A Multistage Sequencing Strategy Pinpoints Many Novel and Candidate Disease Alleles for Orphan Disease Emery-Dreifuss Muscular Dystrophy and Supports Gene Misregulation as its Pathomechanism

Peter Meinke[1,2], Alastair R. W. Kerr[1], Rafal Czapiewski[1], Jose I. de las Heras[1], Elizabeth Harris[3], Heike Kölbel[4], Francesco Muntoni[5], Ulrike Schara[4], Volker Straub[3], Benedikt Schoser[2], Manfred Wehnert[6], and Eric C. Schirmer[1]*

[1]Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, UK

[2]Friedrich Baur Institute at the Department of Neurology, University Hospital, LMU Munich, Germany

[3]John Walton Muscular Dystrophy Research Centre, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

[4]Department of Pediatric Neurology, Developmental Neurology and Social Pediatrics, University of Essen, Germany

[5]Dubowitz Neuromuscular Centre, University College London Great Ormond Street Institute of Child Health, London, UK; and 1 NIHR Great Ormond Street Hospital Biomedical Research Centre, Great Ormond Street Institute of Child Health, University College London, & Great Ormond Street Hospital Trust, London, UK

[6]Institute of Human Genetics, University of Greifswald (retired), Greifswald, Germany

*Corresponding author:

Prof. Dr. Eric C. Schirmer

The Wellcome Centre for Cell Biology

University of Edinburgh, Kings Buildings

Michael Swann Building, Room 5.22

Max Born Crescent

Edinburgh, EH9 3BF, UK

Phone: +441316507075

Fax: +441316507360

E-Mail: e.schirmer@ed.ac.uk

Abstract

Limitations of genome-wide approaches for genetically-heterogenous orphan diseases led us to develop a new approach to identify novel Emery-Dreifuss muscular dystrophy (EDMD) candidate genes. We generated a primer library to genes: (I) linked to EDMD, (II) mutated in related muscular dystrophies, (III) highlighted from limited exome sequencing, (IV) encoding muscle-specific nuclear membrane proteins. Sequencing 56 unlinked EDMD patients yielded confirmed or strong candidate alleles from all categories, accounting for most remaining unlinked patients. Known functions of newly-linked genes argue the EDMD pathomechanism is from altered gene regulation and mechanotransduction through connectivity of candidates from the nuclear envelope to the plasma membrane.

Keywords: Emery-Dreifuss muscular dystrophy; nuclear envelope; nuclear envelope transmembrane protein; primer library; orphan disease

Emery-Dreifuss muscular dystrophy (EDMD) is a rare neuromuscular disorder affecting ~0.3-0.4 in 100,000 people[1,2]. EDMD patients present typically in childhood with early contractures of elbows and Achilles' tendons, progressive wasting of lower leg and upper arm muscles, and later development of cardiac conduction defects and, in a proportion of cases, dilated cardiomyopathy[3]. Features vary considerably in clinical presentation, leading to the usage 'Emery-Dreifuss-like syndromes'[4,5]: patients from the same pedigree can show remarkable phenotypic variation[6-8]. Consistent with this variation, EDMD is also genetically variable: ~half of Emery–Dreifuss-like syndrome cases are linked to mutations in genes encoding 6 nuclear envelope proteins (emerin, lamin A, nesprin 1, nesprin 2, SUN1 and FHL1[8-12]). Variants in desmin and nuclear envelope proteins Tmem43 and SUN2 have been reported to modify the EDMD phenotype[11,13,14]. Roughly half of clinically diagnosed patients remain unlinked[15,16].

The strong nuclear envelope link raised the possibility that remaining unlinked patients might also have mutations in nuclear envelope proteins. The nuclear envelope is linked to >30 inherited diseases and syndromes[17], each with distinct tissue-specific pathologies: for example different lamin A mutations cause muscular dystrophies, neuropathy, lipodystrophy, and multisystemic disorders[18]. How these widely expressed nuclear envelope proteins yield tissue-specific pathologies remains unresolved, but one hypothesis is that tissue-specific nuclear envelope partners mediate the tissue-specificity of effects[18,19].

We previously identified several muscle-specific nuclear envelope transmembrane proteins (NETs)[20]. Of the previously linked proteins emerin, nesprin 1, nesprin 2, SUN1, SUN2, and Tmem43 are all NETs, but these are widely expressed. Several of the muscle-specific NETs identified could contribute muscle specificity to either of the two principly hypothesized EDMD pathomechanisms: mechanical instability and disruption of gene

expression[21]. NETs Tmem214 and KLHL31 track with microtubules on the nuclear surface[20] while NET5/Samp1 contributes actin and centrosome interactions[22,23]. NETs Tmem38A, WFS1, NET39/PLPP7 and, again, Tmem214 and NET5/Samp1 all affect 3D gene positioning and with corresponding effects on expression[24,25]. Many of the genes under muscle-specific NET regulation are recruited to the nuclear periphery to be more tightly shut down during myogenesis and encode proteins that are antagonistic to myogenesis or are from alternative differentiation pathways such as adipocytes. Knockdown of the muscle-specific NETs results in these genes being de-repressed, suggesting a possible gene misregulation mechanism to disease pathology. The potential of gene mispositioning contributing to disease is further underscored by knockdown of Tmem38A, WFS1, and NET39/PLPP7 blocking myotube fusion[24]. Functional overlap of these muscle-specific NETs supports the possibility of their working in a common pathway towards EDMD pathophysiology, making them good candidates for mediators of EDMD muscle pathology at the same time as being novel candidates for causative EDMD alleles.

Therefore, we elected to sequence the genes encoding these muscle-specific NETs in unlinked EDMD patients using a primer library. However, for greater surety, we expanded this primer library to also re-check previously linked genes with complete gene sequencing for possible promoter mutations and to test for mutations in genes linked to related muscular dystrophies. Finally, to also search for candidate alleles in a completely unbiased manner, we performed exome sequencing in families for which material from enough members was available for linkage analysis and added these candidates also to the primer library.

Results

SEQUENCING EDMD FAMILIES

Whole exome sequencing was performed in 12 EDMD patients and 12 unaffected individuals from 5 families with large enough pedigrees for linkage analysis (Fig. 1), finding over 250,000 variants compared to the reference sequence. Variants were filtered using criteria: (a) phenotype co-segregation and modes of inheritance for each family; (b) selecting for SNP frequencies <1%, and filtering for <0.05%; (c) affecting coding sequence; (d) function/tissue-expression of the encoded protein *e.g.* >2-fold higher expression in muscle compared to other tissues. Filtering yielded 213 candidate genes for families 2-5 (Supplemental Table S1).

Family 1 yielded no convincing candidates. As this family had the largest pedigree, we postulated that an unaffected individual was a carrier who had not yet presented or had a distinct sporadic form of disease. Dropping younger individuals who may have not yet presented clinically failed to yield candidates; therefore, genome and transcriptome sequencing was performed in the index patient, resulting in 33 additional candidates (Supplemental Tables S2 and S3). The combined exome, genome and RNA sequencing yielded a total of 252 candidates for the five families.

PRIMER LIBRARY SEQUENCING

A primer library was generated containing (I) the 8 previously-linked EDMD gene ORFs plus the whole genes for *LMNA* and *EMD* (that together account for ~40% of linked alleles), (II) 25 genes from similar muscular dystrophies, (III) the 252 exome sequencing candidates, and (IV) 16 functional candidates, mostly muscle-specific nuclear envelope proteins (Fig. 2A; Supplemental Table S4). Sequencing was performed on 56 additional

unlinked clinically diagnosed EDMD patients unrelated to each other, obtaining on average 3,427,092 reads per patient. The data were analyzed for genes carrying mutations that changed the coding sequence (nonsense, missense, splice sites) with expected altered protein function (e.g. non-conservative substitutions) and SNP frequencies <0.05% (Supplemental Table S5).

Candidate mutations were found in all four categories. Of category I previous EDMD-linked genes, *LMNA* had mutations in three patients that were missed in standard diagnostics (p.R41H, p.R249Q, p.G535fs*; Table 1). These mutations were determined as causative based on similarity to previously linked *LMNA* mutations. Previously EDMD-associated genes *SYNE1*, *SYNE2*, *SUN1* and *TMEM43* also had mutations; however, minor allele frequencies and their combination with other mutations made them unlikely as causative alleles excepting *SYNE1*. Modifying effects, nonetheless, cannot be excluded. No mutations were found in *LMNA* or *EMD* non-coding regions.

Gene category II of related muscular dystrophies yielded 18 patients with mutations considered causative. Four of these patients had combinations of a missense and frameshift mutation in *CAPN3* (Table 1). *GBE1* mutations were found in four other patients: three missense and one splice-site. *VCP* and likely recessive *TTN* were mutated in two patients each; however, *TTN* mutation patients also carried *SYNE1* mutations. Genes with one patient carrying likely disease-causing mutations were *COL6A1*, *CAV3*, *DMD*, *ANO5*, *DYSF* and *POMT1*. The *DMD* mutation created a stop codon at codon three, resulting in possible usage of an alternative start codon and a milder phenotype than Duchenne[26]. For *ANO5*, *DYSF* and *POMT1* the respective patients had two mutations, consistent with the reported inheritance (autosomal recessive for MD-20/*ANO5* and unknown for MD-21/*DYSF* and MD-23/*POMT1*; Table 1). However, lacking DNA from the parents we could not perform segregation studies.

Several category III genes from exome sequencing were elevated to strong candidates if mutated in multiple patients within the primer library cohort based on the assumption that causative genes will be independently mutated in multiple patients. The top candidates were *INTS1*, *ANK2*, *XIRP1* and *USP34*. Heterozygous *ANK2* mutations were identified in family 5 plus six cohort patients with no other obvious disease-causing mutations and so were most likely causative (Table 2). Causation is similarly likely for other genes; however, in some patients there were additional candidate alleles identified. Heterozygous *INTS1* was mutated in four members of family 3 plus five cohort patients, four of whom had no mutation in already associated genes (Table 2). The last patient, MD-23, additionally carried two *POMT1* mutations; however, it is unclear if the likely recessive *POMT1* mutations affected one allele or both so causation remains undetermined. Other good category IV candidates were *USP34* (heterozygous mutations in exome sequenced family 2) and *XIRP1* (mutated in families 2 and 4), each with mutations in an additional five patients. Some patients had additional mutations in already associated genes, but if these other mutations were causative then modifying effects for the new candidates are still possible.

Several category IV functional/tissue candidate genes were mutated in 16 of the 56 primer library cohort patients. These were *WFS1* (4 patients), *TMEM201* (3 patients), *TMEM38A* (3 patients), *PLPP7* (2 patients), *TMEM214* (2 patients), *LPCAT3* (1 patient), *KLHL31* (1 patient), and *BVES* (1 patient). Of these, three patients with *TMEM38A* mutations, two patients with *TMEM214* mutations, one patient with an *LPCAT3* mutation and one patient with a *BVES* mutation were clearly the top candidates with no other reasonable candidates identified and patient MD-32 carried mutations in both *TMEM38A* and *PLPP7*. Other mutations identified were in association with other possible candidates that included likely causative mutations in *GBE1*, *COL6A1*, *LMNA* and *TTN* (details in Table 1). The patient with the combined *LMNA* and *TMEM201* mutations had a very early age of onset (1 year),

suggesting that both mutations contribute to the more severe (congenital) phenotype as the *LMNA* mutation has not been associated with congenital muscular dystrophy.

All in all, sequencing the 56 additional patients with the primer library found mutations in only a subset of the 252 candidates from the exome sequencing and this subset is expected to be much higher confidence because causative genes are more likely to be also mutated in other EDMD patients. In contrast, mutations were found in 19 of 25 related muscular dystrophies and in 11 of 16 functional candidates; so a strong enrichment for these candidate pools was observed (Fig. 2A).

NUCLEAR ENVELOPE LINKS

All previously linked EDMD genes encode nuclear envelope proteins. The functional candidates were also biased towards genes identified in the nuclear envelope by proteomics; however, there was no bias towards the nuclear envelope when selecting genes for the primer library from similar muscular dystrophies or from exome sequencing. Nonetheless, the majority of genes from similar muscular dystrophies encode proteins for which at least a subpopulation associates with the nuclear envelope (Fig. 2B). Interestingly, just considering the candidates from the exome sequencing in which mutations were also found in other patients from the primer library sequencing, the nuclear envelope portion increased from less than 10% to more than 40% - considerable more than the overall genome portion of 5.9% (Fig. 2B). Of note, the proteins encoded by genes linked to other muscular dystrophies such as *COL6A1*, *CAV3*, *DYSF*, *DMD*, *TTN*, and *VCP* and the strongest family sequencing candidates *INTS1* and *ANK2* were all found in nuclear envelope proteomics datasets[20,27]. While these could reflect either a separate pool in the nuclear envelope or connections that were maintained during nuclear envelope isolation, this suggests at least an indirect physical connection of these candidates to the nuclear envelope.

The two top argued mechanisms for how mutations in nuclear envelope proteins can cause pathology are mechanical instability and genome misregulation. Genes in different candidate categories contained Gene Ontology (GO)-terms for functions in gene regulation, cytoskeleton, and both together. Interestingly, the likely candidates from all categories were enriched for genes simultaneously linked to both gene regulation and cytoskeleton GO-terms compared to the overall genome (Fig. 2C). Such genes may be involved in mechanosignal transduction to the genome. Consistent with this idea, most of the proteins encoded by the final candidate genes interact with other candidates according to interactome studies and these interactions form a chain of connectivity between the nuclear envelope and the plasma membrane via cytoskeletal proteins that could support mechanotransduction to the genome (Fig. 2D).

CONFIRMATION OF NOVEL EDMD ALLELES

Thus far only the three *LMNA* mutations, the *CAV3* and one of the *CAPN3* (MD-43) mutations have been confirmed as insufficient numbers of family members have come to clinic for linkage analysis. Therefore, to test the likelihood that other mutations identified cause EDMD disease pathology, we tested two of the gene regulating NETs to determine if the mutations identified disrupt their normal functions in myogenic gene regulation. In keeping with this idea, for the 8 out of 16 functional NET candidates where mutations were found (6 of which have known gene regulation functions), nearly all mutations identified faced the nucleoplasm or were positioned where they could alter membrane topology (Fig. 3A). The two muscle-specific NETs we chose to test were PLPP7/NET39 and Tmem38A. Both recruit partially overlapping, but mostly different sets of genes to the nuclear periphery to enhance their repression and many of the genes targeted are antagonistic to myogenesis or from alternate differentiation pathways[24]. Combined knockdown of PLPP7/NET39,

TMEM38A and WFS1 blocked myogenesis, providing a logical route from their disruption to muscle disease pathology. Therefore, the *PLPP7* and *TMEM38A* mutations were exogenously expressed in C2C12 myoblast cells to determine if they could perform the previously shown gene positioning function of the wild-type in recruiting specific gene targets to the nuclear periphery for enhanced repression. Tmem38A normally repositions the *DDR2* gene locus to the nuclear envelope to repress it during myogenesis, but with mutations p.N260D and p.N260del it fails to do so (Fig. 3B). Similarly, PLPP7/NET39 normally recruits the *PTN* gene locus to the nuclear envelope to repress it during myogenesis, but with mutation p.R252P it could not. PLPP7/NET39 mutation p.M92K also affected the gene positioning, though apparently in the opposite direction which might also affect expression (Fig. 3C).

Discussion

Failure of high throughput genomic approaches to identify new disease alleles can at least in some cases be overcome by our multistage approach. This approach pinpointed candidates in part based on the preferential tissue focus of pathology and in part on the subcellular localization of known alleles. Similarly applying filters in focusing candidates for such a multipronged approach can be applied to other genetically heterogeneous diseases.

With nearly half of EDMD cases previously linked to genes encoding 6 nuclear envelope proteins it was clear that EDMD is a nuclear envelope disease. This is strengthened by enrichment for nuclear envelope proteins amongst our top new candidate alleles. *COL6A1*, *CAV3*, *DYSF*, *DMD*, *TTN*, and *VCP* gene products were found in muscle nuclear envelopes[20]. As most of these proteins have previously been associated with the cytoskeleton or plasma membrane, their association with the nuclear envelope may be indirect through lamin-cytoskeletal connections. However, this association could also be due to splice variants that target to the nuclear envelope or specific translocation to the nucleus under certain conditions as has been shown for *CAV3* family member caveolin 2. In this case, a caveolin 2 subpopulation translocates to the nucleus and interacts with lamin A to regulate histone modifications and gene expression[28].

The gene positioning defects for *TMEM38A* and *PLPP7* mutations not only further link the nuclear envelope to EDMD, but also strengthen the idea that misregulation of myogenic gene expression is the primary cause of EDMD pathology. In addition to Tmem38A and Plpp7, the muscle NETs Tmem214, WFS1, and NET5/Samp1 all have gene-repositioning functions that contribute to gene regulation and the NET MAN1 affects gene regulation through its interactions with Smads as well as binding several chromatin partners[29,30]. The involvement of these muscle gene repositioning NETs, not only as novel causative alleles but also in mediating EDMD pathology caused by mutations in widely

expressed nuclear envelope proteins, is further supported by WFS1, Tmem214, Tmem38A, and NET5/Samp1 being mislocalized in isolated differentiating EDMD muscle cells or muscle biopsy sections[31].

Of the previously EDMD-linked nuclear envelope proteins, Lamin A has both cytoskeletal and genome regulation functions; so its mutation could support both mechanical instability and genome misregulation hypotheses for EDMD pathophysiology[32-37]. Emerin interacts with actin supporting a cytoskeletal role, but it also has many reported contributions to genome regulation through its binding DNA condensing factors BAF and HDAC3, splicing factors, the transcription factor Lmo7, and the transcriptional repressor germ cell-less[38]. FLH1 is linked to signal transduction and splice variant FHL1B targets specifically to the nuclear envelope[39]. Moreover, FHL1 has been linked to other myopathies such as X-linked myopathy with postural muscle atrophy (XMPMA)[40] via its signal transduction function.

As signaling functions could affect both gene regulation and the cytoskeleton, these mechanisms toward pathology were considered equally likely; however, a gene mis-regulation mechanism is much more likely now with the new gene-repositioning candidate alleles identified. Though there are some other disparate functions reported for several of these NETs[20,22,41], WFS1, Tmem38A/TRIC-A, NET39/ PLPP7, Tmem214 and NET5/Samp1 are all at the nuclear envelope preferentially in muscle and all share a common function in directing gene-repositioning for regulation of gene expression during myogenesis[24]. That some of these muscle-specific NETs had overlap in their functions further supports the possibility of their working in a common pathway towards EDMD pathophysiology. At the same time, while there was some overlap in the sets of genes regulated by these muscle NETs, each had also unique gene targets. The links of candidate alleles to gene repositioning and mechanotransduction are the more compelling in this context because the different sets of

genes regulated — all important in myogenesis — thus supports the clinical variation observed in EDMD.

Our sequencing in patients diagnosed with an EDMD-like phenotype identified mutations in several genes linked to muscular dystrophies that share clinical features with EDMD. This might reflect incorrect diagnoses or their involvement in EDMD. The latter case seems likely, considering that *COL6A1*, *CAV3*, *DYSF*, *DMD*, *TTN*, and *VCP* gene products link to the nuclear envelope. Indeed, many of these gene products interact with one another in a way that could form a chain from the plasma membrane to the nuclear envelope (Fig. 2D). This also is compelling to this gene regulation mechanism as this chain could play a role in mechanosignal transduction to the nucleus.

Finally, as the families chosen for exome sequencing all had differences in presentation, there are likely additional mutations picked up in the primer library that eventually might be used as predictors of severity or other aspects of clinical presentation once further sequencing reveals better correlations. Thus it would be beneficial to continue using this primer library diagnostically both to find these correlations and because it is cheaper and faster than standard iterative Sanger sequencing for such a genetically variable disease to identify mutations in known linked genes. In general, this iterative multipronged approach, combining into a primer library a set of preliminary candidates from exome sequencing in which only sufficient pedigrees exist for partial linkage analaysis together with candidates from related disorders and candidates specific to the tissue where pathology is manifested that are associated with linked organelles and functions, might be applied to a wider range of genetically heterogeneous orphan diseases where insufficient numbers of patients are available for standard genome and exome approaches to be effective.

Materials and Methods

PATIENT MATERIALS AND ETHICS

Patient DNA was obtained from the Muscle Tissue Culture Collection (MTCC) at the Friedrich-Baur-Institut (Department of Neurology, Ludwig-Maximilians-University, Munich, Germany), the Institute of Human Genetics, University of Newcastle upon Tyne, Newcastle upon Tyne, UK, the MRC Centre for Neuromuscular Disorders Biobank (CNDB) in London, the Department of Pediatric Neurology, Developmental Neurology and Social Pediatrics at the University of Essen, the Rare Diseases biological samples biobank at the Dubowitz Neuromuscular Centre, Great Ormond Street Hospital for Children NHS Trust, London, UK. All materials were obtained with written informed consent of the donor at the CIND, the CNDB or the MTCC. Ethical approval for the Newcastle MRC Centre Biobank for Neuromuscular Diseases is covered by REC reference 08/H0906/28+5 and IRAS ID 118436 and MTA CT-2166, that 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 CNMDBBL63 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.

EXOME, RNA, AND GENOME SEQUENCING

Genome: 15X clean depth coverage using 90PE Illumina HiSeq2000 technology. RNA-Seq: total RNA from biopsy tissue with rRNA depletion and random-primed cDNA preparation and PE100 sequencing on a Hi-Seq2000 platform with 20 million reads minimum (Otogenetics Corporation, Norcoss, USA).

Exome: Sequencing was performed on the Illumina HiSeq and raw data processed with CASAVA 1.8.

FLUORESCENCE *IN SITU* HYBRIDIZATION

Mutations were generated by Agilent Site-Directed mutagenesis. Plasmids encoding tagged Tmem38a, PLPP7 and mutants were transfected using Lipofectamine 3000 (Invitrogen) into C2C12 cells (ATCC, VA, USA) cultured at 37°C, 5% CO2 in DMEM containing 20% FBS, 50U/mL penicillin and 10mg/mL streptomycin. Fluorescent *in situ* hybridization (FISH) experiments were performed as described in[42].

PRIMER LIBRARY CONSTRUCTION, PROCESSING AND SEQUENCING

A SureSelect^{XT} Custom 1.638 Mbp target enrichment library (5190-4817) containing 25,036 oligonucleotide probes against H. sapiens hg19 GRCh37 sequence as of February 2009 was prepared by Agilent for use with Illumina multiplexed sequencing platforms. Patient genomic DNA was isolated from blood and prepared for sequencing using the SureSelect^{QXT} Reagent Kit (G9681B) according to the manufacturer's instructions. Recommended minimum sequencing per sample was 327.793 Mbp and an average of 3,427,092 was obtained with a range from 442,125 to 7,066,507 using 125 base paired-end sequencing on a Hi-Seq2500.

BIOINFORMATICS

Variant analysis was performed using the Genome Analysis toolkit [GATK] v2.7-2[43] and picard tools v1.74 (http://broadinstitute.github.io/picard/) using GATK Best Practices recommendations[44,45] against human genome assembly hg19. The allele frequencies of variants were cross-referenced with gnomAD version 2.1[46] using both the genome and exome datasets.

RNA-Seq: STAR v2.1.1[47] was used to map reads to the hg19 reference genome, samtools v0.1.19[48] was used for file conversion. Deeptools v1.5.1[49] was used for downstream analysis.

Acknowledgements

We dedicate this study in honor of Alan E.H. Emery who not only described X-linked EDMD as a distinct disease in 1966 but also initiated the longest standing regular Laminopathy meeting in 2006. We further want to thank all patients and their families for their valuable contribution. Funding for this work was principally provided by Wellcome Trust Grants 095209, Muscular Dystrophy UK grant 18GRO-PG24-0248, and MRC grant MR/R018073/1 to ECS and 092076 for the Centre for Cell Biology. Funding was also provided by the European Community's Seventh Framework Programme (FP7/2007-2013) "Integrated European –omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)" (grant agreement n° 2012-305121); the Muscular Dystrophy UK Grant on Gene Identification to FM, and the support of the MRC Neuromuscular Centre Biobank at UCL is also gratefully acknowledged.

Author contributions

ECS, MW and PM designed the project. PM performed the sequencing and data analyses. RC performed the FISH experiments. ARWK and JIH performed the bioinformatics. EH, HK, FM, US, VS and BS contributed patient material and clinical description. PM and ECS wrote the paper. All authors read the manuscript, offered feedback, and approved it before submission.

References

1 Hughes, M. I., Hicks, E. M., Nevin, N. C. & Patterson, V. H. The prevalence of inherited neuromuscular disease in Northern Ireland. *Neuromuscul Disord* **6**, 69-73 (1996).

2 Norwood, F. L. *et al.* Prevalence of genetic muscle disease in Northern England: in-depth analysis of a muscle clinic population. *Brain* **132**, 3175-3186, doi:10.1093/brain/awp236 (2009).

3 Emery, A. E. & Dreifuss, F. E. Unusual type of benign x-linked muscular dystrophy. *J Neurol Neurosurg Psychiatry* **29**, 338-342 (1966).

4 Emery, A. E. Emery-Dreifuss syndrome. *J Med Genet* **26**, 637-641 (1989).

5 Knoblauch, H. *et al.* Contractures and hypertrophic cardiomyopathy in a novel FHL1 mutation. *Ann Neurol* **67**, 136-140, doi:10.1002/ana.21839 (2010).

6 Bonne, G. *et al.* 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 (2000).

7 Canki-Klain, N. *et al.* Clinical variability and molecular diagnosis in a four-generation family with X-linked Emery-Dreifuss muscular dystrophy. *Croat Med J* **41**, 389-395 (2000).

8 Bonne, G. *et al.* Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nat Genet* **21**, 285-288, doi:10.1038/6799 (1999).

9 Bione, S. *et al.* Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat Genet* **8**, 323-327, doi:10.1038/ng1294-323 (1994).

10 Zhang, Q. *et al.* 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, doi:10.1093/hmg/ddm238 (2007).

11 Meinke, P. *et al.* Muscular dystrophy-associated SUN1 and SUN2 variants disrupt nuclear-cytoskeletal connections and myonuclear organization. *PLoS Genet* **10**, e1004605, doi:10.1371/journal.pgen.1004605 (2014).

12 Gueneau, L. *et al.* Mutations of the FHL1 gene cause Emery-Dreifuss muscular dystrophy. *Am J Hum Genet* **85**, 338-353, doi:10.1016/j.ajhg.2009.07.015 (2009).

13 Liang, W. C. *et al.* TMEM43 mutations in Emery-Dreifuss muscular dystrophy-related myopathy. *Ann Neurol* **69**, 1005-1013, doi:10.1002/ana.22338 (2011).

14 Muntoni, F. *et al.* Disease severity in dominant Emery Dreifuss is increased by mutations in both emerin and desmin proteins. *Brain* **129**, 1260-1268, doi:10.1093/brain/awl062 (2006).

15 Wehnert, M. S. & Meinke, P. in *eLS* (John Wiley & Sons, Ltd, 2012).

16 Meinke, P. *Molecular Genetics of Emery–Dreifuss Muscular Dystrophy*. (John Wiley & Sons, Ltd, 2018).

17 Meinke, P. & Schirmer, E. C. The increasing relevance of nuclear envelope myopathies. *Curr Opin Neurol* **29**, 651-661, doi:10.1097/WCO.0000000000000359 (2016).

18 Worman, H. J. & Schirmer, E. C. Nuclear membrane diversity: underlying tissue-specific pathologies in disease? *Curr Opin Cell Biol* **34**, 101-112, doi:10.1016/j.ceb.2015.06.003 (2015).

19 Wilkie, G. S. & Schirmer, E. C. Guilt by association: the nuclear envelope proteome and disease. *Mol Cell Proteomics* **5**, 1865-1875, doi:10.1074/mcp.R600003-MCP200 (2006).

20 Wilkie, G. S. *et al.* Several novel nuclear envelope transmembrane proteins identified in skeletal muscle have cytoskeletal associations. *Mol Cell Proteomics* **10**, M110 003129, doi:10.1074/mcp.M110.003129 (2011).

21 Hutchison, C. J., Alvarez-Reyes, M. & Vaughan, O. A. Lamins in disease: why do ubiquitously expressed nuclear envelope proteins give rise to tissue-specific disease phenotypes? *J Cell Sci* **114**, 9-19 (2001).

22 Buch, C. *et al.* An integral protein of the inner nuclear membrane localizes to the mitotic spindle in mammalian cells. *J Cell Sci* **122**, 2100-2107, doi:10.1242/jcs.047373 (2009).

23 Osorio, D. S. & Gomes, E. R. Connecting the nucleus to the cytoskeleton for nuclear positioning and cell migration. *Adv Exp Med Biol* **773**, 505-520, doi:10.1007/978-1-4899-8032-8_23 (2014).

24 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, doi:10.1016/j.molcel.2016.04.035 (2016).

25 Zuleger, N. *et al.* Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. *Genome Biol* **14**, R14, doi:10.1186/gb-2013-14-2-r14 (2013).

26 Flanigan, K. M. *et al.* Rapid direct sequence analysis of the dystrophin gene. *Am J Hum Genet* **72**, 931-939 (2003).

27 Korfali, N. *et al.* The nuclear envelope proteome differs notably between tissues. *Nucleus* **3**, 552-564, doi:10.4161/nucl.22257 (2012).

28 Jeong, K., Kwon, H., Lee, J., Jang, D. & Pak, Y. Insulin-response epigenetic activation of Egr-1 and JunB genes at the nuclear periphery by A-type lamin-

associated pY19-Caveolin-2 in the inner nuclear membrane. *Nucleic Acids Res* **43**, 3114-3127, doi:10.1093/nar/gkv181 (2015).

29 Osada, S., Ohmori, S. Y. & Taira, M. XMAN1, an inner nuclear membrane protein, antagonizes BMP signaling by interacting with Smad1 in Xenopus embryos. *Development* **130**, 1783-1794, doi:10.1242/dev.00401 (2003).

30 Liu, J. *et al.* MAN1 and emerin have overlapping function(s) essential for chromosome segregation and cell division in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **100**, 4598-4603, doi:10.1073/pnas.0730821100 (2003).

31 Mattioli, E. *et al.* Samp1 Mislocalization in Emery-Dreifuss Muscular Dystrophy. *Cells* **7**, doi:10.3390/cells7100170 (2018).

32 Broers, J. L. *et al.* Both lamin A and lamin C mutations cause lamina instability as well as loss of internal nuclear lamin organization. *Exp Cell Res* **304**, 582-592, doi:10.1016/j.yexcr.2004.11.020 (2005).

33 Broers, J. L. *et al.* Decreased mechanical stiffness in LMNA-/- cells is caused by defective nucleo-cytoskeletal integrity: implications for the development of laminopathies. *Hum Mol Genet* **13**, 2567-2580, doi:10.1093/hmg/ddh295 (2004).

34 Lammerding, J. *et al.* Abnormal nuclear shape and impaired mechanotransduction in emerin-deficient cells. *J Cell Biol* **170**, 781-791, doi:10.1083/jcb.200502148 (2005).

35 Lammerding, J. *et al.* Lamin A/C deficiency causes defective nuclear mechanics and mechanotransduction. *J Clin Invest* **113**, 370-378, doi:10.1172/JCI19670 (2004).

36 Rowat, A. C., Lammerding, J. & Ipsen, J. H. Mechanical properties of the cell nucleus and the effect of emerin deficiency. *Biophys J* **91**, 4649-4664, doi:10.1529/biophysj.106.086454 (2006).

37 Hakelien, A. M., Delbarre, E., Gaustad, K. G., Buendia, B. & Collas, P. Expression of the myodystrophic R453W mutation of lamin A in C2C12 myoblasts causes promoter-

specific and global epigenetic defects. *Exp Cell Res* **314**, 1869-1880, doi:10.1016/j.yexcr.2008.02.018 (2008).

38 Holaska, J. M. & Wilson, K. L. An emerin "proteome": purification of distinct emerin-containing complexes from HeLa cells suggests molecular basis for diverse roles including gene regulation, mRNA splicing, signaling, mechanosensing, and nuclear architecture. *Biochemistry* **46**, 8897-8908, doi:10.1021/bi602636m (2007).

39 Ziat, E. *et al.* FHL1B Interacts with Lamin A/C and Emerin at the Nuclear Lamina and is Misregulated in Emery-Dreifuss Muscular Dystrophy. *J Neuromuscul Dis* **3**, 497-510, doi:10.3233/JND-160169 (2016).

40 Windpassinger, C. *et al.* An X-linked myopathy with postural muscle atrophy and generalized hypertrophy, termed XMPMA, is caused by mutations in FHL1. *American journal of human genetics* **82**, 88-99, doi:10.1016/j.ajhg.2007.09.004 (2008).

41 Tao, S. *et al.* Facilitated hyperpolarization signaling in vascular smooth muscle-overexpressing TRIC-A channels. *J Biol Chem* **288**, 15581-15589, doi:10.1074/jbc.M112.435396 (2013).

42 Zuleger, N. *et al.* Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. *Genome Biol* **14**, 2013-2014 (2013).

43 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

44 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498, doi:10.1038/ng.806 (2011).

45 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33, doi:10.1002/0471250953.bi1110s43 (2013).

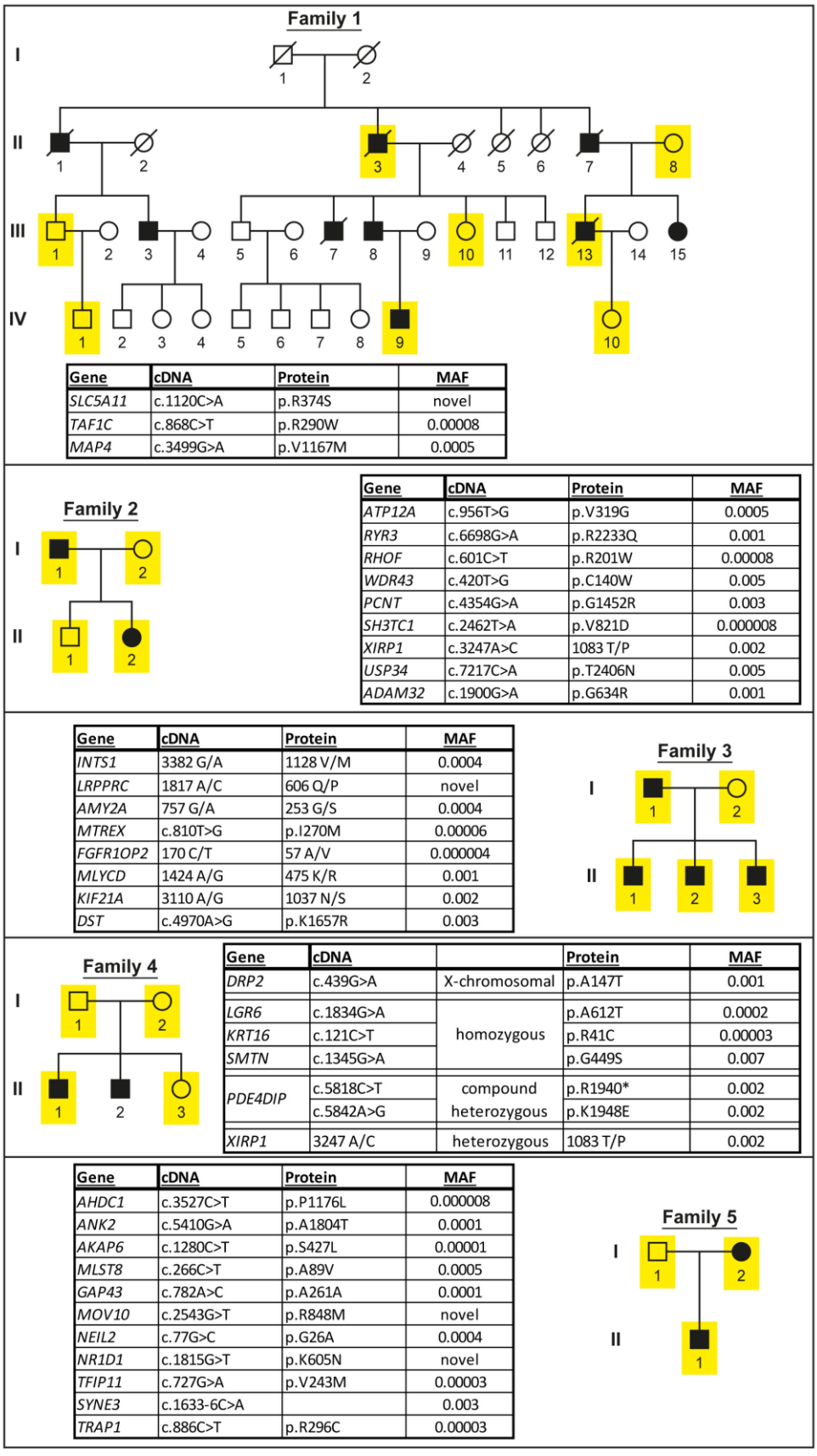
46 Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210, doi:10.1101/531210 (2019).

47 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

48 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

49 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191, doi:10.1093/nar/gku365 (2014).

FIGURES



Gene	cDNA	Protein	MAF
SLC5A11	c.1120C>A	p.R374S	novel
TAF1C	c.868C>T	p.R290W	0.00008
MAP4	c.3499G>A	p.V1167M	0.0005

Gene	cDNA	Protein	MAF
ATP12A	c.956T>G	p.V319G	0.0005
RYR3	c.6698G>A	p.R2233Q	0.001
RHOF	c.601C>T	p.R201W	0.00008
WDR43	c.420T>G	p.C140W	0.005
PCNT	c.4354G>A	p.G1452R	0.003
SH3TC1	c.2462T>A	p.V821D	0.000008
XIRP1	c.3247A>C	1083 T/P	0.002
USP34	c.7217C>A	p.T2406N	0.005
ADAM32	c.1900G>A	p.G634R	0.001

Gene	cDNA	Protein	MAF
INTS1	3382 G/A	1128 V/M	0.0004
LRPPRC	1817 A/C	606 Q/P	novel
AMY2A	757 G/A	253 G/S	0.0004
MTREX	c.810T>G	p.I270M	0.00006
FGFR1OP2	170 C/T	57 A/V	0.000004
MLYCD	1424 A/G	475 K/R	0.001
KIF21A	3110 A/G	1037 N/S	0.002
DST	c.4970A>G	p.K1657R	0.003

Gene	cDNA		Protein	MAF
DRP2	c.439G>A	X-chromosomal	p.A147T	0.001
LGR6	c.1834G>A	homozygous	p.A612T	0.0002
KRT16	c.121C>T		p.R41C	0.00003
SMTN	c.1345G>A		p.G449S	0.007
PDE4DIP	c.5818C>T	compound heterozygous	p.R1940*	0.002
	c.5842A>G		p.K1948E	0.002
XIRP1	3247 A/C	heterozygous	1083 T/P	0.002

Gene	cDNA	Protein	MAF
AHDC1	c.3527C>T	p.P1176L	0.000008
ANK2	c.5410G>A	p.A1804T	0.0001
AKAP6	c.1280C>T	p.S427L	0.00001
MLST8	c.266C>T	p.A89V	0.0005
GAP43	c.782A>C	p.A261A	0.0001
MOV10	c.2543G>T	p.R848M	novel
NEIL2	c.77G>C	p.G26A	0.0004
NR1D1	c.1815G>T	p.K605N	novel
TFIP11	c.727G>A	p.V243M	0.00003
SYNE3	c.1633-6C>A		0.003
TRAP1	c.886C>T	p.R296C	0.00003

Figure 1 Pedigrees of the 5 families used for the initial exome sequencing with top candidate mutations listed in the adjacent boxes (MAF = minor allele frequency). Sequenced individuals: yellow; males: square; females: circles; patients: filled black.

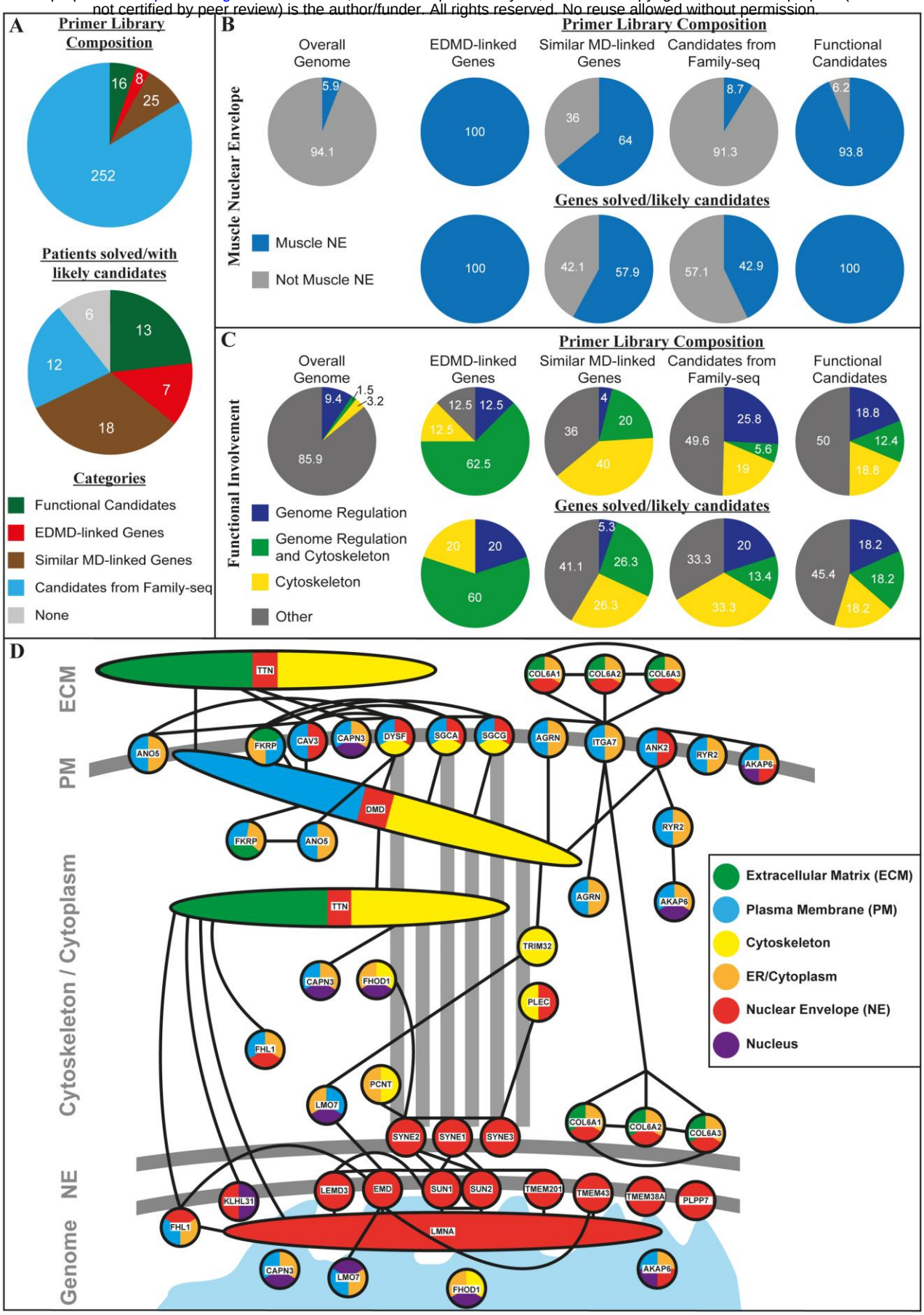


Figure 2 Primer library composition and gene ontology (GO) functions/localizations of all candidate genes from the four categories contributing to the primer library construction and

for the top candidates identified after primer library sequencing. **(A)** Composition of the primer library with number of genes from each of the four categories used in its construction (upper panel) and number of patients solved/with likely candidates from the different categories after primer library sequencing (lower panel). **(B)** Presence in muscle nuclear envelopes for the starting library in comparison to the overall genome (upper panel) and of the remaining candidate genes after primer library sequencing (lower panel) in percent (based on GO-localization terms and/or experimental evidence from appearance in nuclear envelope proteomics datasets[20,28]). **(C)** GO-terms for genome organization, cytoskeleton, and genome organization and cytoskeleton combined functions involvement for the starting library in comparison to the overall genome (upper panel) and of the remaining candidate genes after primer library sequencing (lower panel) in percent, showing an enrichment for the combined category in the top candidate alleles. **(D)** Interactive network of remaining candidate genes after library sequencing based on STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, https://string-db.org/) interactions (high confidence) showing that most candidates are linked to other candidates and that these form connections from the nuclear envelope to the plasma membrane. These connections are consistent with possible mechanotransduction from the extracellular region to the nuclear envelope being the core disrupted function in EDMD. Different described localizations of proteins are displayed by color-coding (based on GO-terms and/or experimental evidence through identification in muscle nuclear envelope proteomics datasets[20]).

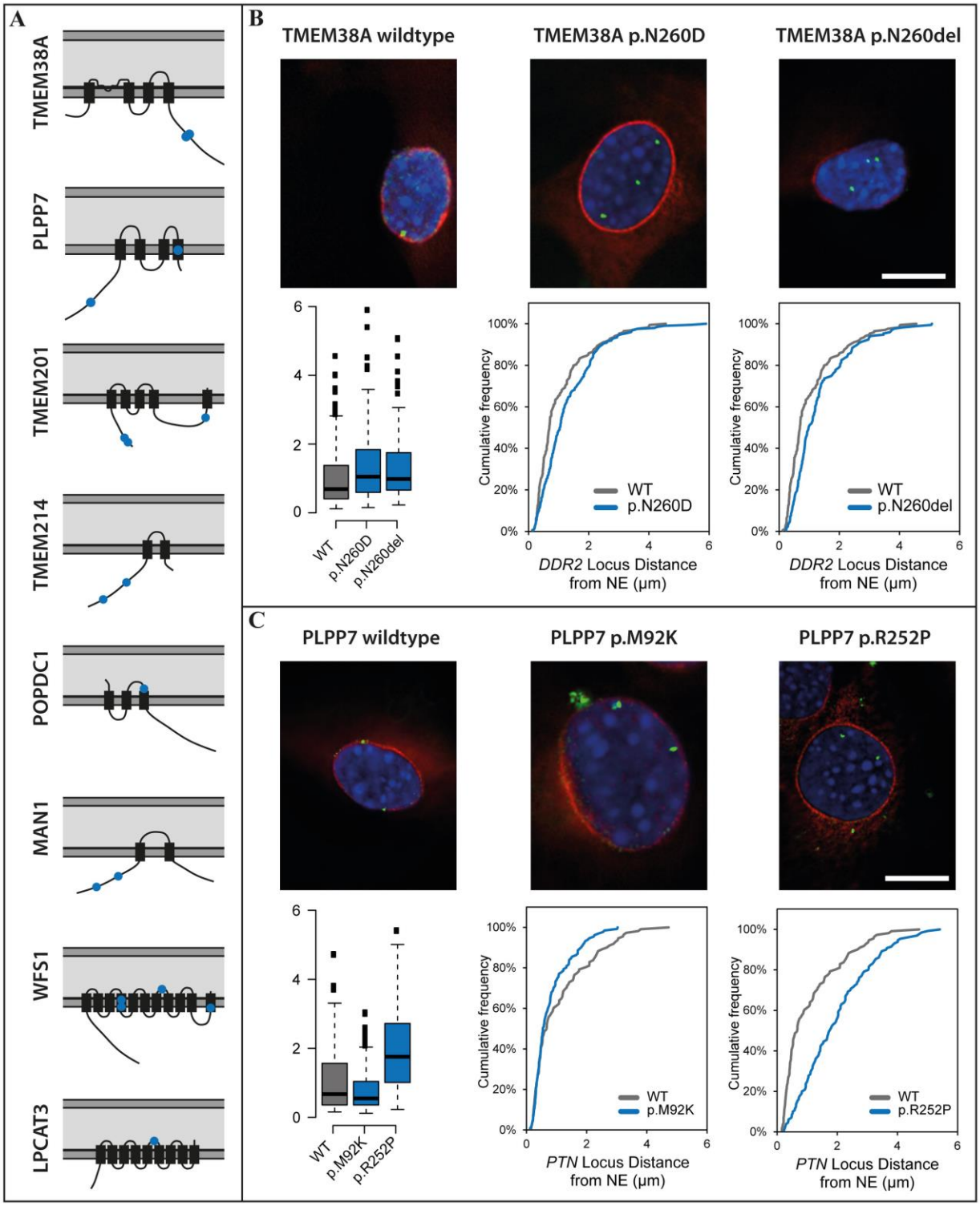


Figure 3 Mutations in muscle gene-repositioning NETs affect their ability to recruit genes to the nuclear envelope (NE). **(A)** Schematic presentation of the topology of further muscle NETs and their mutations identified by the primer library sequencing. The lipid bilayers of the nuclear envelope are shown in dark grey and the lumen of the nuclear envelope in light grey. Transmembrane segments are thicker black rectangles and point mutations identified are

shown in blue. The mutations identified are all positioned in nucleoplasmic regions where they could either interact with the genome or at transmembrane spans where they could disrupt protein topology and hence also genome interactions. **(B)** FISH showing the localization of the *DDR2* gene (green) in C2C12 mouse myoblasts upon the expression of RFP-tagged wild type and mutant TMEM38A that can be seen in both cases to target to the nuclear envelope (red, upper panel). The cumulative frequency of the distance of the gene loci to the NE for each mutation compared to the wild type is shown under each image of the cells expressing the mutant NETs and a whisker plot summary for the distance to the NE of all mutations is given in the lower left corner. Both mutations block the ability of TMEM38A to reposition the *DDR2* locus to the NE. **(C)** FISH showing the localization of the *PTN* gene (green) in C2C12 mouse myoblasts upon the expression of GFP-tagged wild type and mutant PLPP7 (red, upper panel). Cumulative frequency plots of the distance of the gene loci to the NE for each mutation and the summary for the distance to the NE of all mutations are given as in *B*. The mutations also affect the gene repositioning function of PLPP7.

Tables

Table 1: Solved Patients

ID	Sex	Comments/Clinical issues	CK in U/l	Mutation					Age of Onset	Age at Examination	Muscle wasting	Contractures	Heart involvement	Inheritance
				Gene	cDNA	Protein	known disease causing	MAF						
MD-1	**M**	EDMD phenotype with contractures, no cardiac arrhythmia, rigd spine syndrome	1200	***CAPN3***	**c.245C>T**	**p.P82L**	**ar, LGMD (CM080126)**	**0.00006**	6	16	yes	yes	no	ar
				CAPN3	**c.1043delG**	**p.G348Vfs*4**	**ar, LGMD (CD050834)**	**0.0003**						
MD-4	**F**	LGMD, contractures	3000	***CAPN3***	**c.1468C>T**	**p.R490W**	**ar, LGMD (CM950194)**	**0.00009**	20	28	yes	yes	no	ar
				CAPN3	**c.550delA**	**p.T184Rfs*36**	**ar, LGMD (CD951640)**	**0.0003**						
				SUN1	c.281G>A	p.R94H		0.0004						
MD-5	**F**	LGMD, contractures	7000	***CAPN3***	**c.1468C>T**	**p.R490W**	**ar, LGMD (CM950194)**	**0.00009**	8	23	yes	yes	no	ar
				CAPN3	**c.549delA**	**p.T184Rfs*36**	**ar, LGMD (CD951640)**	**0.0003**						
				AKAP6	c.2725C>A	p.P909T		0.00003						
				SYNE3	c.401T>G	p.V134G		0.000004						
MD-		contracures,	normal	***CAPN3***	**c.145C>T**	**p.R49C**	**ar, LGMD (CM076055)**	**0.00002**	cong	50	no	ye	no	spor

41		non-consanguineous, sporadic		***CAPN3***	**c.1821-1825del**	**p.R608Kfs*23**		**novel**	enital		t doc	s		adic
MD-11	**M**	EDMD phenotype with contractures, cardiac arrhythmia, WPW syndrome	1000	***LMNA***	**c.122G>A**	**p.R41H**		**n.a.**	2	15	yes	yes	yes	AD
				SYNE2	c.16178C>T	p.A5393V		0.001						
				TMEM43	c.934C>T	p.R312W		0.004						
MD-19	**F**	EDMD phenotype, pacer	200	***LMNA***	**c.1606delG**	**p.E536Kfs*12**		**novel**	35	55	yes	no	yes	unknown
				SYNE2	c.20161G>A	p.A6721T		0.001						
MD-37		contractures and cardiac conduction defect, non-consanguineous, sporadic	370	***LMNA***	**c.746G>A**	**p.R249Q**	**AD, EDMD (CM000737)**	**n.a.**	1	24	yes	yes	yes	sporadic
				TMEM201	c.44G>C	p.G15A		novel						
MD-13	**M**	mild LGMD, myalgia	2200	***DMD***	**c.9G>A**	**p.W3***	**xr, BMD (CM031161)**	**n.a.**	30	35	yes	no	no	unknown
MD-17	**M**	moderate muscle wasting, hIBMFTD Paget	700	***VCP***	**c.476G>A**	**p.R159H**	**AD, IBMPFD (CM057568)**	**0.000008**	60	72	yes	no	no	AD
				USP34	c.2963T>C	p.L988P		novel						
MD-12	**M**	myalgia, proximal weakness, arrhythmia	700	***COL6A2***	**c.2795C>T**	**p.P932L**	**AD, BTHLM (CM076126)**	**0.002**	45	54	yes	no	yes	AD
				VCP	**c.17A>T**	**p.D6V**		**novel**						
				SYNE2	c.2669C>A	p.T890K		novel						
				SYNE2	c.2647-2A>T			0.00003						
				XIRP1	c.4648A>T	p.I1550F		novel						
				XIRP1	c.3612G>T	p.W1204C		novel						

MD-6	**F**	EDMD phenotype with mild contractures, Polyglucosan bodies	600	***GBE1***	**c.691+2T>C**	**splice donor**	**AD, GSD4 (CS100318)**	**0.001**	27	72	yes	yes	no	unknown
				SYNE2	c.2647-2A>T			0.00003						
				WFS1	c.1316T>G	p.F439C		0.00009						
MD-22	F	distal myopathy	400	***GBE1***	**c.1382T>C**	**p.V461A**		**novel**	16	34	yes	no	no	AD
				TTN	**c.107635C>T**	**p.Q35879***		**0.00002**						
				TTN	**c.22027C>T**	**p.Q7343***		**n.a.**						
				PLPP7	c.275T>A	p.M92K		novel						
				USP34	c.7411C>T	p.H2471Y		0.00003						
MD-25	F			***GBE1***	**c.2017G>A**	**p.A673T**		**0.005**	4 mths	17	yes	no	yes	unknown
				TMEM38A	c.739G>A	p.V247M		0.001						
MD-34		contractures and mild cardiomyopathy, non-consanguineous, sporadic	240	***GBE1***	**c.839G>A**	**p.G280D**		**0.004**	childhood	39	not doc	yes	yes	sporadic
				DYSF	c.5698-5699del	p.S1900Qfs*14	Miyoshi myopathy (CD982604)	0.00004						
				TMEM43	c.934C>T	p.R312W		0.01						
MD-18	M	EDMD phenotype	n.d.	***TTN***	**c.40787-2A>G**			**novel**	unknown	unknown	yes	yes	yes	unknown
				TTN	**c.9047del**	**p.M3016***		**novel**						
				TTN	**c.72409T>C**	**p.S24137P**		**novel**						
				SYNE1	c.16843G>A	p.E5615K		novel						
MD-44		contractures and cardiac conduction defect, non-consanguineous, sporadic	2700	***TTN***	**c.107377+1G>A**			**0.00001**	20	41	not doc	yes	yes	sporadic
				TTN	**c.104952A>C**	**p.E34984D**		**0.00002**						
				TTN	**c.87529A>T**	**p.K29177***		**novel**						
				SYNE1	c.4562G>A	p.R1521Q		0.00004						
				SUN1	c.608C>T	p.A203V		0.002						

				TMEM201	c.1789G>A	p.G597S		0.00001						
MD-8	M	EDMD-like rigid-spine syndrome, Bethlem	500	***COL6A1***	**c.2166dup**	**p.I723Hfs*7**		**novel**	3 mths	20	yes	yes	no	ar
				COL6A1	**c.1784delA**	**p.E595Gfs*7**		**0.00000**						
				COL6A1	**c.1786A>G**	**p.I596V**		**0.00000**						
				WFS1	c.1675G>A	p.A559T		0.001						
MD-20	M		15000	***ANO5***	**c.191dup**	**p.N64Kfs*15**	**ar, LGMD (CI101059)**	**0.001**	25	47	yes	no	no	ar
				ANO5	**c.1391C>A**	**p.A464D**	**ar, LGMD (CM137896)**	**n.a.**						
MD-21	M		200	***DYSF***	**c.3065G>A**	**p.R1022Q**	**ar, LGMD (CM090628)**	**0.014**	20	23	yes	no	no	unknown
				DYSF	**c.3992G>T**	**p.R1331L**	**ar, Dysferlinopathy (CM103814)**	**0.016**						
MD-23	M	axial weakness	500	***POMT1***	**c.1545C>G**	**p.S515R**		**0.0006**	10	20	yes	no	no	unknown
				POMT1	**c.1838G>A**	**p.R613H**		**0.0006**						
				INTS1	c.2395T>C	p.V770A		novel						
MD-43	F	contracures, AD family history	500-800	***CAV3***	**c.136G>A**	**p.A46T**	**AD, LGMD (CM012082)**	**n.a.**	child hood	34	not doc	yes	no	AD

Dark green: known disease associated genes (EDMD or similar MDs) with likely disease causing mutation (category 2 and 3).

Light green: known disease associated genes (EDMD or similar MDs) with unlikely disease causing mutation or two genes of similar likelihood to be the causative disease allele (category 2 and 3).

Yellow: functional candidate gene mutations (category 1).

Purple: mutations in genes from the family sequencing (category 4).

Table 2: Patients with Candidate Genes

ID	Sex	Comments/Clinical issues	CK in U/l	Mutation					Age of Onset	Age at Examination	Muscle wasting	Contractures	Heart involvement	Inheritance
				Gene	cDNA	Protein	known disease causing	MAF						
MD-2	M	EDMD phenotype	1500-2000	*DYSF*	c.1369G>A	p.E457K	ar, MMD (CM074148)	0.008	2	16	yes	yes	no	ar
				Trim32	c.1802A>G	p.H601R		novel						
MD-3	F	distal LGMD	200-800	*PLEC*	c.10562C>T	p.T3521M		0.00003	38	58	yes	no	no	unknown
				COL6A2	c.1769C>T	p.T590M	AD, BTHLM (CM1310895)	0.0004						
				ATP12A	c.9+1G>A			0.001						
				XIRP1	c.3763C>T	p.P1255S		0.002						
				USP34	c.19G>T	p.D7Y		novel						
MD-7	F	EDMD phenotype	350	*FKTN*	c.559G>A	p.G187S		0.00002	2	10	yes	yes	not doc	ar
				KCNJ12	c.109C>A	p.H37N		novel						
				ANK2	c.11791G>A	p.E3931K		0.003						
MD-9	F	Affected father with pacemaker	350	*COL6A3*	c.9508G>A	p.G3170R		0.0002	28	48	yes	no	not doc	AD
				COL6A3	c.1024G>A	p.V342M		0.001						
				Tmem214	c.536G>A	p.R179H		0.0002						
				SYNE3	c.2024G>A	p.R675Q		0.0004						
MD-10	M	EDMD phenyotpe, Father pacemaker	1000	*COL6A2*	c.2102C>A	p.T701N		novel	5	46	yes	yes	no	AD
				AGRN	c.1123G>T	p.A375S		0.005						
				PLEC	c.5638G>A	p.A1880T		0.0005						
				AKAP6	c.2663C>A	p.T888N		0.00002						

				ATP12A	c.1897G>A	p.G633S		0.0002						
				USP34	c.4387A>G	p.S1463G		novel						
MD-14	M	EMD phenotype mild	2000	*DYSF*	c.3191_3196dup	p.A1064_E1065dup	LGMD (CI105954)?	0.039			yes	no		unknown
				PLEC	c.12601G>A	p.E4201K		0.001						
				LEMD3	c.263G>T	p.G88V		0.000000						
				RYR3	c.1508G>C	p.G503A		0.000004						
				ANK2	c.6228G>T	p.K2076N		0.001						
MD-15	M	EDMD phenotype mild	400	*WFS1*	c.1294C>G	p.L432V		0.004	4	6	yes	no	no	unknown
MD-16	M	Hemiatrophia totalis, Parry-Romberg	245	*FKRP*	c.456C>G	p.S152R		0.0005	12	43	yes	no	no	unknown
MD-24		Bethlem phenotype, recently confirmed *STIM1* mutation		*AGRN*	c.4966C>T	p.R1656W		0.001			yes	yes		unknown
MD-26	F	SPS scapuloperoneal syndrome	200	*PLEC*	c.2648G>A	p.R883H		0.001	45	70	yes	no	no	ar
				WFS1	c.2611G>A	p.V871M		0.008						
				AKAP6	c.3335G>A	p.G1112E		0.001						
				FHOD1	c.2714G>A	p.R905Q		0.001						
MD-27	F	distal myopathy	300	*COL6A3*	c.8009C>T	p.A2670V		0.0002	15	25	yes	no	no	AD
				SYNE1	c.19729C>T	p.R6577W		0.00007						
				TTN	c.73705G>C	p.V24569L		0.0002						
				RYR3	c.7249A>G	p.I2417V		0.001						
MD-28	M	distal myopathy:	200	*CAPN3*	c.1678A>G	p.T560A		0.0000	6	48	ye	no	no	un

		Nemaline-Myopathy, regional ichthyosis						2			s			known
				PLEC	c.1715C>T	p.A572V		0.00002						
				Lmo7	c.2797G>A	p.A933T		0.0001						
				KCNJ12	c.127C>T	p.R43C		0.00004						
				KCNJ12	c.758dupT	p.F255Lfs*19		0.001						
MD-29	F	LGMD	up to 10.000	*SYNE1*	c.18137C>T	p.T6046M		0.00002	33	47	yes	no	no	unknown
				LEMD3	c.689G>C	p.R230T		novel						
				INTS1	c.820C>T	p.R274C		0.001						
				TTN	c.39624A>C	p.K13208N		n.a.						
MD-30	M	distal myopathy	300	*DYSF*	c.1877C>T	p.M626T		0.001	57	73	yes	no	no	AD
				DYSF	c.3191_3196dup	p.A1064_E1065dup	LGMD (CI105954)?	0.039						
				PLEC	c.424C>T	p.R142W		0.0003						
				MAGI1	c.977C>A	p.T326K		novel						
MD-31	F	ad MFM	500	*SYNE2*	c.12659A>C	p.Q4220P		0.000004	32	46	yes	no	no	AD
				ATP12A	c.1663A>G	p.T555A		0.0002						
				XIRP1	c.1055G>A	p.R352Q		0.0002						
				INTS1	c.4282C>T	p.P1399L		0.000007						
MD-32	M	LGMD asymmetric, no heart involvement	500	*DYSF*	c.3191_3196dup	p.A1064_E1065dup	LGMD (CI105954)?	0.039	40	70	yes	no	no	unknown
				COL6A3	c.3852C>A	p.F1284L		0.001						
				PLEC	c.7678G>A	p.A2560T		0.001						
				PLEC	c.4697C>T	p.S1566L		0.00005						
				TMEM38	c.778G>A	p.D260N		0.002						

				A										
				PLPP7	c.755G>C	p.R252P		novel						
				INTS1	c.5707C>T	p.P1874L		0.000006						
MD-33	M	SMA-like		*TTN*	c.92755C>T	p.R30919W		0.00002	24	65	yes	no	no	nn
				PCNT	c.4571C>G	p.P1524R		0.002						
				WDR43	c.1075A>G	p.I359V		0.001						
				ANK2	c.11791G>A	p.E3931K		0.003						
MD-35		contractures and mild cardiomyopathy, non-consanguineous, sporadic	4000	*DYSF*	c.509C>A	p.A170E	ar, Dysferlinopathy (CM053208)	0.004	20	27	not doc	yes	yes	sporadic
				COL6A3	c.4156G>A	p.E1386K	AD, BTHLM (CM050230)	0.002						
				AGRN	c.1528G>A	p.G510S		0.008						
MD-36		contractures and cardiac conduction defect, non-consanguineous, sporadic	500-1000	*SYNE2*	c.17191C>T	p.R5731C		0.00004	20s	66	not doc	yes	yes	sporadic
MD-38		contractures and cardiac conduction defect, non-consanguineous, sporadic	normal	*LPCAT3*	c.805C>T	p.R269C		0.000004	congenital	56	not doc	yes	yes	sporadic
				XIRP1	c.3442G>A	p.V1148M		0.001						
MD-39		contracures, non-consanguineous, sporadic	1300	*DMD*	c.1252A>T	p.T418S		0.002	early childhood	21	no	yes	no	unknown
				DYSF	c.3967C>G	p.Q1323E		0.001						
				FKTN	c.373G>A	p.G125S	Walker-Warburg syndrome?	0.037						
				COL6A3	c.4510C>T	p.R1504W		0.001						

				PDE4DIP	c.5486C>T	p.S1829F		0.0001						
				PDE4DIP	c.4063C>T	p.R1355*		0.0001						
				ANK2	c.4744C>T	p.R1582W		0.002						
MD-40		contracures, non-consanguineous, sporadic	normal	*Col6A3*	c.3419C>T	p.T1140M		0.0004	5	46	yes	yes	no	sporadic
				SUN1	c.278A>C	p.Q93P		0.004						
				TTN	c.93768_93769dup	p.K31257Lfs*6		novel						
				TTN	c.107840T>A	p.I35947N		0.000004						
				ATP12A	c.349A>G	p.I117V		0.00001						
				ANK2	c.11231C>A	p.T3744N		0.001						
				INTS1	c.4969C>T	p.R1657C		0.00007						
MD-42		contractures and cardiac arrhythmia, non-consanguineous, sporadic	3000	*DYSF*	c.3065G>A	p.R1022Q	ar, LGMD (CM090628)	0.014	adult	42	not doc	yes	yes	sporadic
				COL6A2	c.1552C>T	p.P518S		0.003						
				COL6A3	c.4727G>A	p.R1576Q		0.004						
				WDR43	c.366T>G	p.S122R		0.0003						
MD-45		contractures and possible mild cardiac involvement, non-consanguineous, sporadic	1000	*DYSF*	c.3992G>T	p.R1331L	Dysferlinopathy (CM103814)	0.016	20s	46	not doc	yes	yes	sporadic
				TRIM32	c.558G>C	p.Q186H		0.001						
				SUN1	c.335C>T	p.T112M		0.0002						
				BVES	c.275A>G	p.D92G		novel						
MD-46				*SGCG*	c.792C>G	p.I264M		0.00006	3 mths	5	yes	yes	no	unknown
				COL6A1	c.347G>A	p.S116N	AD, BTHLM (CM050211)?	0.031						
				ITGA7	c.824G>A	p.R275H		0.033						
				TMEM20	c.52G>A	p.G18S		novel						

				1										
MD-47				*COL6A2*	c.316G>A	p.E106K	AD, BTHLM (CM050217)?	0.002	1	28	yes	yes	yes	unknown
				RYR2	c.1939C>T	p.R647C		0.0001						
MD-48				*POMT1*	c.1233C>A	p.D411E		0.035	5	21	yes	yes	no	unknown
				RYR2	c.2444C>T	p.P815L		0.00003						
				USP34	c.3938G>T	p.G1313V		novel						
				ANK2	c.9854T>C	p.I3285T		0.009						
MD-49				*SGCA*	c.371T>C	p.I124T	LGMD	0.00006	2	9	yes	yes	no	AD
				SGCG	c.320C>T	p.S107L		0.000008						
				COL6A1	c.347G>A	p.S116N	AD, BTHLM (CM050211)?	0.031						
				RYR3	c.6698G>A	p.R2233Q		0.001						
				RYR3	c.9254C>G	p.P3085R		0.0004						
				RYR2	c.3407C>T	p.A1136V		0.007						
MD-50				*DYSF*	c.509C>A	p.A170E	Dysferlinopathy (CM053208)	0.004	4 mths	13	yes	no	yes	unknown
				COL6A1	c.347G>A	p.S116N	AD, BTHLM (CM050211)?	0.031						
				COL6A2	c.2558G>A	p.R853Q	AD, BTHLM (CM050225)?	0.004						
				AGRN	c.5201G>A	p.R1734H		0.004						
				TMEM38A	c.778G>A	p.D260N		0.002						
				XIRP1	c.802_805dup	p.A269Vfs*6		novel						
MD-51				*POMT1*	c.1233C>A	p.D411E		0.035	2	14	yes	no	yes	unknown
				COL6A1	c.347G>A	p.S116N	AD, BTHLM (CM050211)?	0.031						
MD-52				*DMD*	c.3595G>A	p.E1199K		novel	6	24	yes	yes	no	ar
				SYNE2	c.1823A>G	p.E608G		0.000004						

				TMEM43	c.351dup	p.H118Afs*11		0.00003						
MD-53				*SYNE1*	c.17347G>C	p.E5783Q		0.000004	8 mths	5	yes	no	no	unknown
MD-54				*FKTN*	c.116C>T	p.R56C		0.01	8	19	yes	yes	no	unknown
				AGRN	c.1528G>A	p.G510S		0.008						
				PLEC	c.8497_8516dup	p.L2839*		novel						
				PLEC	c.11708G>C	p.R3903P		0.000004						
				PLEC	c.8917G>A	p.D2973N		0.0004						
				SYNE1	c.25403G>A	p.R8468H		0.0002						
				TMEM214	c.787G>A	p.G263S		0.001						
MD-55				*AGRN*	c.5353G>A	p.D1785N		0.006	8	16	yes	yes	no	unknown
MD-56	M	FSHD1	400 U/L	*TMEM38A*	c.778_780del	p.D260del		0.005	15	74	yes	no	no	unknown
				KLHL31	c.1817G>A	p.G606D		novel						
				RYR3	c.5356G>A	p.G1786S		0.001						

Light green: known disease associated genes (EDMD or similar MDs) with unlikely disease causing mutation or two genes of similar likelihood to be the causative disease allele (category 2 and 3).

Yellow: functional candidate gene mutations (category 1).

Purple: mutations in genes from the family sequencing (category 4).

Supplementary Material