in GFP expression (Fig. 3 C, D, Fig. S3 C-C’, E-G’). Additionally, the weep26-GFP reporter line, which is an endogenous tag in a specific isoform of Mhc, is mislocalized (Fig. S3 F-F’), indicating that a previously characterized developmental switch in Mhc isoform expression (Orfanos and Sparrow, 2013) is impaired in *scaf6-IR* IFM. These data confirm our mass spectrometry results and show that the expression level of multiple structural proteins is downregulated in IFM after loss of Scaf6.

To further validate that GO term enrichments observed in the mass spectrometry and mRNA-Seq data accurately reflect biological processes disrupted in mutant muscle, we tested if there was a mitochondrial defect, as suggested by the misregulation of

mitochondrial-associated genes and proteins (Fig. 2 X, Fig. 3 B). We used a Mito-GFP reporter to visualize mitochondria in developing IFM from *scaf6-IR* flies. At 48 h APF, when *scaf6-IR* myofibers are beginning to detach but sarcomere structure is still intact, the mitochondrial load in control and knockdown IFM is similar, and an extensive mitochondrial network ramifies throughout the sarcoplasm (Fig. S3 H, K, O). By 72 h APF, there are stark differences evident in the mitochondrial network. Although the area covered by mitochondria is similar (Fig. S3 O), the mitochondria are no longer evenly distributed and single mitochondrion are much larger than in control IFM (Fig. S3 I, L, N, O). By 1 d adult, IFMs contain significantly fewer mitochondria, and those that are still present are abnormally large and spindly (Fig. S3 J, M, N, O). This progressive disruption and loss of mitochondria mirrors the increasing severity of myofiber and myofibril phenotypes, confirming an impact on mitochondria after loss of Scaf6 and validating this finding from the proteomics.

***scaf6*^{-/-} mutant myofibers detach due to aberrant contractility**

Motivated by our verification of splicing and expression changes in structural genes at the RNA and protein levels, we investigated how misregulation of sarcomere proteins contributes to the phenotypes we observed in *scaf6*^{-/-} IFMs. To identify which sarcomere proteins are most strongly affected, we looked for genes that contain intron retention events, have significant changes in exon use, and are significantly downregulated in our proteomics data. There are 13 structural genes that meet these criteria: *Tropomyosin 1* (*Tm1*), *upheld* (*up*, TnT), *wings up A* (*wupA*, TnI), *Stretchin-Mlck* (*Strn-Mlck*), *Muscle LIM protein at 60A* (*Mlp60A*), *Z band alternatively spliced PDZ-motif protein 66* (*Zasp66*), *Zasp52*, *sallimus* (*sls*, Titin), *bent* (*bt*, Projectin), *Ryanodine receptor* (*RyR*), *Limpet* (*Lmpt*), *Unc-89* and *Mhc* (Fig. 3 E). This list is striking, as it contains Z-disc components critical to assembly and maintenance of sarcomere structure (*Mlp60A*, *Zasp66*, *Zasp52*, *Sls*, *bt*) as well as important regulators of muscle contractility including *RyR*, *Strn-Mlck*, *Mhc* and the tropomyosin complex (*Tm1*, *up*, *wupA*) (Fig. 3 E). Notably, all structural proteins with reported muscle hypercontraction phenotypes in *Drosophila* (*Tm1*, TnT, TnI, *Strn-Mlck*, *RyR*, *Strn-Mlck*, *Mhc*) are misregulated in *scaf6*^{-/-} flies, so we focused on aberrant actomyosin contractility as a possible developmental mechanism leading to the detachment of IFM myofibers.

We used two approaches to test if the detachment of *scaf6*^{-/-} IFM myofibers is a result of hypercontraction, which results from misregulation of actomyosin interactions.

We first evaluated the myofiber phenotype of *scaf6*^{-/-} IFM in the background of the *Mhc*¹⁰ allele, a mutation which results in loss of the IFM-specific hinge domain and dramatically altered actomyosin dynamics (Collier et al., 1990). While the majority of myofibers are detached in 1 d adult *scaf6*^{-/-} IFMs (Fig. 3 F), the detachment phenotype is significantly rescued in *Mhc*¹⁰, *scaf6*^{-/-} IFMs (Fig. 3 G, H). This demonstrates that myofiber detachment is dependent on myosin-mediated contractility. As *scaf6*^{-/-} IFMs are already detached by 48 h APF (Fig. 1 F), we next used a live-imaging assay to monitor spontaneous myofiber contractions, or twitching, during early pupal stages. Myofibrils assemble in developing pupae around 30 h APF, and twitching frequency and multi-twitch dynamics peak around 48 h APF (Spletter et al., 2018a). When we monitored twitching in *scaf6*^{-/-} and *scaf6-IR* flies shortly after myofibril assembly at 36 h APF, we found that IFMs twitched significantly more often than in controls (Fig. 3 K). However, by 48 h APF, IFMs in *scaf6*^{-/-} and *scaf6-IR* flies twitched significantly less often than in control flies (Fig. 3 I, I', J, J', L). We also noted that the total number of *scaf6*^{-/-} pupae where we can observe twitch events is less than in wildtype *w*¹¹¹⁸ flies (Fig. 3 K, L). This data demonstrates that dysregulation of muscle function is already evident before fiber detachment. We therefore conclude that mis-splicing and decreased expression of sarcomere proteins in *scaf6* mutants results in aberrant myosin contractility leading to hypercontraction and detachment of the IFM myofibers.

Behavior defects in *scaf6*^{-/-} flies reflect tissue-specific functions in muscle and neurons

The enrichments for neuronal related terms and splicing defects in synapse-related genes in our mRNA-Seq data (Fig. 2, Fig. S2) suggest that in addition to its role in muscle, Scaf6 might have a physiological function in the nervous system. To test this possibility, we decided to evaluate Scaf6 cellular and behavioral phenotypes in a tissue-specific manner. We first selected a panel of behavioral assays to test muscle and neuronal function, including larval crawling, flight ability, climbing ability, grooming efficiency, the righting reflex, eclosion competence and adult survival (Fig. 4 A, Fig. S4). *scaf6*^{-/-} mutant flies display severe impediments in most of these assays. *scaf6*^{-/-} flies are flightless (Fig. 1 I, Fig. 4 A), have a reduced ability to climb 5 cm in 5 seconds (Fig. S4 F), cannot effectively groom themselves when sprinkled with dust (Fig. S4 J, K), are slow to right themselves when they land on their back (Fig. S4 I), and show high rates of pharate and young adult lethality (Fig. 4 A, Fig. S4 C, D). These defects are restricted to the late pupal and adult

stages, as *scaf6*^{-/-} flies pupate normally and crawl as well as wildtype larvae (Fig. S4 A, B). These data establish a baseline of mobility and coordination in whole animal mutant flies, that allowed us to next evaluate tissue-specific requirements of Scaf6.

To determine which of these behaviors required Scaf6 function in muscle or in neurons, we employed tissue-specific RNAi knockdown via the Gal4-UAS system. Mef2-Gal4 is a muscle-specific driver, while Elav-Gal4 drives selectively in neurons. Mef2-Gal4 driven *scaf6-IR* flies were flightless (Fig. 1 I, Fig. 4 A), and had a minor but significant climbing impairment (Fig. 4 A, Fig. S4 F), but performed as well as control flies in all other behavioral assays (Fig. 4 A, Fig. S4 C, F, I, K). Elav-Gal4 driven *scaf6-IR* flies were weak fliers (Fig. 4 A, Fig. S4 E), and displayed significantly impaired climbing, grooming and righting ability (Fig. 4 A, Fig. S4 F, I, K). Strikingly, when we knocked-down *scaf6* in both muscle and neurons by combining the Mef2-Gal4 and Elav-Gal4 drivers, we replicated the mutant phenotype, including pharate and young adult lethality as well as the full constellation and severity of behavioral defects Fig. 4 A, Fig. S4 C, E, F, I, K). We did not observe defects in flight or climbing ability when *scaf6-IR* was driven by the glial-specific Repo-Gal4 or the fat body-specific ppl-Gal4, supporting the tissue-specificity of the observed phenotypes (Fig. S4 G, H). This data demonstrates that the lethality and severity of phenotypes observed in *scaf6*^{-/-} mutant flies depends on Scaf6 function in both muscles and neurons, where flight ability is strongly influenced by a muscle-specific component and climbing, grooming and the righting reflex have a stronger neuronal-specific component.

Scaf6 is required cell-intrinsically in IFM and motor neurons

To collect further evidence of a tissue-specific requirement for Scaf6, we next evaluated the cellular phenotype in tissue-specific knockdown conditions. In *scaf6*^{-/-} mutant flies at 72 h APF, in addition to severe IFM myofibril and sarcomere defects (Fig. 1 F, Fig. 4 D), we observed a dramatic reduction in the number of higher order motor neuron axon branches (Fig. 4 B, C, K, L). In Mef2-Gal4 driven *scaf6-IR* flies, although myofibril and sarcomere structure are compromised (Fig. 1 H, Fig. 4 G), motor neuron axons branch normally (Fig. 4 E, F, L). By contrast, in Elav-Gal4 driven *scaf6-IR* flies, myofibril and sarcomere structure are intact (Fig. 4 J), but IFM motor neuron axons show a significant reduction in higher order branches (Fig. 4 H, I, L). This data indicates that Scaf6 is required intrinsically in motor neurons for axon branching, and intrinsically in muscle for maintenance of myofibril and sarcomere structure.

To further support this conclusion, we performed tissue-specific rescue in the *scaf6*^{-/-} background. We started by testing rescue flies in our panel of behavioral assays (Fig. 4 M). Mef2-Gal4 driven UAS-myc-scaf6 is not sufficient to rescue eclosion, flight, climbing, righting or grooming behaviors, and performance of these flies does not differ significantly from Mef2-Gal4, Elav-Gal4 and UAS-myc-scaf6 controls in the *scaf6*^{-/-} background (Fig. 4 M, Fig. S4 C, F, I, K). However, since the rescue of flight behavior relies on a near-perfect balance of temporal and spatial protein expression, we also evaluated the cellular phenotype in Mef2-Gal4 rescue flies. Notably, in 1 d adult flies we observed a partial rescue of muscle phenotypes. IFM myofibers were still attached, as compared to *scaf6*^{-/-} flies where 100% of fibers are detached at this stage (Fig. 1 J, Fig. 4 N, O). Further, although defects in myofibril structure were still evident, we observed a significant improvement in myofibril integrity and sarcomere structure (Fig. 4 N', O'). This shows that muscle-specific expression can indeed rescue myofiber and myofibril defects. Additionally, Elav-Gal4 driven UAS-myc-scaf6 was able to significantly rescue climbing, righting and grooming ability (Fig. 4 M, Fig. S4 F, I, K). Neuronal-specific rescue also improved adult survival, although it did not significantly rescue eclosion competence (Fig. 4 M, Fig. S4 C). These data demonstrate cell-intrinsic function for Scaf6 in muscle and neurons, and further illustrate how wildtype behaviors depend on Scaf6 activity in both tissues.

CHERP plays a role in myoblast proliferation and differentiation, but not apoptosis

Since protein structure and domains are conserved between fly Scaf6 and vertebrate CHERP, we investigated CHERP regarding its' function in vertebrate muscle cell lines. We first performed a knockdown in mouse C2C12 cells and tested proliferation and apoptosis rates, the two main functions of CHERP known from previous studies in HEK293T and cancer cell lines (Lin-Moshier et al. 2013, De Maio et al. 2018, Wang et al. 2019, Martin et al. 2021). Proliferation rates, assessed in Ki67 stainings (Figure 7 **A1, B1, C1** and **E**), decreased by nearly 50% in CHERP knockdown cells, which points towards a positive regulation of proliferation through CHERP. This is consistent with previous reports, making CHERP a proto-oncogene potentially involved in various cancer types (Wang et al. 2019, Martin et al. 2021). In line with lower proliferation rates, the overall number of viable cells after 24 hours was reduced by approximately 40% as shown in Figure 7**D**. In contrast, we did not observe increased rates of apoptosis as reported in

(Wang et al. 2019), shown in Figures 7 A2, B2, C2 and F. We thus conclude that the lower number of viable cells is a result of reduced proliferation rather than increased apoptosis or a combination of both, underlining a tissue-specific function of CHERP in muscle compared to other cell types. To gain information about genes regulated by CHERP, either their expression level or splicing state, we performed RNAseq in untreated, no-target and siCHERP C2C12 samples in replicates. In total, 1373 genes were differentially expressed between untreated and CHERP knockdown cells with 408 being down- and 965 being up-regulated ($\log_2FC > 1$ & < -1 and $p\text{-value} < 0.05$, Figure 7 F). The top hits among the downregulated genes are the prolactins Prl2c2 and Prl2c3 which function as growth factors during embryonic development and positively regulate proliferation. We find many sarcomeric proteins up-regulated in CHERP knockdown cells, among others Troponin (Tnnt1, Tnnt3 and Tnnc1), Titin (Ttn), Obscurin (Obscn), Nebulin (Neb) and myosin heavy chain 8 (Myh8). Further, the transcription factor Myogenin (Myog) is the most significant upregulated gene with a $p\text{-value}$ of 2.2×10^{-254} and a \log_2 fold change of 3.7. Myog is a master regulator of myogenesis with numerous target genes and required for proper myoblast differentiation. We tested the expression change of Myog on the protein level using immunofluorescence and indeed confirmed the up-regulation of Myog in CHERP knockdown cells observed in our RNAseq data (Figure 7H, I, J, K). At the same time, Tbx2 is also highly up-regulated in siCHERP cells. The protein encoded by this gene is a transcriptional repressor negatively regulating cell cycle exit and has to be down-regulated upon differentiation (Zhu et al. 2014). This seemed to be contrary to the general notion that proliferation is impaired in siCHERP cells. It was shown that Tbx2 represses p21 (Cdkn1a), a protein that inhibits cell cycle progression – however, in our data set p21 levels are elevated by $\log_2FC = 0.8$. We thus hypothesize that high Tbx2 RNA levels are either 1) not reflected on the protein level, 2) elevated due to a compensation mechanism or 3) Tbx2 plays a more complex role than anticipated, not repressing cell cycle exit exclusively. We thus checked Tbx2 expression in another dataset we generated (data not shown), a time course of differentiating human myoblasts which revealed elevated Tbx2 RNA expression at Day 1 and 2 of differentiation by $\log_2FC = 1.7$, when myoblasts usually start exiting the cell cycle. Although human and mouse myoblast differentiation is not congruent, muscle development is highly conserved among species.

We wondered if the trend of proliferation being down- and differentiation being upregulated proves to be true also on a broader level, so we looked for enriched GO-terms

among the differentially expressed genes running the analysis separately for up- and downregulated genes. Strikingly, the top hits in downregulated genes are cell cycle related with whole networks being affected, e.g. the transcription factor Foxm1 and its targets, the cyclin dependent kinase Cdk1 and cyclin B1 Ccnb1 (Figure 7G). In contrast, cell cycle related terms are not enriched in upregulated genes, while muscle differentiation and contraction are highly enriched, as well as sarcomere organization. We further screened our data using a list of Myog target genes to find out whether the upregulation of Myog actually has an effect on their expression and found many of these genes to be upregulated (Figure 7L, M).

For proper myoblast differentiation in cell culture, two main environmental cues are required: physical contact between the cells and low serum conditions. This leads to cell cycle withdrawal and transcriptional activation of the differentiation program. Cell cycle exit is an effect of stress pathways induced by starvation (low serum), which are mediated by MAPK and p53 pathways, which – under certain circumstances – also induce apoptosis, a pathway that was previously associated with CHERP. Indeed, we find several MAP kinases and interacting proteins mis-regulated in siCHERP cells. It is thus conceivable that CHERP regulates cell cycle progression and apoptosis, and thus also differentiation, through regulation of the MAPK pathway, however, further experiments are necessary to investigate this theory. We further hypothesize that CHERP knockdown induced cell cycle exit is sufficient to partially activate the differentiation program as shown by the upregulation of myogenic factors as Myog, Mef2a and sarcomere proteins. However, at this point we cannot exclude the possibility that CHERP regulates differentiation by direct targeting of relevant genes.

CHERP targets muscle specific genes through interaction with the U2 spliceosome

To test exon usage in CHERP knockdown cells, we performed a DEXSeq analysis and found 1664 exons in 1219 genes used differentially ($\log_2FC > |1|$, $p\text{-value} \leq 0.05$). The majority of exons (~ 70%) had positive \log_2 fold changes, i.e. are included, suggesting a major role for CHERP in masking splice sites. Again, there are sarcomeric proteins like Tropomyosin (Tpm), Nebulin (Neb) and Troponin (Tnnt) among the mis-regulated genes (Figure 8A). Only 8.5% of the DE genes identified with DESeq2 are also alternatively spliced in DEXSeq and thus potentially regulated through alternative exon usage. We then checked AS events using MAJIQ, which identifies intron retention and alternative splice sites in addition to exon usage. Here we find 986 genes alternatively spliced with 1436

AS events (exon skipping, alternative 3' and 5' splice sites, intron retention), increasing the proportion of DE genes potentially regulated by splicing to 11%. When checking all AS events detected in control and siCHERP, we find more total events in the latter while only 30% overlap, consistent with our findings in drosophila muscle (Figure **8B**). In MAJIQ, 60% of the alternatively used exons are included and 40% excluded, reflecting the same trend as DEXSeq. While alternative splice sites are evenly in- and excluded in CHERP knockdown cells, there is a significantly higher amount of retained introns (Figure **8C**). We then performed a GO term enrichment analysis on our DEXSeq dataset, revealing metabolism related genes being strongly enriched. Further, genes regulating gene expression and signaling are differentially spliced, which could explain the high amount of DE genes that are not regulated through splicing (Figure **8D**). These results strongly indicate that CHERP regulates muscle-specific genes and pathways through alternative splicing. We then wondered if the mechanism of splicing and splice site recognition is also muscle-specific, i.e. different from non-muscle tissue. We thus performed immunoprecipitation (IP) of CHERP in mouse myoblasts and myotubes, 2 primary human muscle cell lines as well as in HEK293T cells, followed by mass spectrometry to find interactors. The pulled proteins were very similar between the samples with the biggest difference between two different antibodies (sc-100650 in HEK293T and myoblasts, ab15951 in myotubes). Notably, a pulldown of CHERP using the polyclonal Abcam antibody ab15951 was highly specific in mouse and human myotubes, but failed to pull down CHERP in non-muscle cells and myoblasts, suggesting a differing protein conformation in myotubes compared to proliferating cells. 13 proteins were found in every IP - one of them is CHERP – the remaining 12 can be considered as potential direct interactors compared to the other enriched proteins (Figure **8E**). Consistent with previous reports about the function of CHERP, these 12 proteins are all part of the U2 spliceosomal complex. 6 belong to the U2 core complex, mostly the Sf3b subcomplex which is responsible for the branch point adenosine (BPA) recognition. Besides CHERP, there are 3 further U2 accessory proteins: Rbm17 and U2surp, which form a subcomplex with CHERP, and Dhx15. Rbm5, Rbm10 and Sugp1 are recruited to the 3' SS at A complex formation (Figure **8F**). This makes a role in 3' SS recognition or stabilizing 3' SS binding likely, especially when considering that intron retention is one of the dominant AS events in CHERP knockdown cells. We conclude from these results, that the mechanism of splicing through CHERP is rather general than tissue-specific.

Since our findings of Scaf6/CHERP function in muscle suggest a vital role, we wondered if it contributes to muscle diseases. Cachexia is defined as muscle loss and wasting accompanying severe diseases like cancer and AIDS. Here, we find CHERP mRNA upregulated indicating a potential contribution to the cachexia phenotype (Figure 8G). Similarly, CHERP is also transcriptionally upregulated in myotonic dystrophy type 1 (DM1) as validated by qPCR (Figure 8H). DM1 is one of the best studied splicing diseases with two splicing factors, MBNL1 and CELF1, being mis-regulated leading to mis-splicing of muscle-specific genes. However, these two factors alone do not explain all aspects of the DM1 symptomology making other factors playing a role likely. We then checked CHERP expression on the protein level in myoblast cells derived from two different DM1 patients and, surprisingly, found it strongly downregulated by more than 80%, in contrast to mRNA expression (Figure 8I). This suggests a negative feedback loop assuring CHERP levels below a certain threshold. This is reasonable considering that CHERP positively regulates cell cycle and proliferation which leads to increased cell growth if not tightly regulated. Consistently, CHERP was suggested to be a proto-oncogene before. Using a publicly available RNAseq dataset generated in DM1 muscle biopsies (Wang et al. 2019, GSE86356), we performed a GO-term enrichment analysis under the same conditions as for our CHERP knockdown experiment and compared the results. Notably, the enriched terms of siCHERP treated cells nearly completely overlap with the ones in DM1 with only 18 terms not enriched in DM1 (Figure 8J). Among the overlapping terms are signaling, differentiation and muscle contraction. We thus think that a contribution of CHERP to the DM1 phenotype is conceivable and should be investigated in future experiments.

Materials and Methods

Fly stocks and husbandry

Fly work was performed with approval in Germany according to §15 GenTSV (license number 55.1-8791-14.1099). Fly stocks were maintained at room temperature using standard culture conditions. Experimental crosses were maintained at 27 °C, unless otherwise noted. Fly food was prepared in a water-jacketed cooker by combining 16 L water, 150 g soy flour, 1,300 g corn flour, 300 g yeast, 130 g agar, 1,300 g molasses, and 650 g malt extract. After cooling, food was supplemented with 415 mL 10% Nipagin and 295 mL acid mix containing 3% phosphoric acid and 21% propionic acid. Food was

aliquoted with a peristaltic pump, allowed to set at room temperature (RT) and stored at 4 °C until use.

w¹¹¹⁸ (BDSC:3605) was used as the wild-type control background. *scaf6^{12M9}* is a point mutation generated by EMS mutagenesis resulting in a CAG > TAG transition at aaXXX and has been characterized previously (Hayashi et al., 2014). *scaf6^{M2M1}* is a CRISPR allele resulting in a single base deletion at Tyr100 causing a frame shift and early stop. It is an indel generated by Cas9 targeting with sgRNA TTTGTAGTTGTAATCCTCCAG, according to the protocol published in (Ni, Perrimon et al, Nat Methods, 2011, PMID 21460824). Other stocks used include *UAS-mito-HA-GFP* (BDSC:8443), *Him-nuc-eGFP* (Soler and Taylor, 2009), *rhea-C-YPet* (Lemke et al., 2019), and *UAS-reaper; UAS-hid* (kind gift of Dr. Ulrike Gaul). GFP trap lines included *weeP26-GFP* (Clyne et al., 2003), *Zasp66-GFP* (BDSC:6824) and *Zasp52-GFP* (BDSC:6838) (Buszczak et al., 2007; Morin et al., 2001). GFP-tagged fosmid reporter lines included *Mhc-GFP* (fTRG500), *Act88F-GFP* (fTRG10028), *sls-GFP* (fTRG477), *unc-89-GFP* (fTRG1046), and *Strn-Mlck-IsoR-GFP* (*Strn4*) (Sarav et al., 2016; Spletter et al., 2015). Gal4 driver lines used include: *Mef2-Gal4* ((Ranganayakulu et al., 1996), BDSC:27390), expressed continuously in all muscle; *Act88F-Gal4* (Bryantsev et al., 2012), expressed from about 24 h APF and largely restricted to IFM; *Salm-Gal4* (Schönbauer et al., 2011), expressed in IFM from 8 h APF; *1151-Gal4* (Roy and Vijay Raghavan, 1997) and *Him-Gal4* (Spletter et al., 2018a), expressed in myoblasts; *Elav-Gal4*, expressed in neurons; *repo-Gal4*, expressed in glial cells; and *ppl-Gal4* (BDSC:58768) outcrossed to wildtype (BDSC:2202), expressed in fat body.

Behavioral assays

Eclosion and survival. To assay pupal eclosion, at least 50 staged pupae from each genotype were collected onto wetted filter paper in 30 mm Petri dishes and incubated at 27°C. The eclosion rate was calculated as the number of successfully eclosed adults divided by the total number of pupae. To monitor adult survival, newly eclosed adult male flies were collected and kept on standard fly food at 27°C. The number of surviving flies was monitored daily for 5 days. At least 5 groups of 50 flies were assayed for each genotype.

Larval crawling assay. 3rd instar larvae were collected, and 4 larvae at a time were placed into Petri dishes (90 mm x 15 mm) filled with a 1% agarose gel. Petri dishes were illuminated from above, and larvae were allowed to move freely for 30 seconds and

recorded using a Dcx USB compact camera using the provided uc480 software (ThorLabs, #DCC1645C). All experiments were performed at 25°C, and 42-72 larvae per genotype were tested. Video files were analyzed using the ImageJ plugin Trackmate to track individual larva and calculate mobility rates as described previously (Tinevez et al., 2017).

Adult behavior assays. For all adult behavior assays, 1 d adult male flies were collected on CO₂ and recovered at least 24 hours before testing. **Flight ability:** Flight test assays were performed as described previously (Schnorrer et al., 2010). Flies were introduced at the top of a 1-meter-long cylinder divided into five zones. Flies that land in the top two zones are considered to have normal flight ability, those landing in the middle two zones are weak fliers, and flies falling to the bottom are flightless. At least 50 flies were assayed for each genotype. **Climbing ability:** Climbing was assayed using a modified rapid iterative negative geotaxis (RING) assay, as described previously (Nichols et al., 2012). Flies were flipped in groups of 10 into a 50 ml Falcon tube (Greiner bio-one, #E211233C) with a line marking the 5 cm mark. Flies were knocked to the bottom, and the number of flies able to climb 5 cm in 3 or 5 seconds was recorded. Each vial was tested three times, with a 1-minute recovery period between tests. Climbing ability was calculated as the ratio of successful climbers to the total number of tested flies. At least 50 flies and 5 independent test groups were measured per genotype. **Righting reflex:** 3-5 male flies were transferred to a 50 ml transparent cylinder (Nerbeplus, #11-881-0051). Flies were knocked to the bottom to stimulate the righting reflex, and recorded with a USB compact camera at a frame rate of 30 frames per second. The righting reflex was determined based on the number of frames until a fly that landed on its wings was standing on its legs. At least 50 flies were measured per genotype. **Grooming assay:** The grooming assay was adapted from (Barradale et al., 2017). Each well of a 24-well plate (Greiner Bio-one, Cellstar, #E20053R8) was pre-loaded with 0.05 mg of brilliant yellow dye (Sigma-Aldrich, #MKCB2318V). A single, ice-anesthetized fly was placed in each well of the plate, and the plate was vortexed for 30 seconds on high to spread the dye homogeneously over each fly. 12 flies were sacrificed immediately ($t = 0$), while the other 12 flies were incubated at 27°C for 30 min ($t = 30$) to allow flies to groom and remove the dye. Flies were sacrificed in 1 ml of 100% ethanol and incubated for 5 hours to completely dissolve the dye. A 5-fold dilution of the ethanol solution was measured at an absorbance of 397 nm in an Infinite M1000 Pro plate reader (Tecan). Grooming ability was calculated as the normalized ratio of $1 - \text{absorbance}(\text{individual}^{t=30} / \text{average}^{t=0})$. At least 30 flies were tested per genotype.

Cell culture

Human embryonic kidney 293T (HEK293T) cells were a gift of A. Ladurner. Cell lines were cultured at 37 °C and 5% CO₂. HEK293T and C2C12 mouse myoblast cells were grown in high glucose Dulbecco's modified Eagle medium (DMEM) supplemented with 10% FBS, 100 U/mL penicillin-streptomycin and 2.2 mM Glutamine. Cells were trypsinized (EDTA-Trypsin, Sigma) and passaged 1:10 every two days. Automated cell counting was performed with a Vi-Cell XR (Beckman Coulter).

C2C12 cells were differentiated to myotubes over a 5-to-7-day time course in high glucose DMEM supplemented with 2% horse serum, 1 µM Insulin, 100 U/mL penicillin-streptomycin and 2.2 mM Glutamine, with fresh media provided every other day. As C2C12 differentiation efficiency is around 40-60% (Parker et al., 2016), highly enriched myotube samples (90-95% pure) were obtained by partial trypsinization and differential centrifugation. Culture dishes were rinsed and treated with 4 mL of 1x Trypsin (~ two minutes), and monitored under a microscope until myotubes had detached, but undifferentiated myoblasts remained attached. Myotube cells were then pelleted by centrifugation at 150 rpm for 1 minute, and supernatant containing co-trypsinized myoblasts was removed.

Primary human myoblasts were obtained from the Muscle Tissue Culture Collection (MTCC) at the Friedrich-Baur-Institute (Department of Neurology, Ludwig-Maximilians-Universität München, Munich, Germany). All control and patient materials were obtained with written informed consent of the donor. Ethical approval for this study was obtained from the ethical review committee at the Ludwig-Maximilians-Universität München, Munich, Germany (reference 45-14). Primary myoblast lines were grown in Skeletal Muscle Cell Basal Medium Kit including Supplement with FBS (PELOBiotech, Cellovations, PB-BH-272-0090). Cells were passaged daily at a ratio of 1:2 or 1:3 to maintain a confluency of 30-80%. Myoblast identity was monitored by immunostaining against Desmin and Ki67. Differentiation was induced after 3-4 passages at 100% confluency in Skeletal Muscle Cell Growth medium (PELOBiotech, Cellovations, PB-MH272-0000) supplemented with 5% horse serum, 30 U/mL penicillin-streptomycin and 1x GlutaMAX (Gibco, 35050-061). When differentiation was complete (6-7 days), 90-95% pure myotube samples were isolated using partial trypsinization and differential centrifugation as described above.

CHERP knockdown

CHERP knockdown was performed with an ON-TARGETplus SMARTpool containing four different siRNAs targeting mouse CHERP (Dharmacon, L-058389-01-0005). An ON-TARGETplus Non-targeting Control Pool (Dharmacon, D-001810-10-05) was used as a negative control. C2C12 cells were seeded 30,000 cells per well in a 24-well plate. After 24 hours, cells were transfected with DharmaFECT 1 (Dharmacon, T-2001-07A) plus siRNA to a concentration of 25, 40 or 50 nM per well, following the manufacturers protocol. Based on Western blot and RT-qPCR, knockdown was as effective at 25 nM as at higher 40 and 50 nM siRNA concentrations. After 72 hours, cells were processed for subsequent experiments. For RNA isolation, cells were washed, lysed in TRIzol and stored at -80 °C. For protein assays, cells were trypsinized, washed 3 times in 1x PBS and resuspended in 100 µL RIPA lysis buffer (1% NP40, 0.1% SDS, 50mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.5% sodium deoxycholate, 1 mM EDTA, 1 mM PMSF, 1x cOmplete™ Protease-Inhibitor-Cocktail (Sigma-Aldrich, #11836170001)). Lysates were centrifuged for 40 minutes at 15,000 rpm at 4 °C, and supernatants stored at -20 °C. Protein concentration was measured using the Bradford Assay, and 100 µg of total protein was used for Western blotting.

Immunohistochemistry and microscopy

Drosophila wing disc, leg disc and brain were dissected in 1X PBS with Biology-grade #5 forceps (Fine Science Tools). Samples were fixed in 4 % PFA in 0.5 % PBS-T (1X PBS + Triton X-100) for 30 minutes. Hemi-thorax sections to visualize IFM morphology were performed as described previously (Weitkunat and Schnorrer, 2014). After washing in 1X PBS-T, all samples were blocked for 1 hour with 5% NGS (Jackson ImmunoResearch #105315) in 0.5% PBS-T. Samples were then incubated at 4°C overnight on a rocking shaker with primary antibody in 0.5% PBS-T. Primary antibodies included rat anti-Elav (1:50, DSHB), mouse anti-Myc (1:500, Abcam #GR171077-8), rabbit anti-H2AZ (1:2000, (Schauer et al., 2013)), rabbit anti-phospho-histone H3 (1:200, Cell Signaling #9701s), rabbit anti-cleaved *Drosophila* Dcp-1 (1:200, Cell signaling #9578s), mouse anti-FLAG (1:200, clone M2, Sigma Aldrich, #F1804), goat anti-HRP conjugated with Alexa 647 (1:100, Jackson ImmunoResearch #134899), GFP-Booster Atto488 (Chromotek, #gba488-100). Samples were washed for 1 hour at room temperature and then incubated in 0.5% PBS-T with secondary antibody at 4°C overnight. Secondary antibodies, all obtained from Invitrogen Molecular Probes, included Alexa488 goat anti-mouse (1:500), Alexa488 goat anti-rabbit (1:500), Alexa546 goat anti-rabbit

(1:500), Alexa647 goat anti-rat (1:200), and rhodamine phalloidin (1:500). Samples were washed with 1X PBS for 30 min and mounted in Vectashield containing DAPI (#ZH1021). Images were acquired with a Leica TCS SP8X WLL upright confocal microscopy running Leica LAS X software at the Core Facility Bioimaging of the LMU Biomedical Center.

For cell culture lines, cells were grown on coverslips and fixed and permeabilized in ice cold methanol. Coverslips were washed with 1X TBS. C2C12 cells were blocked in 3% NGS in TBS-T (0.3% Triton-X100 in 1x TBS) for 1 hr at RT, while human cells were blocked in 0.1% cold water fish gelatine solution diluted in deionized water (Sigma-Aldrich, #9000-70-8). Samples were incubated with primary antibodies in blocking solution overnight at 4 °C. Primary antibodies included: rabbit anti-CHERP (1:500, Abcam ab15951), mouse anti-Desmin CloneD33 (1:25, Dako), rabbit anti-KI67 (1:100, Thermofisher), mouse anti-Lamin-A 4A7 (1:50, Glenn Morris). Samples were washed in TBS-T, and incubated at least 1 hour at RT in blocking solution with secondary antibodies from Invitrogen Molecular Probes including: Alexa488 goat anti-rabbit (1:500), rhodamine phalloidin (1:500), Alexa594 goat anti-rabbit (1:500). Coverslips were washed and mounted in Vectashield containing DAPI. Images were acquired on a Leica TCS SP8X WLL upright confocal or on an Olympus FluoView 1200.

Spontaneous twitching assay

Spontaneous contractions during IFM development were monitored as described previously (Lemke and Schnorrer, 2018; Spletter et al., 2018a). *Rhea-C-YPet*, a fluorescent tagged Talin construct that localizes to myofiber tips (Lemke et al., 2019), was crossed into a wild-type, *scaf6^{-/-}* or *scaf6-IR* background. A small window was removed from the pupal case using forceps, and pupae were mounted in 50% glycerol in slotted slides. Live-image movies were recorded every 0.65s for 10 min using a 20X water objective on a Leica SP8X WLL upright confocal microscope. Twitches were manually scored in Image J and tabulated in Excel. Data were normalized per 5 min interval and plotted in GraphPad Prism.

Induction of osmotic stress and apoptosis

To induce acute osmotic stress in *w¹¹¹⁸* IFM, after removal of head, wings and abdomen in 1X PBS, thoraxes were incubated for 30 min. at room temperature on a

rocking shaker in either 250 mM NaCl (Roth, #3957.5) or 500 mM sucrose (Roth, #4661.3) in 1X PBS. Control samples were incubated in 1X PBS. Samples were then fixed, stained and prepared for imaging as described above. To induce apoptosis in *w¹¹¹⁸* myoblasts, 3rd instar larval wing discs were dissected in 1X PBS and incubated for 6 hours at room temperature on a rocking shaker in 10 nM Actinomycin D (Sigma-Aldrich, #A1410) in 1X PBS. Samples were then fixed, stained and prepared for imaging as described above.

Image analysis

Confocal image analysis was performed in Image J. Muscle detachment was scored manually from 10x z-projection images of whole thorax. Sarcomere length and myofibril width were measured based on rhodamine-phalloidin staining with the MyofibrilJ plugin (Spletter et al., 2018b). Motor neuron axon complexity was scored manually from HRP-647 staining in z-projections of IFM images that measured 88 μm x 88 μm x 40 μm with a step-size of 2 μm . Myoblast numbers were quantification using Cell-Count-OpenCV designed by Annie Yim (<https://github.com/anniepyim/Cell-Count-OpenCV>). This script employs Python OpenCV to count cell numbers in .lif files. Plotting and statistical tests of significance were performed in GraphPad Prism.

RNA isolation and RT-qPCR

IFM, brain and leg tissues were dissected from pupa at 72 h APF in cold 1X PBS as described previously (Kao et al., 2019). After centrifugation and buffer removal, tissues were homogenized in TRIzol (Ambion, #16655301) with a microcentrifuge tube pestle. For cell lines, cells were washed in 1x TBS, and then lysed in 1-2 mL TRIzol. Samples were stored at -80 °C in TRIzol. RNA was isolated following the manufacturer's instructions. RNA concentration was measured using a Qubit 2.0 Fluorometer (Invitrogen) and RNA integrity evaluated on a Bioanalyzer (Agilent Technologies). cDNA was synthesized from equal starting concentrations of total RNA with the SuperScript III kit (Invitrogen, #18989) following the manufacturer's protocol, after pre-treatment of total RNA samples with DNaseI (New England Biolabs, #M0303). For RT-PCR, cDNA template was amplified for 35-40 cycles using One-Taq Quick-Load DNA polymerase (New England BioLabs, M0509L), and the resulting PCR products were visualized on a 1-2% agarose gel with a 100 bp DNA ladder (New England BioLabs, N3231L) or a 1kb

Plus DNA ladder (New England BioLabs, N3200L). For RT-qPCR, cDNA template was amplified for 40 cycles using PowerTrack SYBR Green Master Mix (Thermo Fisher Scientific, #01094810) on a QuantStudio™ 3 Real-Time PCR system (Thermo Fisher Scientific). Relative expression was quantified using the $-2^{\Delta\Delta CT}$ method using normalization to expression of *RNA polymerase II subunit C*. All primers are provided in Supplementary Table XXX.

CHERP immunoprecipitation

CHERP was immunoprecipitated from HEK293T cells (one 10 cm plate), mouse myoblast C2C12 cells (one 15 cm plate), differentiated C2C12 myotubes (two 10 cm plates), and primary human myotubes (two 10 cm plates). After reaching confluency, cells were trypsinized, washed in fresh medium followed by chilled 1x PBS and resuspended in 1 mL NP40 lysis buffer (1% NP40, 50 mM Tris-HCl, 150 mM NaCl) containing 1x cOmplete™ Protease-Inhibitor-Cocktail (Sigma-Aldrich, #11836170001), 1 mM PMSF and 0.025 U/μL Benzonase Nuclease (Sigma-Aldrich, #E1014). After a 30-minute incubation on ice, cells were homogenized 20 strokes with a dounce homogenizer. Lysates were centrifuged for 40 minutes at 15,000 rpm at 4 °C. Protein concentration was measured by Bradford Assay (Bio-Rad, #5000006), with final concentrations ranging from 12-18 mg/mL. 4 μg of mouse anti-CHERP (HEK293T, C2C12 myoblasts; SantaCruz, sc-100650) or rabbit anti-CHERP (C2C12 myotubes, human primary myotubes; Abcam, ab15951) was added to each lysate and incubated at 4 °C overnight on a rotator. Samples were incubated with 1 mg Dynabeads™ Protein G magnetic beads (ThermoFisher, #10003D) for one hour at 4°C, then the beads were washed in 0.02% PBS-T (1x PBS, pH 7.4, 0.02% Tween, x mM NaCl) for 3x 10 min + 3x 5 min. Final salt concentrations in the wash buffer for each sample are as follows: HEK293T, 134 mM NaCl; C2C12 myoblasts, 250 or 500 mM; C2C12 myotubes, 250 mM; human primary myotubes, 250 mM). Proteins bound to the beads were eluted with 2x Laemmli buffer for Western Blot or Silver staining, or processed for mass spectrometry.

Western Blotting and Silver Staining

Samples were denatured by incubation in 2x Laemmli buffer for 10 minutes at RT, and boiled at 95 °C for 5 minutes. Magnetic beads were removed if necessary, and denatured proteins were separated together with a PageRuler™ Prestained Protein Ladder

(ThermoFisher, #26616) on a 10% SDS polyacrylamide stacking gel at 200 V for 45 minutes. Proteins were blotted onto a nitrocellulose membrane for one hour at 100 V. The membrane was blocked in 5% non-fat milk solution in 0.5% Tween-TBS buffer (TBS-T) for at least one hour at RT or overnight at 4 °C. The membrane was incubated with rabbit anti-CHERP (1:1000, Abcam, ab15951) for at least 1 hour at RT or overnight at 4°C. The membrane was washed 3x 10 min. in TBS-T, and incubated for one hour at RT with HRP-conjugated goat anti-rabbit secondary antibody (1:10,000, Bio-Rad). Membranes were washed 3x 10 minutes, incubated with Pierce ECL Western Blotting Substrate (Thermo Scientific, #32106) and exposed to X-ray films (Fuji medical X-ray, Super RX-N). Films were digitized and semi-quantitative analysis was performed in Image J.

For silver staining, the gel was fixed for one hour in fixation buffer (50% MeOH, 12% glacial acetic acid, 2% formaldehyde), washed 2x 10 minutes in 50% EtOH, incubated for 1 minute in 0.02% (w/v) sodium thiosulfate pentahydrate ($\text{Na}_2\text{S}_2\text{O}_3 \times 5 \text{H}_2\text{O}$), rinsed with water, incubated with 0.2% (w/v) silver nitrate (AgNO_3) in 3% formaldehyde and rinsed with water. The silver stain was developed for ~ 3 minutes in 6% (w/v) sodium carbonate, containing 2 mL 0.02% sodium thiosulfate pentahydrate and 50 μL formaldehyde.

mRNA transcriptome sequencing

Tissue dissection and sample preparation from *Drosophila* was performed as described previously (Kao et al., 2019; Spletter et al., 2018a). IFMs from at least 100 flies per replicate were dissected from control w^{1118} and mutant $scaf6^{-/-}$ flies at 30 h and 72 h APF. Legs and brains were dissected at 72 h APF from at least 100 and 200 flies per replicate, respectively. Total RNA was isolated using TRIzol from two biological replicates per genotype. Library preparation and sequencing were performed by Genewiz (South Plainfield, NJ, USA) and LC Sciences (Houston, TX, USA). Poly-A-selected and stranded libraries were sequenced as 100 bp Illumina paired-end reads with a depth >60 million reads per library.

Mass-spectrometry

IFMs from control w^{1118} and mutant $scaf6^{-/-}$ flies were dissected in cold 1X PBS at 72 h APF. Samples were spun-down and snap frozen on dry ice after removing buffer.

IFMs from 40 flies were pooled for each biological replicate, with in total 4 replicates per genotype. After thawing on ice, samples were processed using the PreOmics iST Sample Preparation Kit (Preomics, #0000.0061) following the manufacturers instructions. Prepared peptides were submitted to the Protein Analysis Unit (ZfP) at the LMU Biomedical Center, and label-free analysis was performed on an LTQ Orbitrap XL with Ultimate 3000 (Thermo Fischer Scientific). Peptides and protein groups were identified using MaxQuant (Cox et al., 2014) and data processing and differential expression analysis were performed in Perseus (Tyanova and Cox, 2018). To retain biologically relevant protein groups with “true negative” missing intensities between mutant and control samples (ie missing not at random, MNAR), we filtered the data by requiring at least three replicates in either group to contain a value, and we imputed missing values by replacement with a constant value (lowest observed intensity – 1). Differential expression was tested by t-test with FDR = 0.05. Results were exported and further analysis performed in R.

For CHERP immunoprecipitation mass spectrometry, samples in 4 biological replicates were prepared by on-bead tryptic digestion. Beads were washed three times with 100 μ L NH_4HCO_3 (50 mM), and proteins were eluted by adding 100 μ L trypsin solution (10 ng/ μ L trypsin, 1M urea, 50 mM NH_4HCO_3) and incubated at a rotation rate of 1,400 rpm for 30 min. at 25 °C. The supernatant together with two 40 μ L NH_4HCO_3 washes was transferred to a low-bind tube (Eppendorf). DTT was added to a final concentration of 1 mM and samples were incubated overnight at 25 °C at a rotation rate of 500 rpm. Samples were treated with 3 μ L of 130 mM iodoacetamide (IAA) and incubated for 30 min. at 25 °C protected from light. Alkylation was quenched with 1M DTT and trifluoroacetic acid (TFA) for 10 min. at 25 °C. Samples were desalted using SDB-RPS StageTips (3M Empore PN 2241), eluted with 60 μ L freshly prepared elution buffer (80% CAN, 25% NH_4OH), and dried in a SpeedVac for 1 hour at 45°C and 2000 rpm. Samples were resuspended in 15 μ L MS loading buffer (2% ACN, 0.3% TFA), resuspended by water bath sonication for 5 minutes, and analyzed at the Protein Analysis Unit (ZfP) at the LMU Biomedical Center on a short gradient using a ThermoFisher QExactive mass spectrometer. Proteins were identified using MaxQuant (Cox et al., 2014) and normalized intensity values (iBAQs) were analyzed using the Bioconductor Differential Expression analysis of Proteomics data package (DEP) (Zhang et al., 2018). We required significant genes to be enriched in all four replicates ($p = 0.05$), and used a cut-off of $p = 0.001$ for inter-cell line comparison.

Bioinformatics

mRNA-Seq data from *Drosophila* IFM, leg and brain from w^{1118} and $scaf6^{-/-}$ were generated as part of this work and are available from GEO under accession numbers GSE194199 and XXXX. Reads were remapped with STAR to BDGP6.22, indexed with SAMtools and reads summed with featureCounts. Analysis and visualization were performed in R. Differential gene expression and exon use were analyzed with DEseq2 (Love et al., 2014) and DEXseq (Anders et al., 2012b), respectively. Thresholds are listed in figure legends, but standardly were defined as $p\text{-value} \leq 0.05$ and $\text{abs}(\log_2 \text{fold change}) \geq 1.2$. GO term analysis was performed with Metascape (Zhou et al., 2019) or GOrilla (Eden et al., 2009). Packages are listed in Supplemental Table YYY. Membership lists for GO term categories “x” and “y” were downloaded from Flybase and are listed in Supplemental Table Y. Gene categories “sarcomere proteins” and “mitochondrial proteins” are from (Nikonova et al., 2022; Spletter et al., 2018a).

mRNA-Seq data from mouse C2C12 myoblasts (control, non-target and CHERP-siRNA treated) were generated as part of this work and are available from GEO with accession number XXXXXX. Data was mapped using STAR 2.7.9a to the mouse genome assembly (GRCm39), and features counted using FeatureCounts. One of the non-Target replicates (non-target replicate 1) deviated significantly from other samples based on dispersion estimation and principal component analysis (PCA), and was excluded from further analysis. Further analysis was performed in R version 4.1 and Python 2.7. Differential gene expression was evaluated with DEseq2 (Love et al., 2014), differential exon use with DEXseq (Anders et al., 2012b), and differential splicing with rMATS (Shen et al., 2014). DE genes were thresholded as $\text{padj} \leq 0.05$ and $\text{abs}(\log_2 \text{fold change}) \geq 1$. GO term enrichment was performed with InterMine (Smith et al., 2012). A list of MyoG and MyoD targets was obtained from Harmonizome (Rouillard et al., 2016).

Data availability

Raw data used to generate plots are available in the accompanying source data files for each figure. All mRNA-Seq data used in this manuscript are publicly available from GEO under accession numbers GSE63707, GSE107247, GSE143430, GSE194199, GSE184001, and XXXXX. Mass spectrometry data are publicly available from XXXX.

Acknowledgements

We sincerely thank Frank Schnorrer, Aynur Kaya-Çopur, Carla Margulies, Andreas Ladurner, Ulrike Gaul, the Bloomington Drosophila Stock Centre (BDSC) and the Vienna Drosophila Resource Center (VDRC) for providing flies and reagents. MLS is grateful to Andreas Ladurner for generous support. We thank Sandra Esser for excellent technical assistance. Our work relied on the support of core facilities at the LMU Biomedical Center (Martinsried, DE), including the Core Facility Bioimaging, the Protein Analysis Unit (Zentrallabor für Proteinanalytik (ZfP)), the Core Facility Biophysics, the Department of Molecular Biology and the Core Facility Bioinformatics. We acknowledge the Deutsche Forschungsgemeinschaft (MLS, 417912216; PM, 470092532), the Friedrich-Baur-Stiftung (MLS), the Center for Integrated Protein Science Munich (CIPSM) at the Ludwig-Maximilians-University München (MLS), and the Deutsche Gesellschaft für Muskelkranke e.V. (MLS, VT) for financial assistance.

Author Contributions

Investigation (S-YK, VT, PW, KR, OW, AG, LH, PMM, IA, MB, ADRC, RH,),

Validation (S-YK, VT, PW, MB, ADRC,),

Data curation (MLS, IF, TS, PMM),

Formal analysis (S-YK, MLS, VT, LH, PMM, IA, IF, MM, TS, PM, RH, SC,),

Visualization (S-YK, MLS, VT, LH, PMM,),

Methodology (LH, PMM, IA, IF, MM, TS, PM, SC),

Resources (RP, PM, RH, SC),

Software (LH, PMM, IA, TS, SC),

Writing – original draft (S-YK, MLS, VT),

Writing – review & editing (S-YK, VT, PW, KR, OW, AG, LH, PMM, IA, MB, ADRC, IF, MM, TS, RP, PM, BS, RH, SC, MLS),

Conceptualization (MLS, RH, SC, MM, PM, BS),

Supervision (MLS, SC, MM, RH, MM, RP, PM, BS),

Funding acquisition (MLS, SC, RH, MM, RP, PM, BS)

Conflict of interest

The authors declare they have no conflicts of interest.

Figure legends

Figure 1. Scaf6 is necessary for development of the indirect flight muscles.

A) Scheme of the *scaf6* genomic locus at 3L: 17042574-17036859, which encodes a conserved RS-protein with SWAP/U2-Surp (blue), CID (pink) and G-patch (light orange) domains. Available genetic tools include the *scaf6*^{M2M1} and *scaf6*^{12M9} alleles (red arrows), an RNAi construct (dark orange), an endogenous C-terminal FLAG-tag (cyan), and a UAS-*scaf6-RA* line with an N-terminal myc-tag (magenta). **B)** Plot of *scaf6* mRNA expression level in IFM from mRNA-Seq data at eight timepoints (myoblasts, myo; hours after puparium formation, APF; day, d). Normalized counts and significance values from DESeq2 (not significant, ns; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$). **C-D)** UAS-myc-Scaf6 (green) localizes to the nucleus when driven by Mef2-Gal4 in IFM (C; phalloidin stained actin, magenta; DAPI, blue) and elav-Gal4 in the ventral nerve cord (VNC) (D; Elav, magenta). Scale bars = 10 μ m. **E-H)** Confocal time-course of IFM development in control (E, G), mutant *scaf6*^{M2M1/12M9} (*scaf6*^{-/-}, F), and RNAi (*scaf6-IR*, H) flies at 48 h, 72 h, 88 h APF and 1 d adult. Upper panel: Z-projections of myofibers in thorax hemi-sections, scale bar = 100 μ m. Lower panel: single plane images of sarcomere structure, scale bar = 5 μ m. **I)** Plot of flight ability in 1 d adults, quantified as the percent of flies ($n > 50$) that fly normally (white), are weak fliers (cyan) or flightless (blue). Genotypes as labeled. **J)** Quantification of myofiber integrity from 1 d adults; all myofibers attached (white), all detached (mauve), 1-5 detached (pink). **K-L)** Quantification from thorax hemi-sections in E-H of sarcomere length (K) and myofibril width (L) from control (white bars) and *scaf6* mutant or knockdown flies (red bars). Boxplots are shown with Tukey whiskers, with outlier datapoints marked as dots. Significance determined by ANOVA and post-hoc Tukey (not significant, ns; *** = $p < 0.001$).

Figure 2. *scaf6*^{-/-} IFMs display temporal-dependent defects in gene expression and alternative splicing.

A) Plot of the number of significantly DE genes (adjusted p-value (p-adj) ≤ 0.05) evaluated with DESeq2 in *scaf6*^{-/-} versus *w¹¹¹⁸* IFM at 30 h and 72 h APF (upregulated, yellow; downregulated, purple). **B)** Scatter plot of the correlation in gene expression changes between 30 h and 72 h APF in *scaf6*^{-/-} versus *w¹¹¹⁸* IFM. Only genes with a p-adj ≤ 0.05 at either 30 h or 72 h APF are shown. Grey shading marks \log_2 (fold change, FC) $\geq |1.2|$ and p-adj ≤ 0.05 . Sarcomere genes, red dots; Pearson's correlation, blue line. **C)** Percent bar chart of the overlap between 30 h and 72 h APF in all expressed genes (normalized counts > 100), all DESeq2 DE genes (p-adj ≤ 0.05) and all genes with DEXSeq DE exons (p-value ≤ 0.05) (unique to 30 h, yellow; both 30 h and 72 h, green; unique to 72 h, blue). **D)** Violin and accompanying boxplot of the *scaf6*^{-/-} versus *w¹¹¹⁸* IFM \log_2 FC values at 30 h APF for all genes that are normally significantly upregulated (orange) or downregulated (purple) in wildtype muscle between 24 h and 30 h APF. **E)** Heatmap of enrichment p-value for select cellular component (CC) and biological process (BP) GO terms. Enrichments were calculated for significantly differentially expressed genes and exons, as well as for novel alternative splice (AS) and intron retention (IR) events, in *scaf6*^{-/-} IFM at 30 h and 72 h APF. **F)** Plot of the total number of significantly DE exons (p-value (p-val) ≤ 0.05) identified with DEXSeq in *scaf6*^{-/-} versus *w¹¹¹⁸* IFM at 30 h and 72 h APF (upregulated, yellow; downregulated, purple). **G)** Scatter plot of the correlation in change in exon use between 30 h and 72 h APF in *scaf6*^{-/-} versus *w¹¹¹⁸* IFM. All exons with a p-val ≤ 0.05 at either 30 h or 72 h APF are shown. Grey shading marks \log_2 (FC) $\geq |1.2|$ and p-val ≤ 0.05 . Sarcomere gene exons, red dots; Pearson's correlation, blue line. **H)** Plot of the number of novel AS events detected by fortuna and IR events

identified with a novel script that are observed uniquely in w^{1118} (white) or $scaf6^{-/-}$ (red) IFM at 72 h APF. **I)** Plot of novel AS event type for events detected selectively in $scaf6^{-/-}$ IFM at 30 h or 72 h APF (alternative acceptor (AA), purple; alternative donor (AD), light purple; alternative pair (AP), blue; exon skip (ES), turquoise; intron in exon (IE), green; intron retention (IR), light green; other (XX), yellow). **J-M)** RT-PCR verification of splicing changes in *Zasp66* (J), *Tm1* (K), *sls* (L) and *Unc-89* (M). Top: Schematic of select splice isoforms and mRNA-Seq read-count tracks for w^{1118} (black) and $scaf6^{-/-}$ (red) IFM at 72 h APF. Primer locations are marked by black arrows and expected RT-PCR product length in base pairs (bp) is labeled to the right. Exons (Ex) are numbered according to FB2022_01. Bottom: Representative RT-PCR gel image. Colored triangles are coordinated with schematic above to denote bands corresponding to annotated isoforms preferentially expressed in IFM (dark blue) or other tissues (orange), as well as novel isoforms resulting from alternate junction use (green) and IR events (pink).

Figure 3. Protein expression defects in $scaf6^{-/-}$ IFM result in aberrant contractility and myofiber detachment.

A) Volcano plot of protein group differential expression from LC-MS analysis of $scaf6^{-/-}$ versus control w^{1118} IFM. Significance calculated in Perseus with FDR = 0.05 (significant, blue; non-significant, grey). Sarcomere proteins are identified in red. **B)** Plot of select GO terms from the Cellular Component (CC) and Biological Process (BP) ontologies that are significantly enriched in the downregulated (Difference < 0) or upregulated (Difference > 0) protein sets. **C)** Single-plane confocal images of GFP-tagged reporters for sarcomere proteins *Zasp66*, *Sls* and *Unc-89* in 1 d adult IFM. Control (top, *Mef2-Gal4, reporter > w^{1118}*) and *scaf6* knockdown (bottom, *Mef2-Gal4, reporter > scaf6-IR*) samples were imaged with the same settings and pseudo-colored to reflect signal intensity. Scale bar =

5 μm . **D)** Quantification of fold change in fluorescence signal intensity for all tested GFP-tagged reporters in *scaf6-IR* versus control IFM. Significance determined by ANOVA and post-hoc Tukey (not significant, ns; *** = $p < 0.001$). **E)** Venn diagram (left) of the overlap in sarcomere proteins that are differentially expressed (magenta), have changes in exon use (blue) and have intron retention events (green). A matrix shows select GO term associations of the 10 overlapped genes (right), as well as which genes have reported hypercontraction phenotypes. **F-G)** Confocal z-projections of IFM myofiber structure in 1 d adult *scaf6^{-/-}* (F) and double mutant *Mhc¹⁰; scaf6^{-/-}* (G) flies. Scale bar = 100 μm . **H)** Quantification from F and G of the percent of flies with all myofibers attached (white) or detached (mauve), or only some myofibers detached (pink). Significance determined by Chi-squared test ($\chi^2=10.39$, $df=2$, $p\text{-value}=0.0055$). **I-J)** Still images from live confocal recording of spontaneous contraction (twitch) events in 48 h APF IFM from control *w¹¹¹⁸* (I, I') and mutant *scaf6^{-/-}* (J, J') flies. Resting position ($t=0$) is pseudo-colored green, while the extended twitch position ($t=0.64$ seconds) is pseudo-colored magenta, such that a lack of movement will appear white in the overlay image (I', J'). Dashed lines labels the tip of the myofiber. Scale bar = 50 μm . **K-L)** Quantification of twitch events at 36 h APF (K) and 48 h APF (L). Upper bar represents the number of pupae with (white) or without (red) an IFM twitch event ($N = \text{total number of pupae}$). Lower panel shows the average frequency of twitches in single myofiber in 5 minutes (single twitch instance, purple; multi-twitch instance, green). Significance determined by ANOVA and post-hoc Tukey (not significant, ns; * = $p < 0.05$; *** = $p < 0.001$).

Figure 4. Scaf6 is required cell intrinsically for motor neuron axon branching and maintenance of sarcomere structure.

A) Heatmap of performance ability for behaviors including grooming, righting reflex, climbing, flight and eclosion, as well as adult survival, in mutant and tissue-specific RNAi knockdown conditions. Genotypes as listed. Additional behavior data available in Fig. S4. **B-J)** Confocal images of IFM motor neuron (MN) axon branches (HRP, grey) and sarcomere structure (phalloidin labeled actin, grey) in control (B, E, H), mutant (C, D), and tissue-specific *scaf6-IR* (F, G, I, J) flies at 72 h APF. Muscle-specific *Mef2-Gal4 > scaf6-IR* results in loss of sarcomere structure without affecting axon branching (E, F, G), while neuronal-specific *Elav-Gal4 > scaf6-IR* results in loss of axon branches without affecting sarcomere structure (H, I, J). Scale bars = 10 mm (C, I), 5 mm (D, J). **K)** Schematic of MN higher-order axon branching (secondary, black; tertiary, blue; quaternary, red). **L)** Quantification of MN axon branches at 72 h APF in mutant and tissue-specific knockdown conditions. Genotypes as labeled. Significance from Student's T-test (not significant, ns; ** = $p < 0.01$; *** = $p < 0.001$). **M)** Heatmap of behavioral performance ability and survival in control and tissue-specific rescue conditions. **N-P)** Confocal projections of IFM myofiber structure (N, O, P) and single-plane images of myofibril structure (N', O', P') in 1 d adult hemi-thoraxes from *scaf6^{-/-}* (N, N') and muscle (O, O') or neuronal (P, P') specific rescue. Note the attached myofibers and periodic sarcomere pattern in the muscle-specific (*Mef2-Gal4 > UAS-Myc-Scaf6; scaf6^{-/-}*) but not the neuronal specific (*Elav-Gal4 > UAS-Myc-Scaf6; scaf6^{-/-}*) rescue. Scale bars = 100 mm (P), 5 mm (P').

Figure 5. RNA sequencing reveals differences in gene expression and alternative splice site usage in *Scaf6*^{-/-} IFM, leg and brain

A) Heatmap of genes expressed in IFM, leg and brain in *w¹¹¹⁸* and *Scaf6*^{-/-} flies. Genes are clustered after z-scores. **B)** PCA analysis of differential usage of 3' splice sites (acceptor) in *w¹¹¹⁸* and *Scaf6*^{-/-} IFM, leg and brain. Wildtype and mutant samples clearly group together on PC1 and PC2, constituting as much as 70% of variance as shown in the PCA scree plot. **C)** PCA analysis of differential usage of 5' splice sites (donor) in *w¹¹¹⁸* and *Scaf6*^{-/-} IFM, leg and brain. Wildtype and mutant samples group together on PC1 and PC2, constituting as much as 65% of variance as shown in the PCA scree plot.

Figure 6. *Scaf6* regulates myoblast proliferation *in vivo*

A) Quantification percentage of myofibers number in temporal conditional knockdown of *scaf6* with different Gal4 driver. **B)** Turkey box plot and whiskers of DESeq2 log₂FC changes data for *scaf6*^{-/-} versus WT IFM at 72h APF in selective GO term category. **C1-E1)** Images of myoblast visualization with GFP staining in 3rd instar larval wing disc in control (C1), knockdown of *scaf6* (D1) or overexpression of *scaf6* (E1). **C2-E2)** Images of myoblast marker (PH3) staining in 3rd instar larval wing disc in control (C2), knockdown of *scaf6* (D2) or overexpression of *scaf6* (E2). **C3-E3)** Images of merged channel with DAPI (blue), myoblast (green), and mitosis marker (magenta) in 3rd instar larval wing disc in control (C3), knockdown of *scaf6* (D3) or overexpression of *scaf6* (E3). Scale bar represents 50mm. **C4-E4)** X-Z projection of C3-E3. **F)** Quantification of myoblast number in 3rd instar larval wing disc in control, knockdown of *scaf6* or overexpression of *scaf6* in muscle. Student's T test comparison p-value<0.01**, p-value <0.001***. **G)** Quantification of PH3 positive cells within myoblast pool in 3rd instar larval wing disc in control, knockdown of *scaf6* or overexpression of *scaf6* in muscle.

Student's T test comparison p -value $<0.05^*$. **H1-J1**) Images of apoptosis marker (Dcp1) staining in 3rd instar larval wing disc in control (C2), knockdown of *scaf6* (D2) or overexpression of *scaf6* (E2). **H2-J2**) Images of merged channel with DAPI (blue), myoblast (green), and apoptosis marker (magenta) in 3rd instar larval wing disc in control (C3), knockdown of *scaf6* (D3) or overexpression of *scaf6* (E3). Scale bar represents 50mm. **K**) Quantification of apoptosis positive cells within in 3rd instar larval wing disc in control, knockdown of *scaf6* or overexpression of *scaf6* in muscle. Student's T test comparison n.s.= no significance.

Figure 7. *CHERP* regulates proliferation and differentiation in mouse C2C12 myoblasts

A-D) Untreated and no-target siRNA treated C2C12 myoblasts are proliferative while *CHERP* knockdown cells proliferate less (**A** upper panel Ki67 in magenta, quantified in **C**). Accordingly, cell numbers are reduced by ~40% after 24 hours knockdown compared to control and no-target siRNA treated cells as shown in **B**. In contrast, *CHERP* knockdown does not lead to impaired double-strand break repair as assessed through TUNEL staining (**A** lower panel and quantified in **D**). **E**) Volcano plot of differentially expressed genes in *CHERP* knockdown vs. control C2C12 myoblasts (blue: p -value < 0.05 and $\log_2FC > |1|$). The expression of sarcomeric protein encoding genes is upregulated in *CHERP* knockdown cells as marked in red. The differentiation marker *Myog* is the most significantly upregulated gene. **F**) GO enrichment analysis of differentially expressed genes in *CHERP* knockdown cells, divided in up- and downregulated terms. **G-H**) Immunofluorescent staining and quantification of *Myog* upregulation in *CHERP* knockdown cells compared to untreated and no-target siRNA treated C2C12 cells. **I**) Chord plot and heatmap of proliferation and differentiation associated genes. Genes related to proliferation are downregulated, while genes related

to differentiation are upregulated in CHERP knockdown myoblasts. **J**) Violinplot of upregulated pathways: differentiation, Myog targets, muscle contraction and sarcomere organization.

Figure 8. CHERP regulates muscle specific genes through target-specific splicing but not through differential protein-protein interaction

A) Volcanoplot of genes differentially spliced in CHERP knockdown C2C12 myoblasts as analysed using DEXSeq (blue: $p\text{-value} < 0.05$ and $\log_2\text{FC} > |1|$, red: sarcomeric proteins). **B-C)** MAJIQ identifies splicing events in untreated and CHERP knockdown cells. Overlap of all detected splicing events between untreated and CHERP knockdown C2C12 (**B**). Distribution of alternative splice events ($\psi\text{-value} > |0.1|$) in control and CHERP knockdown cells. IR = intron retention, EI = exon inclusion, Alt3 = alternative 3' splice site, Alt5 = alternative 5' splice site (**C**). **D)** GO term enrichment analysis of genes alternatively spliced in CHERP knockdown condition. **E)** Immunoprecipitation followed by mass spec (IP-MS) of CHERP in non-muscle cells (HEK293T), mouse C2C12 myoblasts (mMB) and myotubes (mMT) and human primary myotubes from two healthy individuals (hMT1, hMT2). **F)** CHERP interacts with components of the U2 spliceosomal machinery in muscle and non-muscle cells alike. Namely, with 3 U2 accessory proteins (in blue), 3 proteins recruited at A complex (in yellow) and 6 U2 complex core proteins (in green). **G-J)** CHERP is misregulated in disease. In colorectal cancer and myotonic dystrophy type I (DM1), CHERP is transcriptionally upregulated as tested via microarray and qPCR, respectively (**G** and **H**). CHERP upregulation on the RNA level in DM1 does not lead to an upregulation on the protein level in DM1 (Western Blot), but to a downregulation, suggesting a negative feedback loop (**I**). CHERP is involved in pathways that are also affected in DM1 (**J**).

SUPPLEMENTARY INFORMATION

Figure S1. *Scaf6* is expressed in *Drosophila* muscle and neurons.

A) Top: Scheme of *scaf6* mRNA isoforms. Note that RA is the only full-length isoform. Black arrows denote locations of RT-PCR primers. Bottom: mRNA-seq counts over the *scaf6* locus at 30 h APF (light blue) and 72 h APF (dark blue). **B)** Normalized counts of *scaf6* expression in mRNA-Seq data from IFM, brain, leg and ovary samples. Significance values from pairwise DESeq2 analysis (not significant, ns; *** = $p_{adj} < 0.001$). **C-E)** *Scaf6*-RA-FLAG (greyscale, C, D, E; green, C', D', E') is nuclear localized (white arrows) in single-plane confocal images of IFM (C-D; DAPI, blue; phalloidin-stained F-actin, magenta) or ventral nerve cord (VNC) (E, anti-Elav, magenta) from 1 d adult flies. Scale bars=5mm. Note that the background staining of T-tubules in IFM with anti-FLAG is also present in control *w¹¹¹⁸* flies, while the nuclear signal is only observed in the *Scaf6*-RA-FLAG genotype. **F-H)** Overexpressed *Scaf6*-PA (greyscale, F, G, H; green, F', G', H') is nuclear localized in brain (*Elav-Gal4 > UAS-Myc-Scaf6*) (F-G) and in leg muscle (*Mef2-Gal4 > UAS-Myc-Scaf6*) (H). Scale bars = 50 mm (F') and 5 mm (G', H'). **I)** RT-PCR verification of *scaf6* isoform expression in dissected IFM (top), brain (middle) and leg (bottom) at 48 and 72 h APF and in 1-day old adults. Bands corresponding to specific *scaf6* isoforms as labeled based on expected size labeled on the left (bp = DNA ladder). **J)** \log_2 Fold Change values reflecting differential expression (DE) of *scaf6* between control *w¹¹¹⁸* and mutant *scaf6^{-/-}* flies in mRNA-Seq data from 30 h APF IFM and 72 h APF IFM, leg and brain samples. Significance is from DESeq2 (not significant, ns; *** = $p_{adj} < 0.001$). **K-L)** RT-qPCR verification of *scaf6* mRNA expression levels in IFM and brain. Significance was calculated by ANOVA with post-hoc Tukey test (not significant, ns; * = $p < 0.05$; *** = $p < 0.001$). **M)** Heatmap summarizing lethality timepoint (early adult or pharate, light yellow; late pupal, gold; early pupal, orange; pre-pupal, red) for flies from all combinatorial crosses between *scaf6^{12M9}*, *scaf6^{M2M1}* and the *Df(3L)ED4674* deficiency line covering the *scaf6* locus. **N-Q)** Confocal z-projection images of IFM myofibers (top, scale bar = 100 mm) and single-plane images of fibrillar IFM (middle) and tubular leg (bottom) muscle sarcomere structure (scale bars = 5 mm) in *scaf6^{12M9/Df(3L)ED4674}* (N), *scaf6^{M2M1/Df(3L)ED4674}* (O) and *scaf6^{M2M1/M2M1}* (P) at 90 h APF, and in *UAS-Dicer2*, *Mef2-Gal4 > scaf6-IR* 1 d adult flies (Q).

Figure S2. Temporal dynamics of gene expression and alternative splicing in *scaf6*^{-/-} IFMs.

A-B) Volcano plots from DESeq2 analysis of differential gene expression in mRNA-Seq data from *scaf6*^{-/-} versus *w¹¹¹⁸* IFM at 30 h APF (A) and 72 h APF (B). Genes with a log₂(FC) ≥ |1.2| and p-adj ≤ 0.05 (grey boxes) are colored blue. Sarcomere genes are labeled in red. **C)** Boxplot of change in gene expression for mitochondrial genes, genes with an RNAi phenotype in muscle (muscle pheno), genes encoding sarcomere proteins (SPs) or genes annotated with the GO term “synapse.” *scaf6*^{-/-} versus *w¹¹¹⁸* log₂(FC) values are shown for significantly DE genes (p-adj ≤ 0.05) at 30 h (yellow) and 72 h (blue) APF. Red line marks boxplot median. **D)** Violin plot with boxplot overlay of *scaf6*^{-/-} versus *w¹¹¹⁸* IFM log₂FC values at 30 h APF (left) or 72 h APF (right) for all genes that are normally significantly upregulated (orange) or downregulated (purple) in wildtype muscle between 24 h and 30 h APF or between 30 h and 72 h APF. **E)** Dot plot of log₂(FC) values comparing 72 h versus 30 h APF IFM in *w¹¹¹⁸* (black) or *scaf6*^{-/-} (orange). Genes with a significant temporal change in expression in *w¹¹¹⁸* are shown in ascending order along the X-axis. The majority of temporally regulated genes show similar changes in expression in *scaf6*^{-/-} IFM. **F)** Percent bar plot of the overlap in DE genes and genes with DE exons in *scaf6*^{-/-} versus *w¹¹¹⁸* IFM at 30 h and 72 h APF (only gene level, yellow; only exon level, purple; both gene and exon level, teal). **G-H)** Volcano plots from DEXSeq analysis of differential exon use in mRNA-Seq data from *scaf6*^{-/-} versus *w¹¹¹⁸* IFM at 30 h APF (G) and 72 h APF (H). Exons with a log₂(FC) ≥ |1.2| and p-val ≤ 0.05 (grey boxes) are colored blue. Sarcomere gene exons are labeled in red. **I)** Boxplot of change in exon use for mitochondrial, muscle phenotype, sarcomere protein and synapse genes. *scaf6*^{-/-} versus *w¹¹¹⁸* log₂(FC) values are shown for significantly DE exons (p-val ≤ 0.05) at 30 h (yellow) and 72 h (blue) APF. Red line marks boxplot median. **J)** Plot of the number of reads supporting observed novel alternative splice (AS) events as detected by fortuna. More reads are detected in *scaf6*^{-/-} samples at 30 h (orange) and 72 h (red) APF than in *w¹¹¹⁸* samples (blue, light blue). **K)** Plot of the number of novel AS events detected by fortuna and IR events identified with a novel script that are observed uniquely in *w¹¹¹⁸* (white) or *scaf6*^{-/-} (red) IFM at 30 h APF. **L)** Plot of the number of novel AS events from fortuna observed in mitochondrial, muscle phenotype, sarcomere protein and synapse genes in *w¹¹¹⁸* (blue) or *scaf6*^{-/-} (red) at 30 h APF (circle) and 72 h APF (triangle).

Figure S3. Loss of Scaf6 causes downregulation of sarcomere protein expression and mitochondrial defects.

A) Venn diagram of the overlap between all genes with a significant change in exon use (blue, DEXSeq, p -value ≤ 0.05), that contain novel splice events (purple), that are significantly regulated on the gene level (red, DESeq2, p -adjusted ≤ 0.05), or that are significantly regulated at the protein level (yellow, LC-MS, FDR = 0.05). **B)** Correlation dot plot for all significantly differentially expressed proteins (LC-MS, FDR = 0.05) compared to and their corresponding fold change in mRNA expression (mRNA-Seq, DESeq2) between *scaf6*^{-/-} and control *w¹¹¹⁸* IFM at 72 h APF. mRNAs with significant \log_2 (fold change) values (purple, DESeq2, p -adjusted ≤ 0.05) and genes with novel splice events or differential exon use (gold, DEXSeq, p -value ≤ 0.05) are highlighted. The regression line (red) plus 95% confidence interval (light blue) are illustrated, corresponding to a Pearson's correlation coefficient of 0.3640722 (Spearman = 0.4417035). **C-G)** Expression of GFP-tagged reporter genes under endogenous enhancer control for Mhc (C, C', C''), F, F', F''), Act88F (D, D', D''), Strn-Mlck-Isoform R (E, E', E'') and Zasp52 (G, G', G'') in control (*Mef2-Gal4, reporter* x *w¹¹¹⁸*) and *scaf6* knockdown (*Mef2-Gal4, reporter* > *scaf6-IR*) IFM from 1 day adult flies. Single-plane confocal images of control (C, D, E, F, G) and *scaf6* knockdown (C', D', E', F', G') samples were imaged with the same settings and pseudo-colored to reflect signal intensity. Scale bar = 5 μ m. Representative line scan of GFP reporter intensity in arbitrary fluorescence units (a.u.) shows localization of reporters across a single sarcomere from z-disc to z-disc in control (black line) and *scaf6-IR* (red line) (C'', D'', E'', F'', G''). **H-M)** Single plane confocal images of Mito-GFP reporter expression in IFMs of control (*Mef2-Gal4* x *w¹¹¹⁸*, H, I, J) and knockdown (*Mef2-Gal4*>*scaf6-IR*, K, L, M) flies at 48 h APF, 72 h APF, and 1 d adult. Dashed lines in merged channel images (H', I', J', K', L', M') outline a single mitochondrion (Mito-GFP, green; phalloidin-stained actin, magenta). Scale bar = 5 μ m. **N-O)** Quantification of single mitochondrion size (N) measured in square microns (μ m²) and mitochondrial load (O) measured as the percent GFP positive area in control (white) and *scaf6* knockdown (red) IFM. Significance from Student's T-test (not significant, ns; *** = $p < 0.001$).

Figure S4. Scaf6 function in muscle and neurons is required for multiple behaviors.

A) Analysis of larval crawling ability. 3rd instar larvae (left, yellow dots) were tracked for 30 seconds using Trackmate (right). **B)** Quantification of crawling speed (mobility in millimeters per second) from (A) for controls (white boxes) and *scaf6*^{-/-} (red box).

Significance determined by ANOVA and post-hoc Tukey (not significant, ns). **C)** Quantification of pupal eclosion ability in controls (white), mutant (red), tissue-specific knockdown (orange) and tissue-specific rescue (yellow) conditions. Plot shows percent of eclosed pupae (N = 61 - 468 per genotype). Significance from ANOVA with post-hoc Tukey test (not significant, ns; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$). **D)** A representative image of pharate lethality in *scaf6*^{-/-}. Pupal development is complete, but flies are too weak to fully eclose. **E)** Quantification of flight ability, recorded as the percentage of flies that display normal flight (white), are weak fliers (cyan) or are flightless (blue). Genotypes as labeled. **F)** Quantification of climbing ability, presented as the percent of flies able to climb 5 centimeters in 5 seconds. Genotypes, labeling and significance as in (C). **G-H)** Assay of flight ability (G) and climbing ability (H) for controls and tissue-specific knockdown of *scaf6* in glia (Repo-Gal4) and fat body (ppl-Gal4). Climbing scored as in (F); significance from Student's T-test (not significant, ns). **I)** Quantification of the time in seconds necessary for a fly to right itself after landing on its back. Genotypes, labeling and significance as in (C). **J)** Representative images of fly grooming, showing how well a wildtype (*w¹¹¹⁸*) and *scaf6*^{-/-} mutant fly are able to remove a dusting of brilliant yellow dye in 30 minutes. **K)** Quantification of grooming ability, calculated as the ratio of absorbance at 397nm after 30 minutes grooming to absorbance measured directly after dusting ($t = 30 / t = 0$). Genotypes, labeling and significance as in (C).

Figure 1

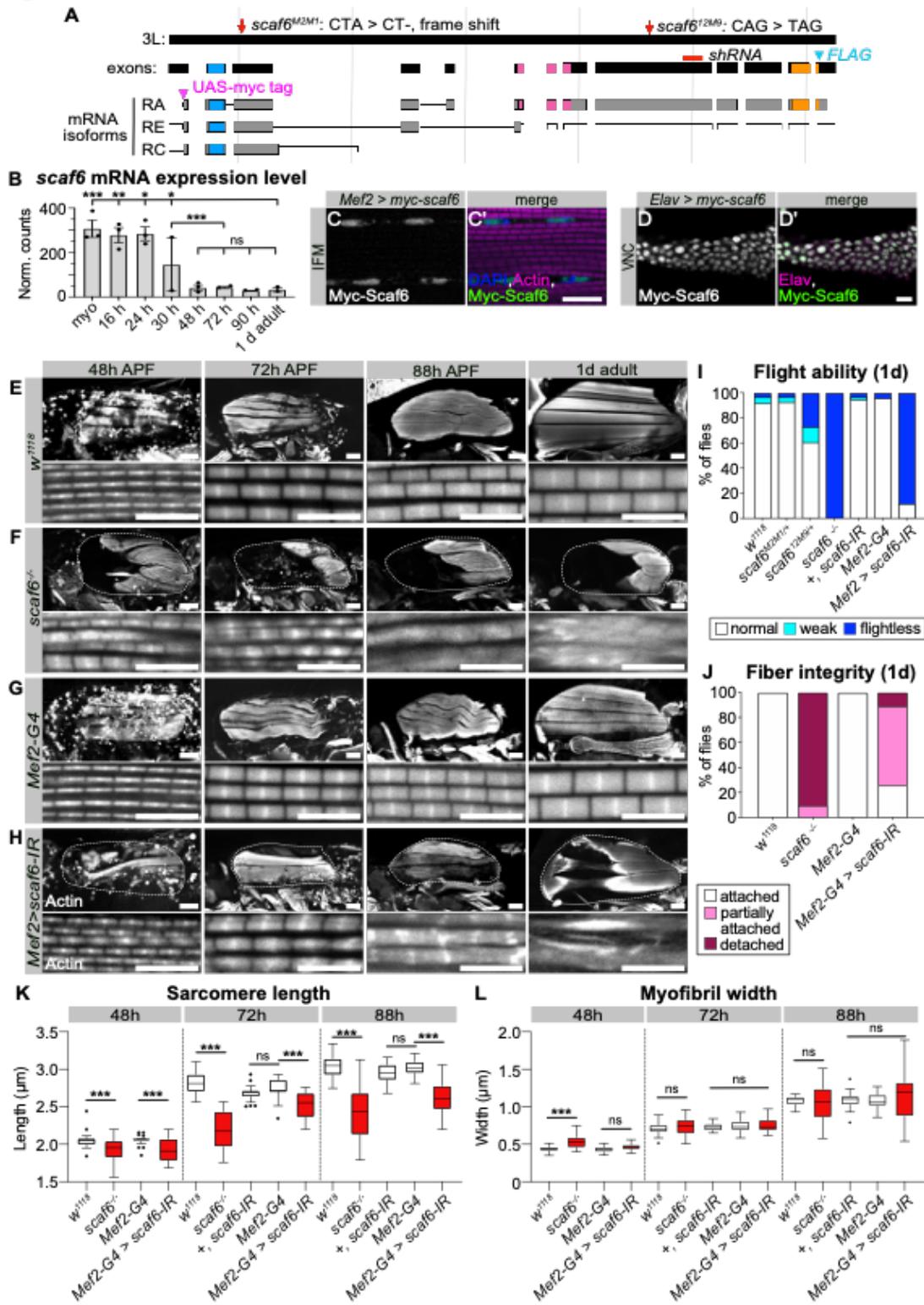


Figure 2

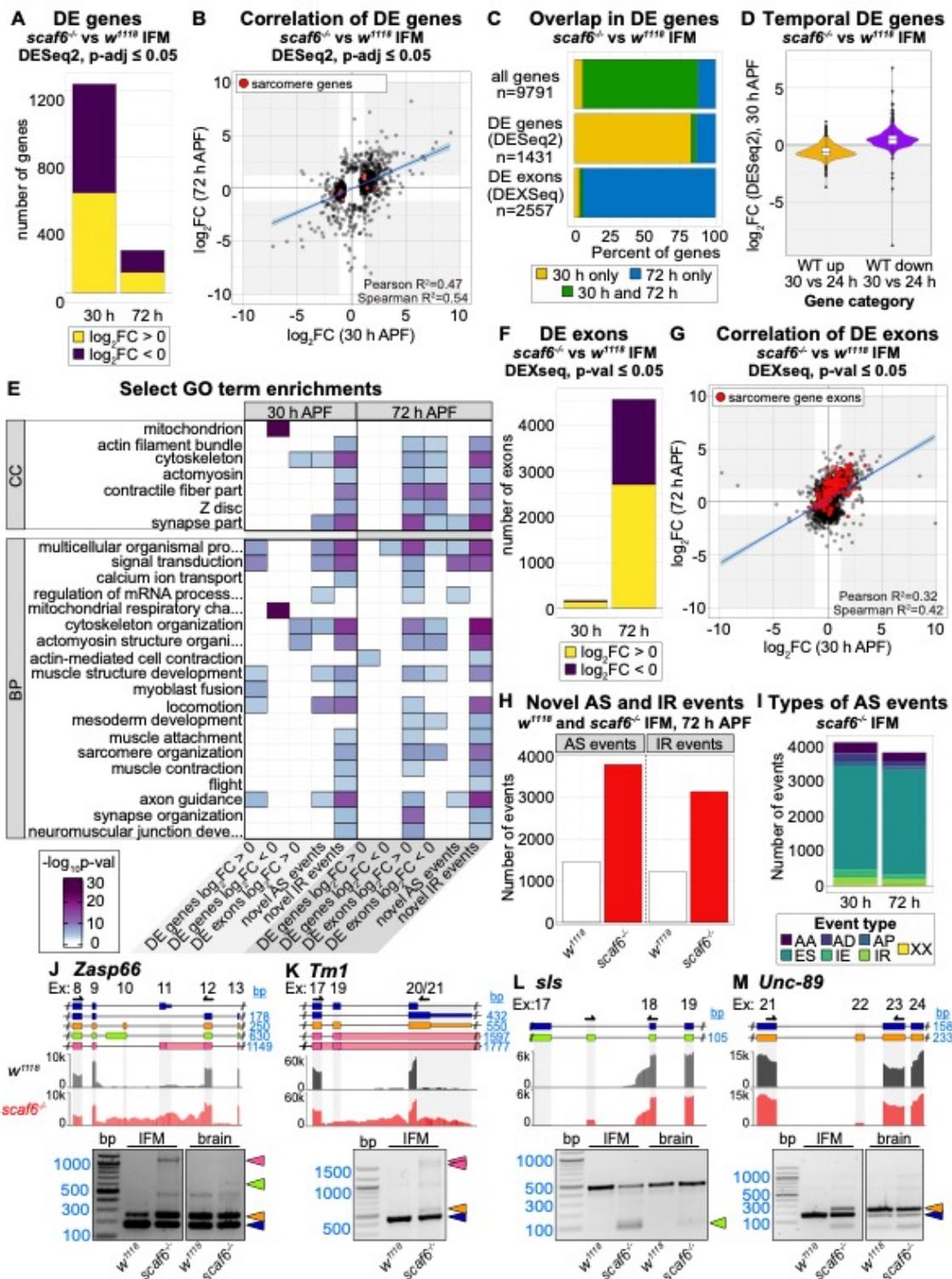
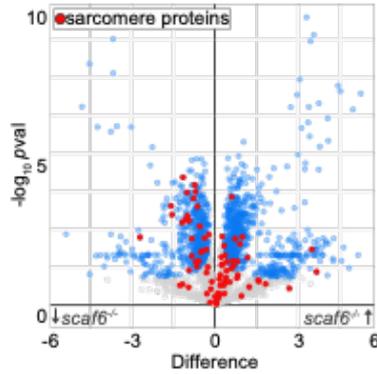
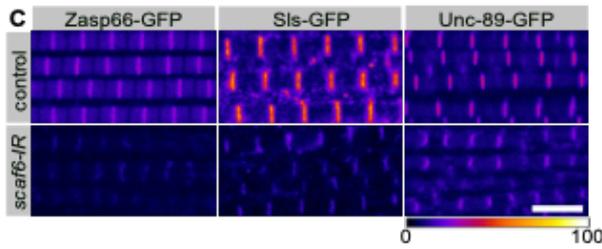
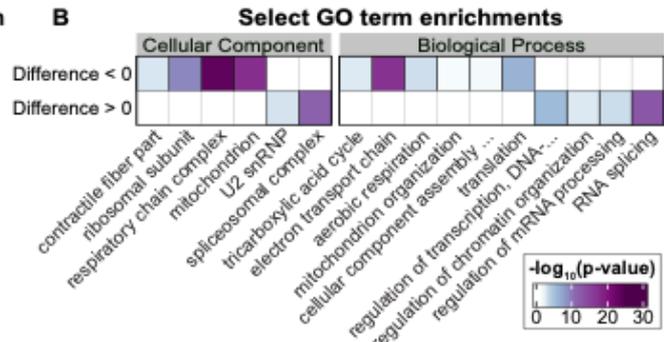


Figure 3

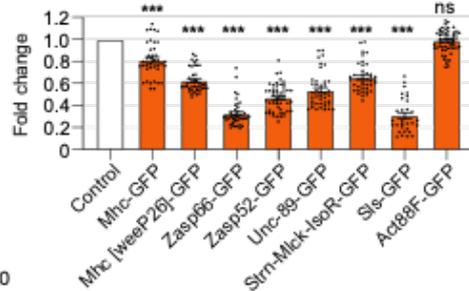
A Volcano plot of *scaf6*^{-/-} vs *w*¹¹¹⁸ at 72h
LC-MS, FDR=0.05



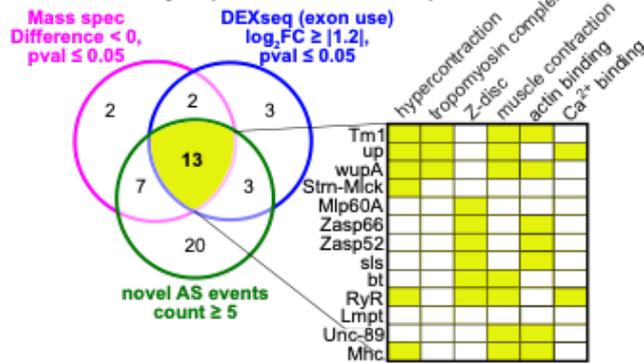
B



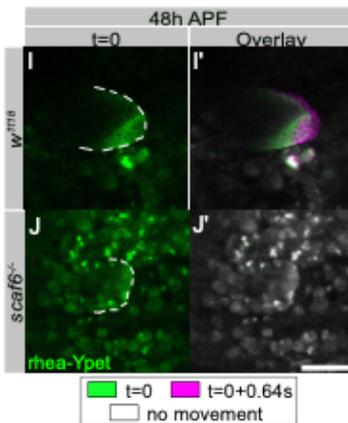
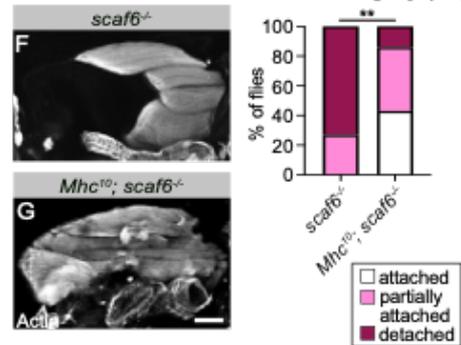
D Sarcomere Protein Expression Level



E Differentially expressed sarcomere proteins



H Fiber integrity (1d)



Twitching Events

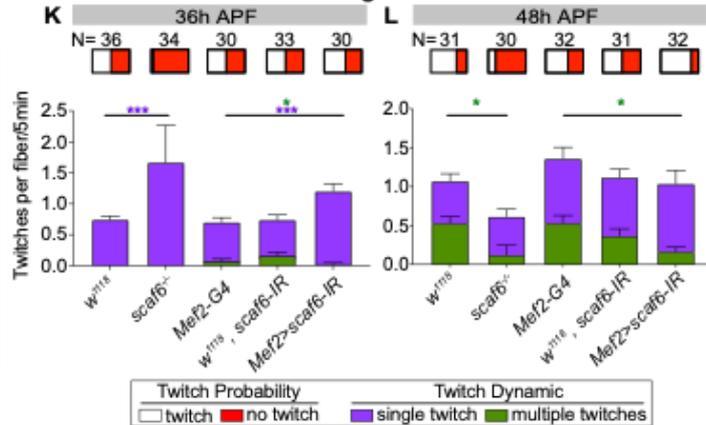


Figure 4

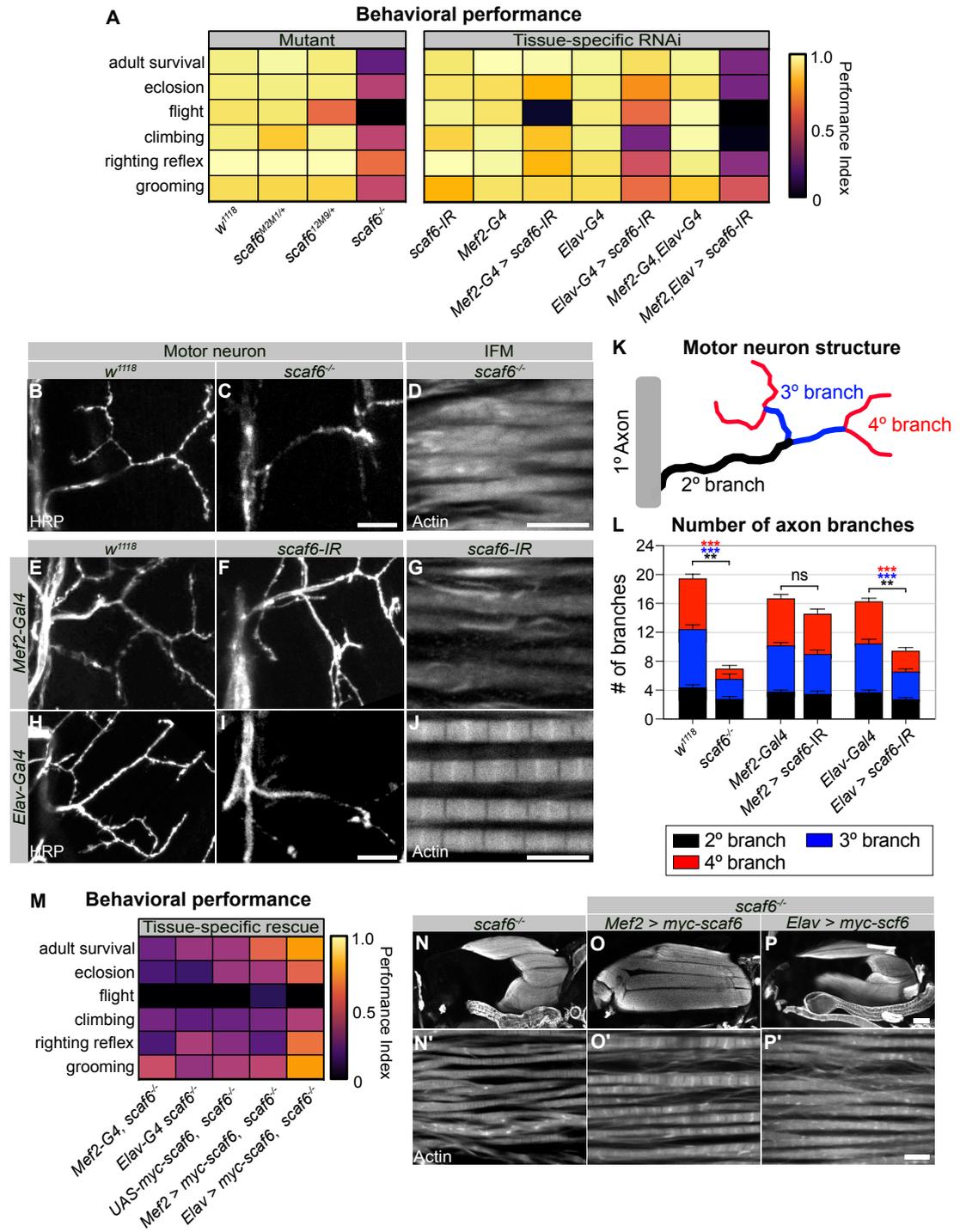


Figure 5

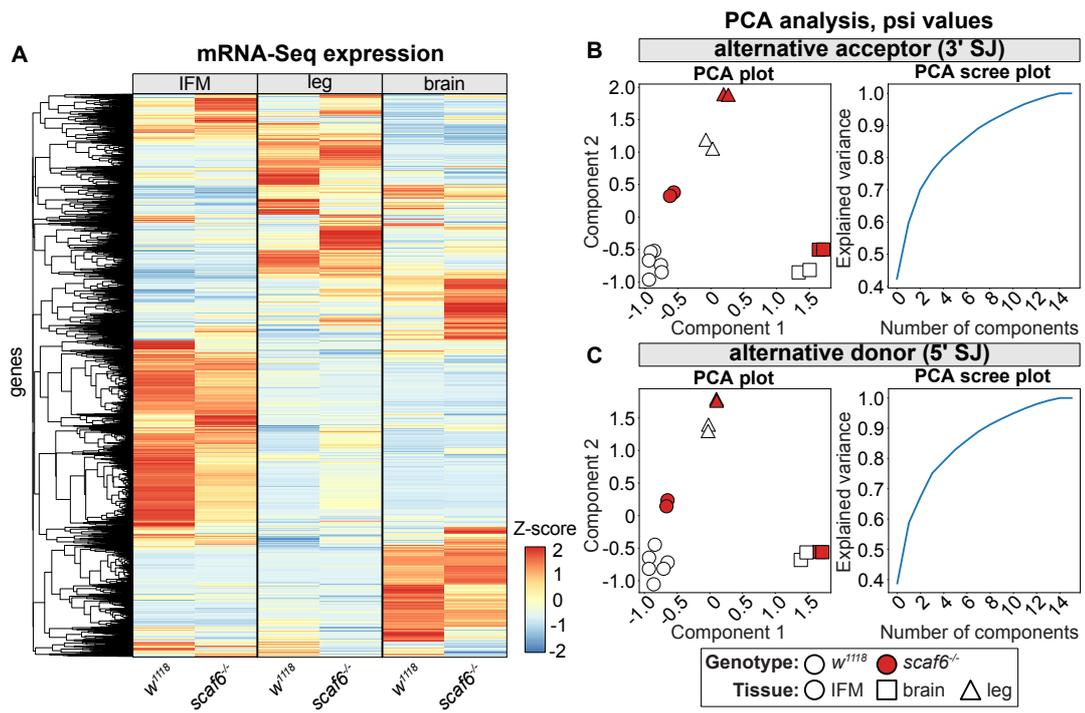


Figure 6

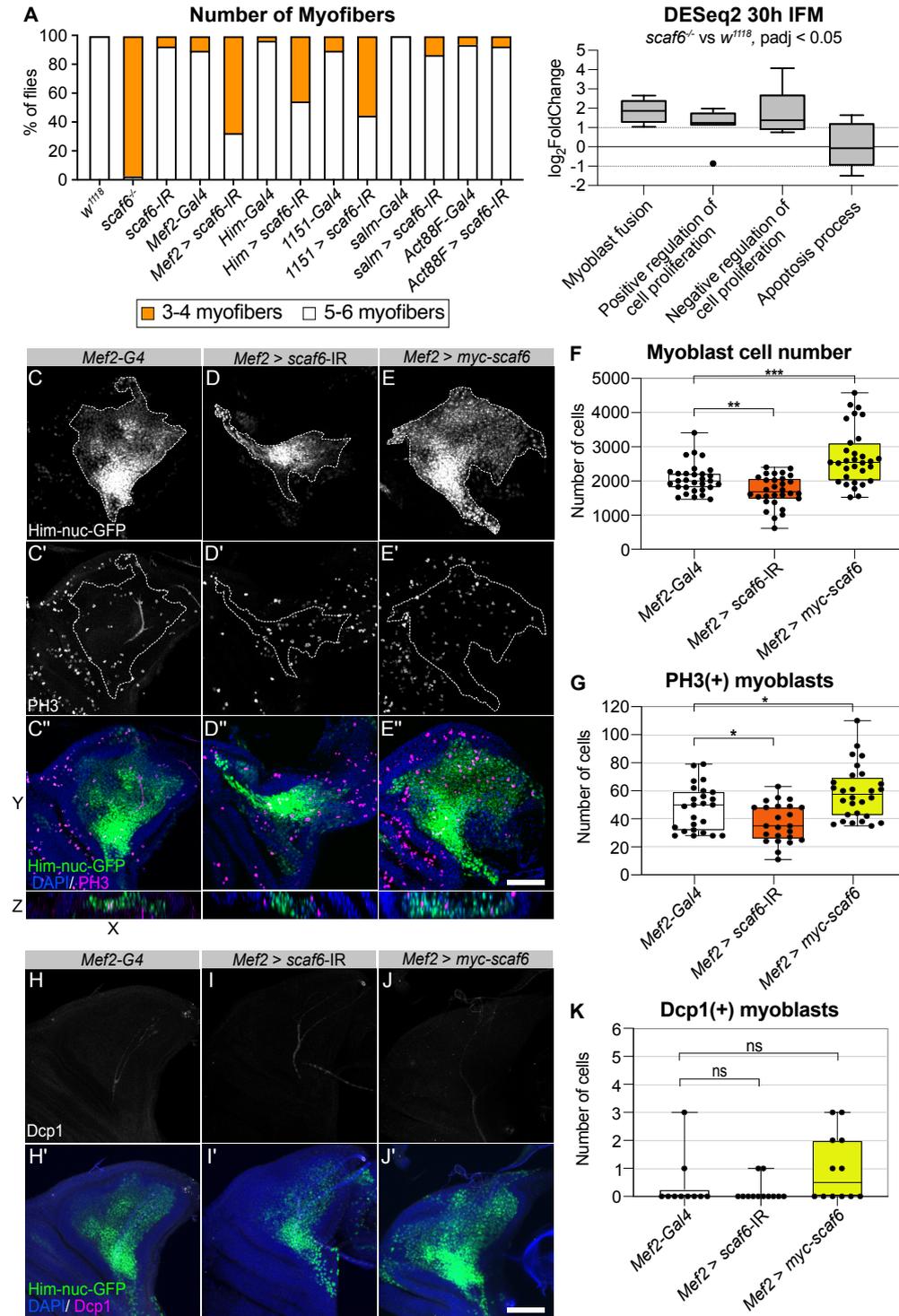
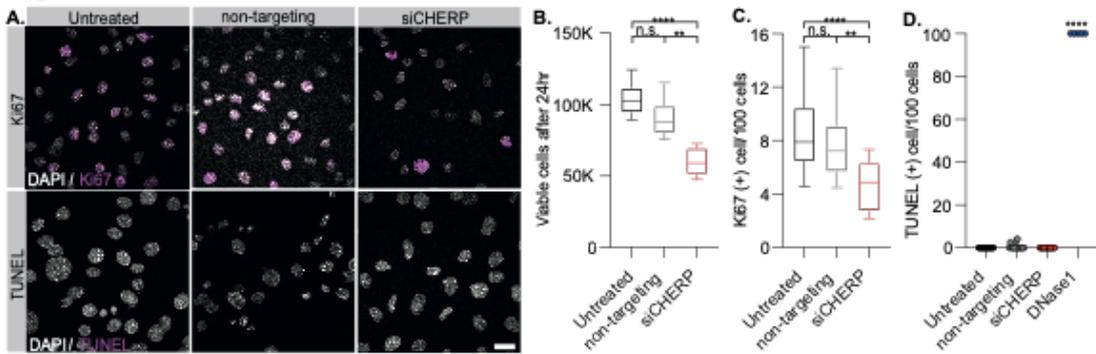
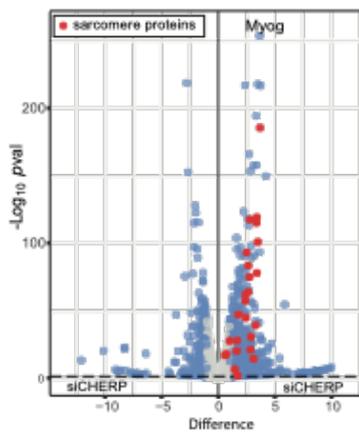


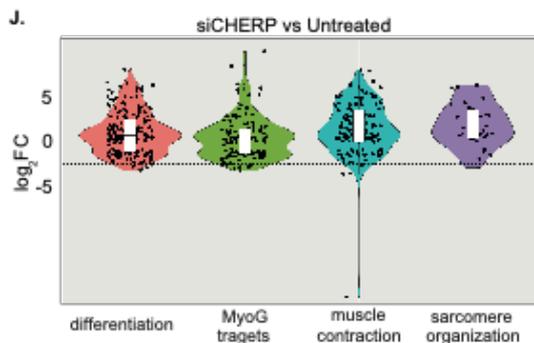
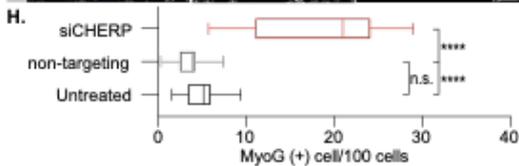
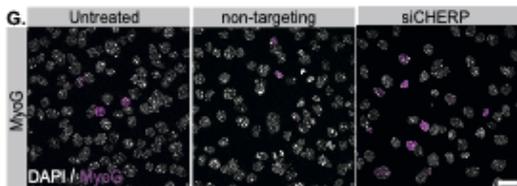
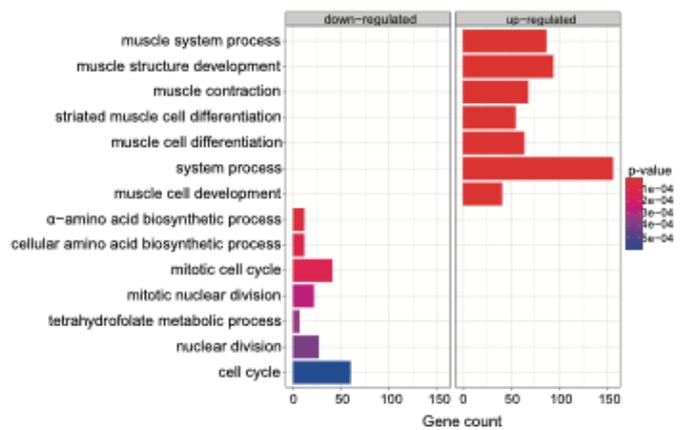
Fig. 7



E. Volcano plot of siCHERP vs control



F. GO enrichment in selective term



I. Cell cycle and differentiation related genes

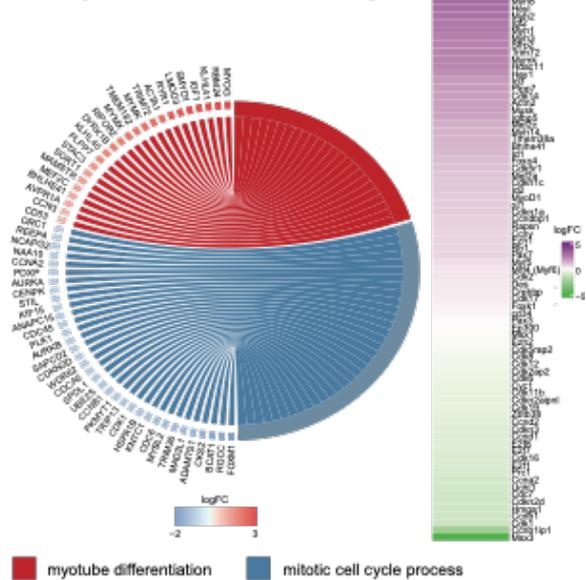
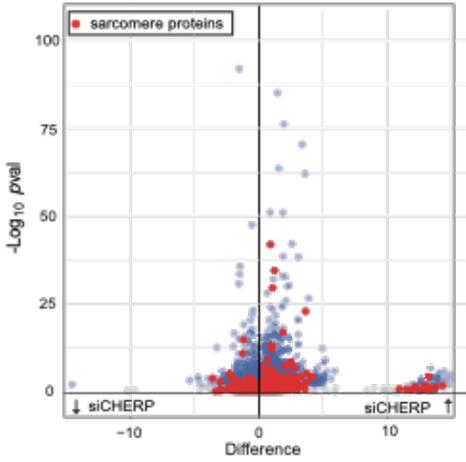
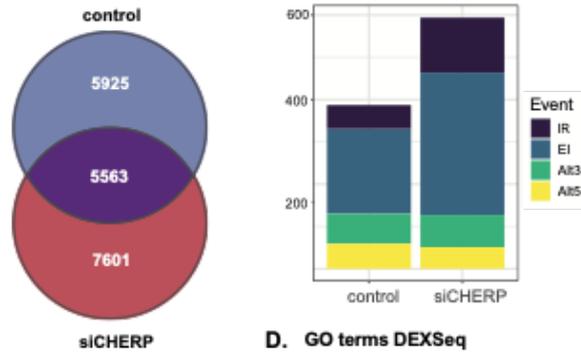


Fig. 8

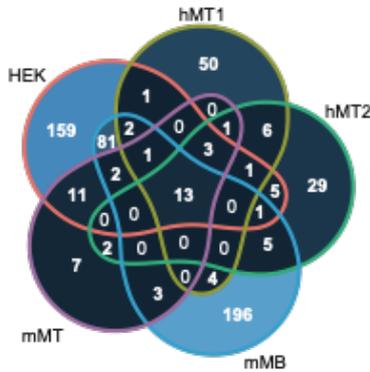
A. Exon usage in siCHERP cells



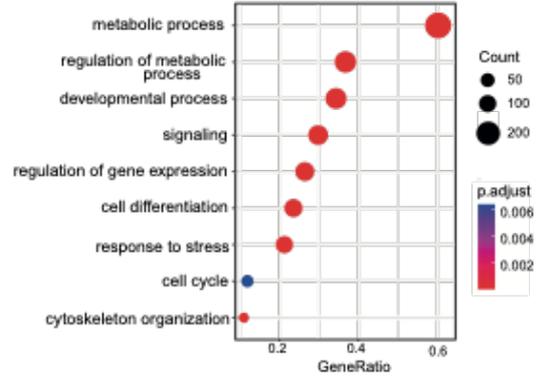
B. AS Events detected with MAJIQ C. AS Events with $\Delta\Psi > 0.1$



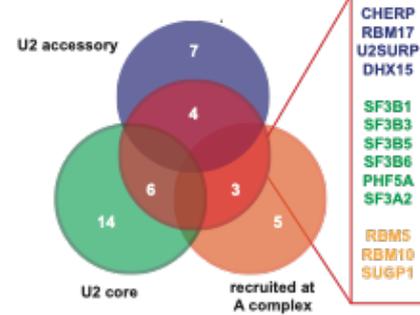
E. CHERP immunoprecipitation in different cell lines



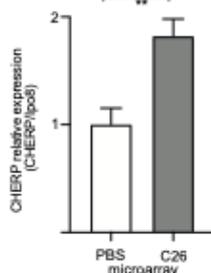
D. GO terms DEXSeq



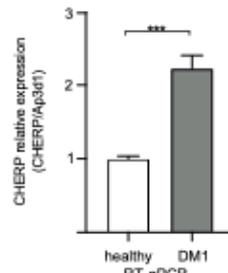
F. U2 component v.s. CHERP IP



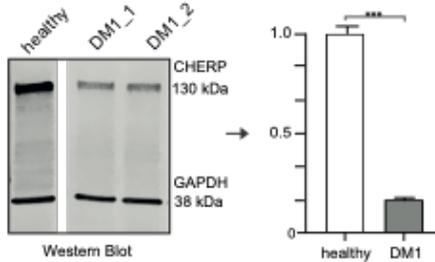
G. CHERP expression in colon cancer



H. CHERP expression in myotonic dystrophy type 1



I. CHERP expression in myotonic dystrophy type 1



J. Comparison of affected GO terms in myotonic dystrophy type 1 and siCHERP

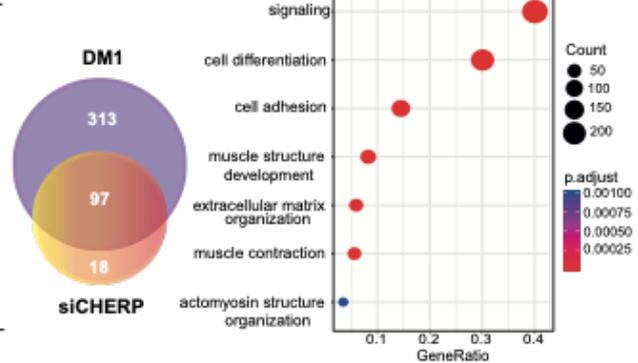


Figure S1

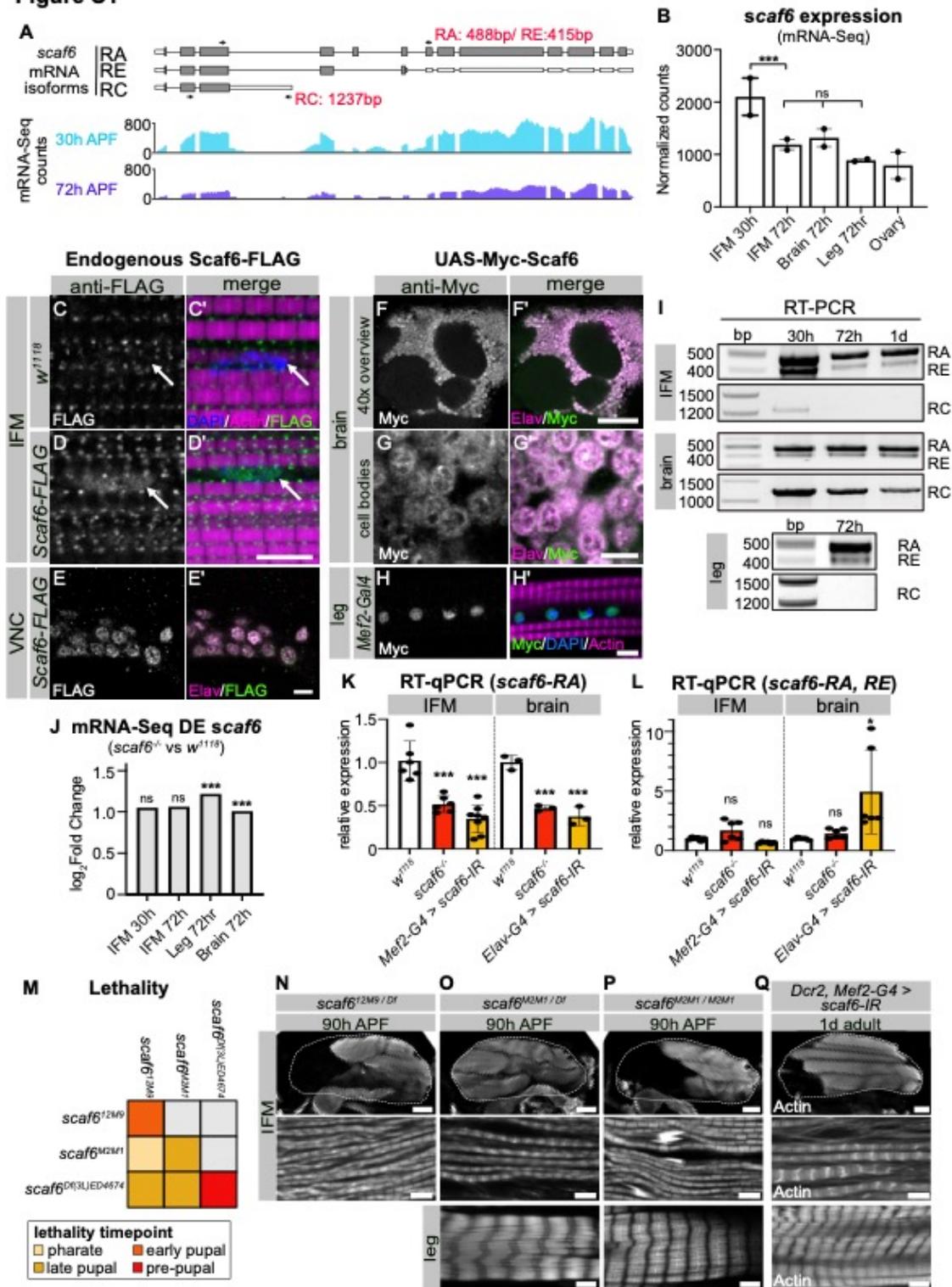


Figure S2

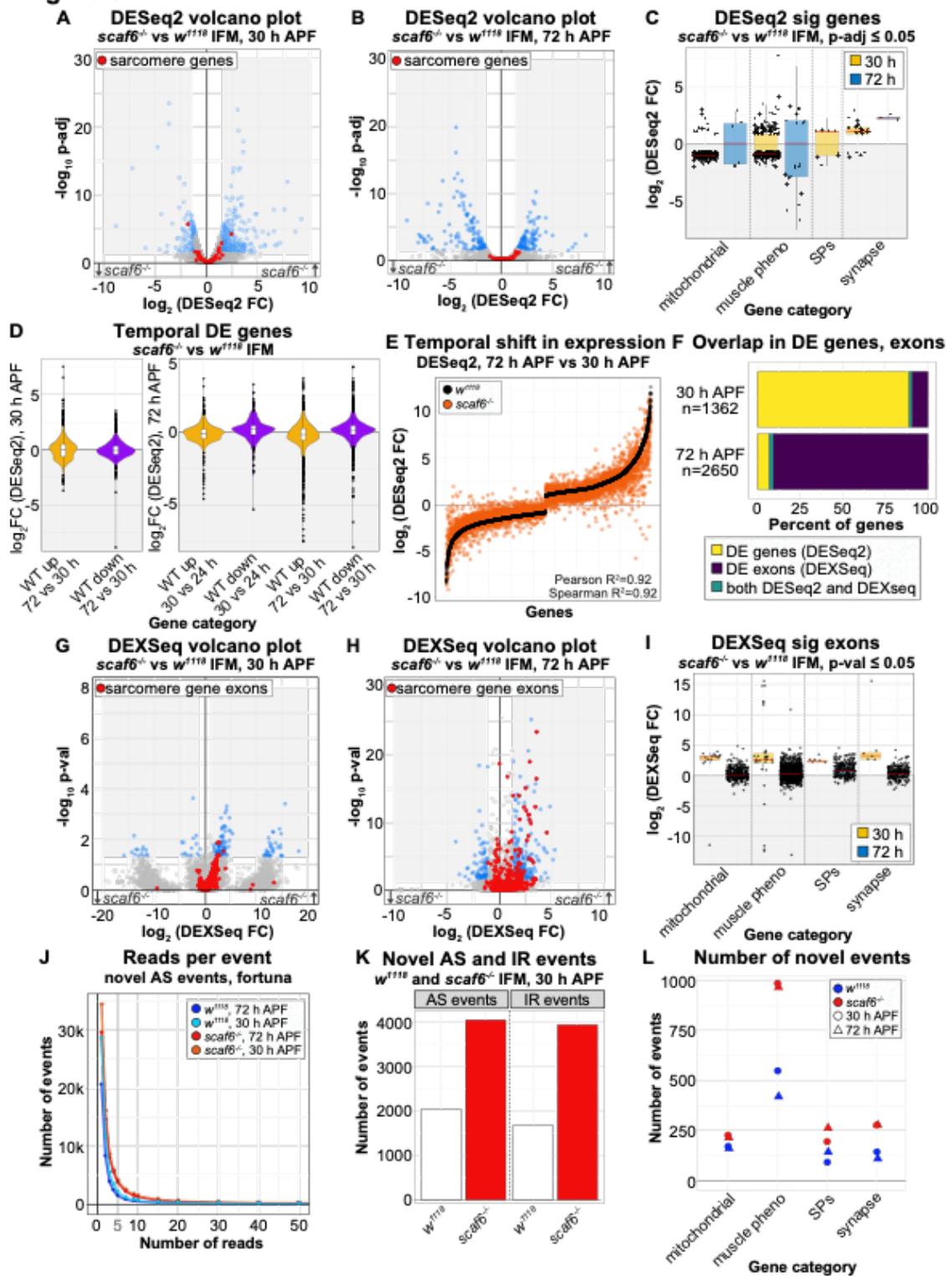


Figure S3

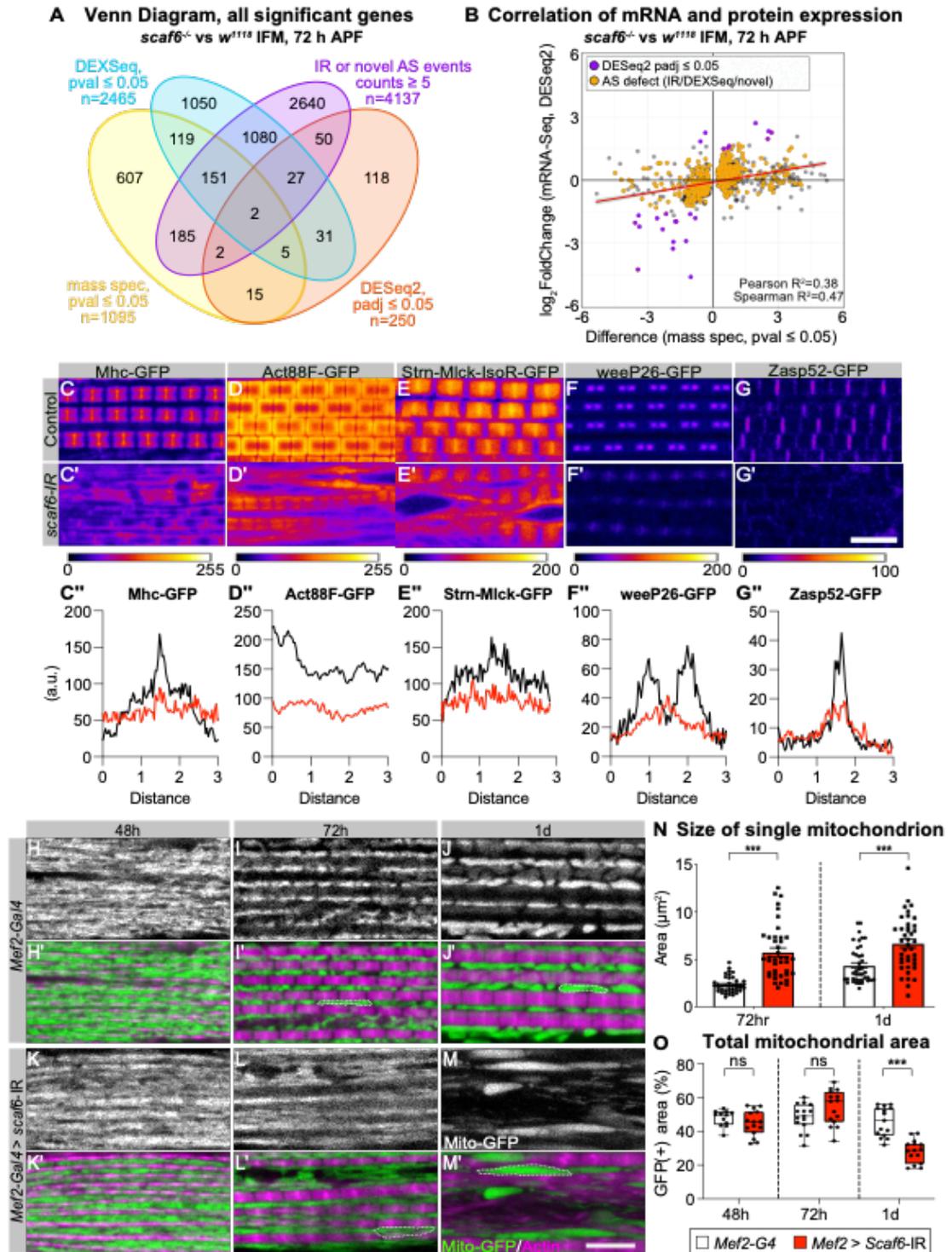
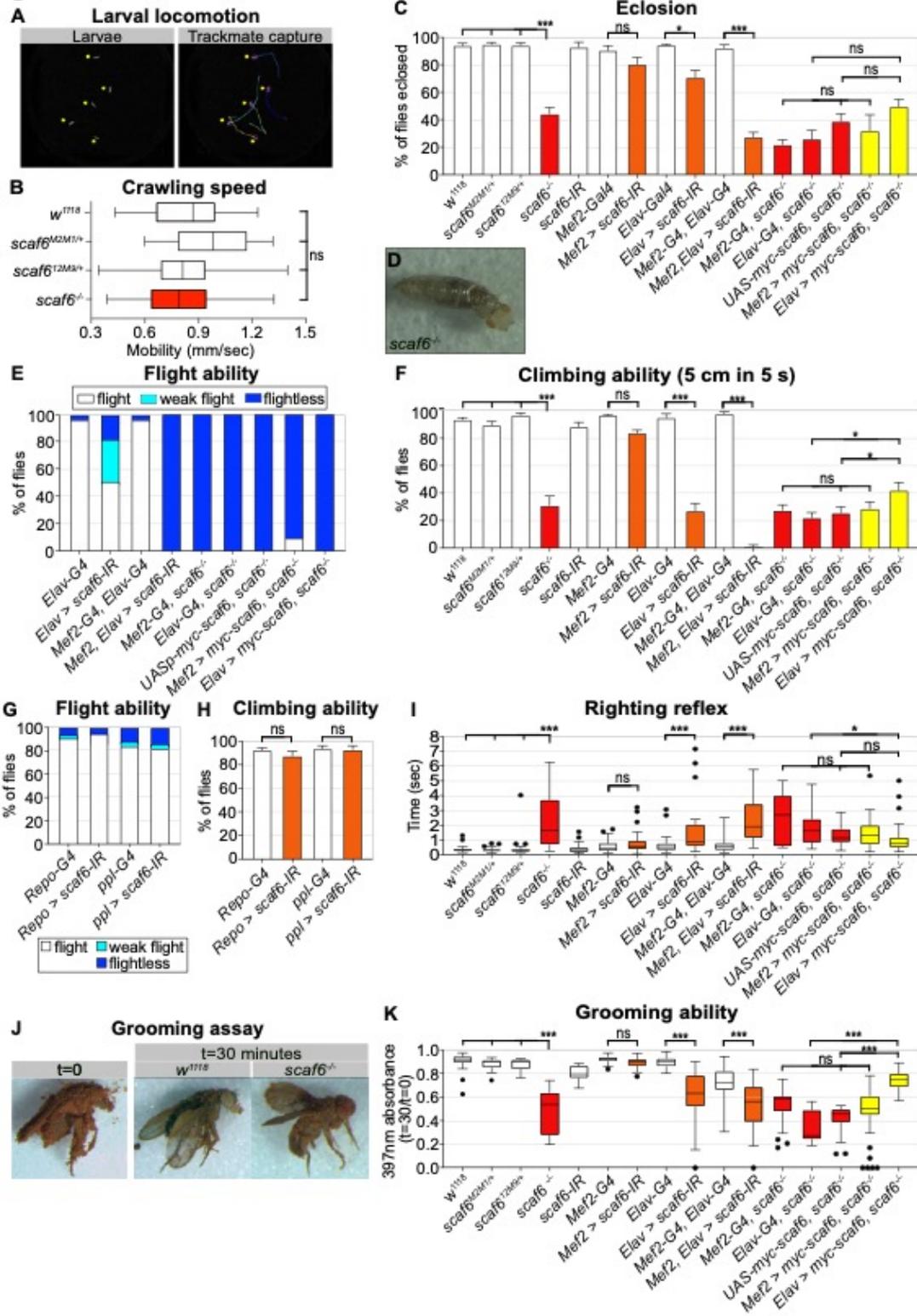


Figure S4



ACKNOWLEDGEMENTS

First and foremost, I want to thank Dr. Peter Meinke who guided me through this exciting time. He is the captain of our tiny boat, navigating through the storms of science and politics while we are busy with lighting the oil lamps, piercing the dead of night. And who is better in lighting the lamps and keeping them lit in the storm than Dr. Stefan Hintze, who is our first officer on board? He is also the frame and the planks and the glue without everything is prone to fall apart. Watching over us from ashore is our Admiral Prof. Dr. Benedikt Schoser, providing the coordinates of this turbulent journey, pulling the strings and weaving the nets. But the cruise would come to an untimely end without our crew members, Alexandra and Leijla, preparing the wax and the oil and singing the shanties that keep everyone motivated and cheerful. The Meinke crew may sail on a tiny boat now, but I can already feel the polished planks of the grand sailing ship it will be in the future.

I want to express my greatest respect, gratitude and affection for Prof. Dr. Maria Spletter, who supervised me in my Master's thesis and beyond. I am not the kind to idolize, but if I'd be, I know who I'd choose.

It was my explicit wish to be granted the *Doctor rerum naturalium*. Without the consent, interest and wise support of Prof. Dr. Hans Straka, this would have been impossible, and I want to thank him for that. I am very sad that he passed away so suddenly and untimely and I hope, this dissertation would have gained his approval. I am glad and grateful that Prof. Dr. Benedikt Grothe agreed to step in as my supervisor. Life has many twists and turns, and as chance would have it, Prof. Grothe was the first professor whose scientific drive triggered mine, back then during my Bachelor's.

Many thanks and love shall also be shared with my academic grandfathers Prof. Dr. Manfred Wehnert and Prof. Dr. Eric Schirmer whose wisdom and excellent supervision reach me through our captain. I will carry their wise words and concepts to the next generation.

Finally, there are the people I come back to after a long and dangerous navigation, who never grow tired of my adventures and who firmly believe in me finding India and would be very well satisfied if it is America instead. Namely, this is my family: my Mom and Dad and lovely sisters and the best man in the world, Dr. Chris Mahl, who is my Senpai in many aspects and my muse in all. Mom and dad are two sides of the same coin, one advises me to trust and the other to question – both is necessary on the long journey ahead. Cast off!

CURRICULUM VITAE

EDUCATION

- since 04/2020 PhD in Neurobiology at the Ludwig-Maximilians-Universität München (LMU)
- 10/2017 – 04/2020 Studies in Molecular and Cellular Biology at the LMU, degree: Master of science, final grade: 1.2, Studies of Philosophy of Science at LMU
- 10/2013 - 10/2017 Studies in biology at LMU, degree: Bachelor of science (240 ECTS points), final grade: 1.6
- 05/2011 Abitur at Ludwigsgymnasium München, final grade: 1.6

RESEARCH EXPERIENCE

- 03/2023 – 06/2023 Research project in the laboratories of Prof. Dr. Ichizo Nishino at the National Center for Neurology and Psychology (NCNP) in Tokyo, topic: Single-nuclei RNAseq in myotonic dystrophy patient muscle biopsies, contact information: nishino@ncnp.go.jp
- since 04/2020 PhD in the research laboratories of Prof. Dr. Benedikt Schoser and Dr. Peter Meinke at the Friedrich-Baur-Institute LMU, topic: Comparison of splicing and nuclear envelope alterations in muscular dystrophies, contact information: benedikt.schoser(at)med.uni-muenchen.de, peter.meinke(at)med.uni-muenchen.de
- 05/2019 – 03/2020 Master's thesis in Dr. Maria Spletter's research laboratories at the Biomedical Center LMU, topic: The role of the splicing factor CHERP in muscle, grade 1.0, contact information: maria.spletter@umkc.edu
- 10/2016 – 08/2017 External Bachelor's thesis in the research laboratories of Prof. Dr. Jeroen Krijgsveld at the European Molecular Biology Laboratory (EMBL), topic: Stem cells and proteomics, grade 1.0
- 02/2016 – 05/2016 Practical course in the research laboratories of Prof. Dr. Udo Jeschke of the Frauenklinik LMU, topic: Estrogen receptor promotor analysis in endometrial cancer

FURTHER QUALIFICATIONS

Language	English fluent in speech and writing (C1) Japanese Basics (A2) German (first language)
Computer literacy	Linux, Mac, Windows experienced R professional Excel experienced Python Basics Illustrator experienced
Special Courses	CCPBioSim workshop EMBL Cambridge, protein function prediction Data science at the Barcelona Graduate School of Economics (GSE) Laboratory animal Science course EU Data science using R (Helmholtz-Center Munich) Philosophy of Science, Logic P2P mentoring program LMU (as a mentor)

HONORS AND AWARDS

March 2023 (for October 2023)	Invitation to the DM (myotonic dystrophy) workshop for young investigators by Andy Berglund, New York DM Center
July 2023	Award for Excellent Shotgun Poster presentation at the Splicing Conference, Lisbon
July 2023	Presentation at the SEB centenary conference, Edinburgh
2022/2023	DAAD/JSPP postdoctoral research fellow (short-term) for a research project in Tokyo (National Center for Neurology and Psychiatry)
2020	Research Grant of the DGM (German muscle dystrophy society)
2019	Lehre@LMU award
2015 - 2018	Scholarship of the "Kölner Gymnasial-und Stiftungsfonds"

RESEARCH INTERESTS AND EXPERTISE

Topics	Muscle development and disease Splicing regulation Nuclear envelope and genome organization
Methods	Bioinformatics RNA-sequencing Mass spectrometry Fluorescence microscopy

Primary and stem cell culture
Basic molecular-biological methods (Western Blot, qPCR etc.)

RESEARCH PUBLICATIONS

- 2023 Donandt, T., **V. Todorow**, S. Hintze, A. Graupner, B. Schoser, M. C. Walter, and P. Meinke. 2023. 'Nuclear Small Dystrophin Isoforms during Muscle Differentiation', *Life (Basel)*, 13.
- 2022 **Todorow**, V., S. Hintze, B. Schoser, and P. Meinke. 2022. 'Nuclear envelope transmembrane proteins involved in genome organization are misregulated in myotonic dystrophy type 1 muscle', *Frontiers in Cellular and Developmental Biology*, 10: 1007331.
- 2022 de Las Heras, J. I., **V. Todorow**, L. Krečinić-Balić, S. Hintze, R. Czapiewski, S. Webb, B. Schoser, P. Meinke, and E. C. Schirmer. 2022. 'Metabolic, fibrotic, and splicing pathways are all altered in Emery-Dreifuss muscular dystrophy Spectrum patients to differing degrees', *Human Molecular Genetics*.
- 2022 Hintze, S., L. Baber, F. Hofmeister, S. Jarmusch, **V. Todorow**, S. Mehaffey, F. Tanganelli, U. Ferrari, C. Neuerburg, D. Teupser, M. Bidlingmaier, J. G. Marques, B. Koletzko, B. Schoser, M. Drey, and P. Meinke. 2022. 'Exploration of mitochondrial defects in sarcopenic hip fracture patients', *Heliyon*, 8: e11143.
- 2021 **Todorow**, V., S. Hintze, A. R. W. Kerr, A. Hehr, B. Schoser, and P. Meinke. 2021. 'Transcriptome Analysis in a Primary Human Muscle Cell Differentiation Model for Myotonic Dystrophy Type 1', *International Journal of Molecular Sciences*, 22.
- 2018 **Todorow**, V., M. Rahmeh, S. Hofmann, V. Kirn, S. Mahner, U. Jeschke, and V. von Schonfeldt. 2018. 'Correction to: Promotor analysis of ESR1 in endometrial cancer cell lines, endometrial and endometriotic tissue', *Archives of gynecology and obstetrics*, 298: 457.
- In submission Schätzl, T., **V. Todorow**, L. Kaiser, H. Weinschrott, B. Schoser, H. Daigner, P. Meinke, M. Kohl. 2023. 'New insights in FSHD pathology: meta-analysis identifies involvement of the neuromuscular junction, nuclear envelop and spliceosome'