



Research paper

A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism



Peter Meinke^{a,b}, Alastair R.W. Kerr^a, Rafal Czapiewski^a, Jose I. de las Heras^a, Charles R. Dixon^a, Elizabeth Harris^c, Heike Kölbel^d, Francesco Muntoni^{e,f}, Ulrike Schara^d, Volker Straub^c, Benedikt Schoser^b, Manfred Wehnert^g, Eric C. Schirmer^{a,*}

^a Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, UK

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

^c John Walton Muscular Dystrophy Research Centre, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

^d Department of Pediatric Neurology, Developmental Neurology and Social Pediatrics, University of Essen, Germany

^e Dubowitz Neuromuscular Centre, University College London Great Ormond Street Institute of Child Health, London, UK

^f 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

^g Institute of Human Genetics, University of Greifswald (retired), Greifswald, Germany

ARTICLE INFO

Article History:

Received 14 October 2019

Revised 18 November 2019

Accepted 26 November 2019

Available online 17 December 2019

Keywords:

Emery-Dreifuss muscular dystrophy
Nuclear envelope
Nuclear envelope transmembrane protein
primer library
Orphan disease

ABSTRACT

Background: As genome-wide approaches prove difficult with genetically heterogeneous orphan diseases, we developed a new approach to identify candidate genes. We applied this to Emery-Dreifuss muscular dystrophy (EDMD), characterised by early onset contractures, slowly progressive muscular wasting, and life-threatening heart conduction disturbances with wide intra- and inter-familial clinical variability. Roughly half of EDMD patients are linked to six genes encoding nuclear envelope proteins, but the disease mechanism remains unclear because the affected proteins function in both cell mechanics and genome regulation.

Methods: A primer library was generated to test for mutations in 301 genes from four categories: (I) all known EDMD-linked genes; (II) genes mutated in related muscular dystrophies; (III) candidates generated by exome sequencing in five families; (IV) functional candidates – other muscle nuclear envelope proteins functioning in mechanical/genome processes affected in EDMD. This was used to sequence 56 unlinked patients with EDMD-like phenotype.

Findings: Twenty-one patients could be clearly assigned: 18 with mutations in genes of similar muscular dystrophies; 3 with previously missed mutations in EDMD-linked genes. The other categories yielded novel candidate genes, most encoding nuclear envelope proteins with functions in gene regulation.

Interpretation: Our multi-pronged approach identified new disease alleles and many new candidate EDMD genes. Their known functions strongly argue the EDMD pathomechanism is from altered gene regulation and mechanotransduction due to connectivity of candidates from the nuclear envelope to the plasma membrane. This approach highlights the value of testing for related diseases using primer libraries and may be applied for other genetically heterogeneous orphan diseases.

Funding: The Wellcome Trust, Muscular Dystrophy UK, Medical Research Council, European Community's Seventh Framework Programme "Integrated European –omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)".

© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

1. Introduction

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

* 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.

E-mail address: e.schirmer@ed.ac.uk (E.C. Schirmer).

Research in context

Evidence before this study

Emery-Dreifuss muscular dystrophy (EDMD) is a genetically heterogeneous orphan disease with clinical variability presenting even in family members carrying the same mutation. The lack of large pedigrees in combination with its genetic heterogeneity (~50% of patients are solved with mutations in 6 genes), clinical variability, several known modifier genes, limited patient numbers, and an unsolved pathomechanism make the usage of genome-wide association sequencing studies ineffective. EDMD also presents a conundrum in that all previously known alleles are widely expressed nuclear envelope (NE) proteins, so how a muscle focused pathology is achieved remains obscure. Moreover the proteins involved have known functions in nuclear/cellular mechanical stability, genome regulation, and cell cycle regulation, so that the EDMD pathomechanism remains unclear.

Added value of this study

The new fully solved and candidate alleles encode proteins that form a physical connection from the plasma membrane to the nuclear envelope, further elaborating genome regulation and mechanosignal transduction as the pathomechanism and supporting the validity of new candidates. This connectivity together with that several of the new candidate alleles are muscle-specific resolves the conundrum of how a muscle-focused pathology could come from widely-expressed proteins. The range of proteins that can be disrupted through these connections moreover can help explain the clinical variability and also the overlaps to similar muscular dystrophies. Notably, the larger set of candidates from the exome sequencing step in our approach likely contains also modifying alleles that may be revealed with further use of the primer library.

Implications of all the available evidence

If this patient cohort is representative of the general patient population this approach likely solves >90% of the ~50% remaining genetically unsolved patients, and in doing so strongly implicates altered gene regulation and mechanotransduction as its pathomechanism. Furthermore, in addition to the potential application of this approach to other genetically heterogeneous orphan diseases, the primer library once generated provides a better and cheaper sequencing strategy for newly diagnosed patients compared to traditional Sanger sequencing of one linked gene at a time. The EDMD patient association has noted a problem with the time from clinical presentation to diagnosis that will be resolved with the wider application of this primer library and allow earlier interventions to improve disease management.

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 [14], each with distinct tissue-specific pathologies: for example different lamin A mutations cause muscular dystrophies, neuropathy, lipodystrophy, and multisystemic disorders. 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 [15].

We previously identified several muscle-specific nuclear envelope transmembrane proteins (NETs) [16]. 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 principally hypothesized EDMD pathomechanisms: mechanical instability and disruption of gene expression. NETs Tmem214 and KLHL31 track with microtubules on the nuclear surface [16] while NET5/Samp1 contributes actin and centrosome interactions [17]. NETs Tmem38A, WFS1, NET39/PLPP7 and, again, Tmem214 and NET5/Samp1 all affect 3D gene positioning with corresponding effects on gene expression [18,19]. 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 adipogenesis. 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 [18]. 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 (Fig. 1a).

2. Materials and methods

2.1. Patient materials

All patient DNA used for sequencing was obtained from blood samples. RNA was obtained from deltoid muscle from the family 1 index patient and matching control. The sources of patient samples were: 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.

2.2. Ethical approval and consent to participate

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 118,436 and MTA CT-2166, that of the Rare Diseases biological samples biobank for research to

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]. 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 [7–11]). Variants in desmin and the nuclear envelope proteins SUN1 and SUN2 have been reported to modify the EDMD phenotype [10,12]. Roughly half of clinically diagnosed patients remain unlinked [13].

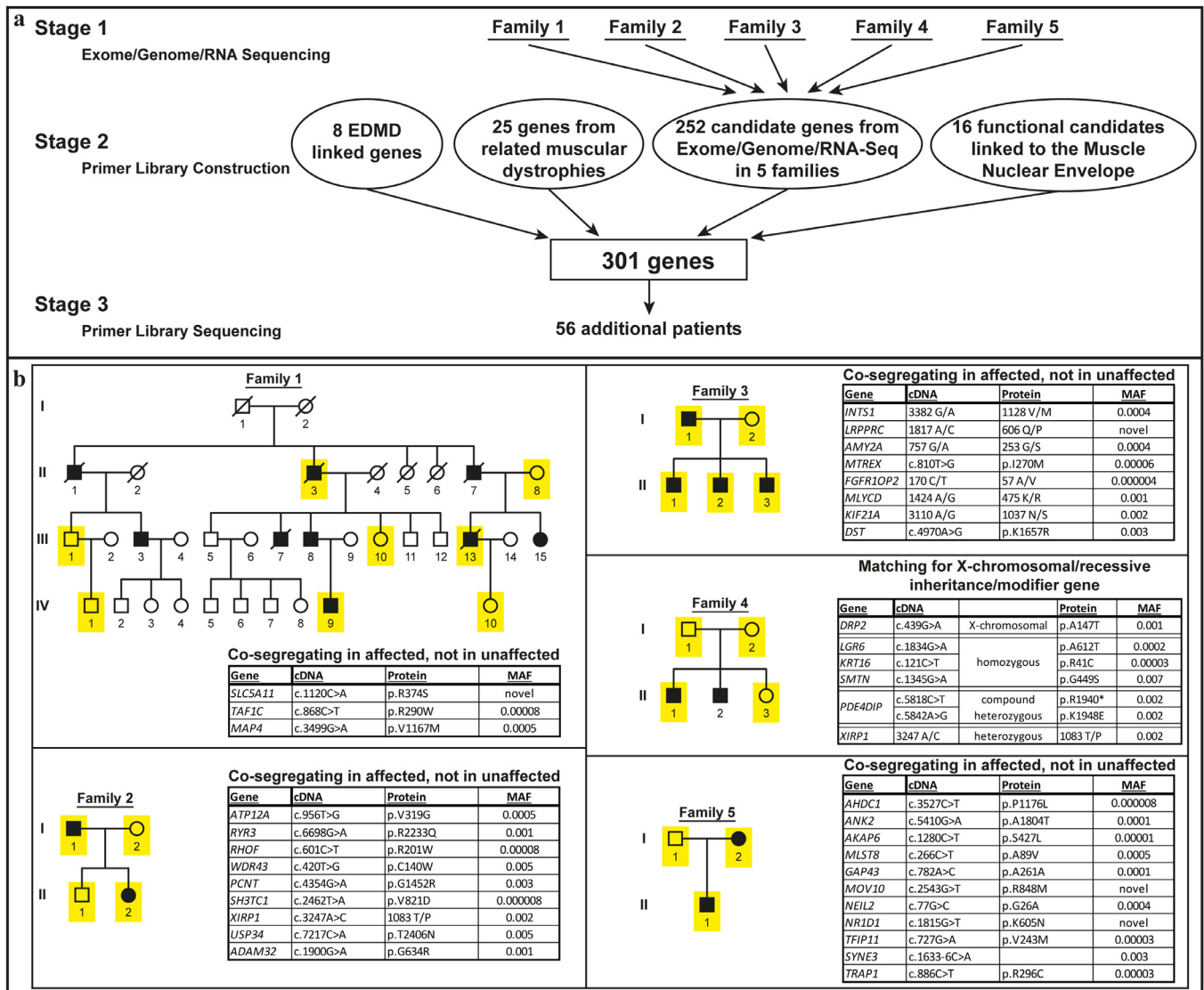


Fig. 1. Methodology and top candidate mutations from exome sequencing. (a) Flow chart describing the study methodology. (b) Pedigrees of the 5 families used for the initial exome sequencing with top candidate mutations listed in the adjacent boxes (MAF = minor allele frequency). The presented top candidate mutations were identified in all sequenced affected family members, but not in unaffected family members. Sequenced individuals: yellow; males: square; females: circles; patients: filled black. The presented top candidate mutations were identified in all sequenced affected family members, but not in unaffected family members.

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 (WoRES) with REC reference 15/WS/0069 and IRAS project ID 177,946.

2.3. 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, Norcross, USA).

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

2.4. Fluorescence *in situ* hybridisation

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, 50 U/mL penicillin and 10 mg/mL streptomycin. Fluorescent *in situ* hybridisation (FISH) experiments were performed as described in [20].

2.5. 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 3427,092 was obtained with a range from 442,125 to 7066,507 using 125 base paired-end sequencing on a Hi-Seq2500.

2.6. Bioinformatics and analysis

Variant analysis was performed using the Genome Analysis toolkit [GATK] v2.7–2 [21] and picard tools v1.74 (<http://broadinstitute>.

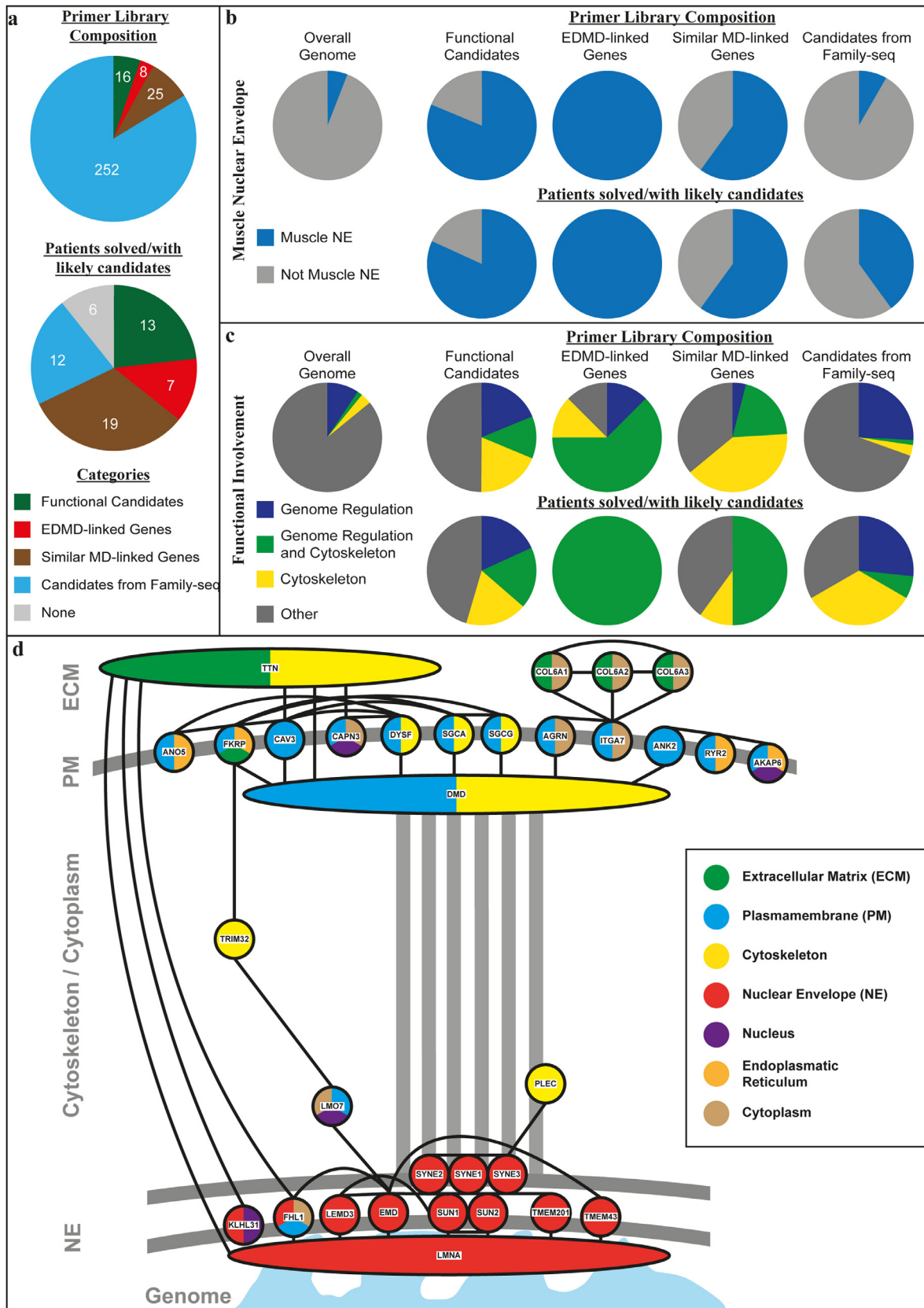


Fig. 2. Primer library composition and gene ontology (GO) functions/localisations of all candidate genes from the four categories contributing to the primer library construction and for the top candidates identified after primer library sequencing. (a) Composition of the primer library with number of 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-localisation terms and/or experimental evidence from appearance in nuclear envelope proteomics datasets^{20,28}). (c) GO-terms for genome organisation, cytoskeleton, and genome organisation and cytoskeleton combined functions involvement for the starting library in comparison to the overall genome (upper panel) and of the remaining

github.io/picard/) using GATK Best Practices recommendations [22,23] against human genome assembly hg19. The allele frequencies of variants were cross-referenced with gnomAD version 2.1 [24] using both the genome and exome datasets.

RNA-Seq: STAR v2.1.1 [25] was used to map reads to the hg19 reference genome, samtools v0.1.19 [26] was used for file conversion. Deepools v1.5.1 [27] was used for downstream analysis.

3. Results

3.1. 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. 1b), 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 on 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 from the five families.

3.2. 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 3427,092 reads per patient. The data were analysed 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 [28]. 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.

Category III genes from exome sequencing were elevated to "likely-causative" candidates if also 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).

3.3. Nuclear envelope links

All previously linked EDMD genes encode nuclear envelope proteins. The functional candidates were also biased towards genes

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 localisations of proteins are displayed by colour-coding (based on GO-terms and/or experimental evidence through identification in muscle nuclear envelope proteomics datasets²⁰).

Table 1
Solved Patients.

ID	Sex	Comments/Clinical issues	CK in U/I	Mutation					Age of Onset	Age at Examination	Muscle wasting	Contractures	Heart involvement	Inheritance	
				Gene	cDNA	Protein	known disease causing	MAF							SIFT
MD-1	M	EDMD phenotype with contractures, no cardiac arrhythmia, rigid 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	tolerated (0.96)						
MD-4	F	LGMD, contractures	3000	CAPN3	c.1468C>T	p.R490W	ar, LGMD (CM950194)	0.00009	deleterious (0)	20	28	yes	yes	no	ar
				CAPN3	c.550delA	p.T184Rfs*36	ar, LGMD (CD951640)	0.0003	deleterious (0)						
				SUN1	c.281G>A	p.R94H		0.0004	tolerated (0.34)						
MD-5	F	LGMD, contractures	7000	CAPN3	c.1468C>T	p.R490W	ar, LGMD (CM950194)	0.00009	deleterious (0)	8	23	yes	yes	no	ar
				CAPN3	c.549delA	p.T184Rfs*36	ar, LGMD (CD951640)	0.0003	deleterious (0)						
				AKAP6	c.2725C>A	p.P909T		0.00003	tolerated (0.11)						
				SYNE3	c.401T>G	p.V134G		0.000004	deleterious (0.02)						
MD-41		contractures, non-consanguineous, sporadic	normal	CAPN3	c.145C>T	p.R49C	ar, LGMD (CM076055)	0.00002	---	congenital	50	not documented	yes	no	sporadic
				CAPN3	c.1821-1825del	p.R608Kfs*23		novel	---						
MD-11	M	EDMD phenotype with contractures,	1000	LMNA	c.122G>A	p.R41H		n.a.	deleterious (0)	2	15	yes	yes	yes	AD
				SYNE2	c.16178C>T	p.A5393V		0.001	tolerated						

(continued)

		cardiac arrhythmia, WPW syndrome						(1)							
				<i>TMEM43</i>	c.934C>T	p.R312W		0.004	---						
MD-19	F	EDMD phenotype, pacemaker	200	<i>LMNA</i>	c.1606delG	p.E536Kfs*12		novel	---	35	55	yes	no	yes	unknown
				<i>SYNE2</i>	c.20161G>A	p.A6721T		0.001	tolerated (1)						
MD-37		contractures and cardiac conduction defect, non-consanguineous, sporadic	370	<i>LMNA</i>	c.746G>A	p.R249Q	AD, EDMD (CM000737)	n.a.	deleterious (0)	1	24	yes	yes	yes	sporadic
				<i>TMEM201</i>	c.44G>C	p.G15A		novel	deleterious (0)						
MD-13	M	mild LGMD, myalgia	2200	<i>DMD</i>	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	<i>VCP</i>	c.476G>A	p.R159H	AD, IBMPFD (CM057568)	0.000008	tolerated (0.17)	60	72	yes	no	no	AD
				<i>USP34</i>	c.2963T>C	p.L988P		novel	deleterious (0)						
MD-12	M	myalgia, proximal weakness, arrhythmia	700	<i>COL6A2</i>	c.2795C>T	p.P932L	AD, BTHLM (CM076126)	0.002	tolerated (0.39)	45	54	yes	no	yes	AD
				<i>VCP</i>	c.17A>T	p.D6V		novel	tolerated (0.09)						
				<i>SYNE2</i>	c.2669C>A	p.T890K		novel	tolerated (0.09)						
				<i>SYNE2</i>	c.2647-2A>T			0.00003	---						
				<i>XIRP1</i>	c.4648A>T	p.I1550F		novel	deleterious (0)						
				<i>XIRP1</i>	c.3612G>T	p.W1204C		novel	deleterious (0.02)						
MD-	F	EDMD	600	<i>GBE1</i>	c.691+2T>	splice donor	AD, GSD4 (CS100318)	0.001	---	27	72	yes	yes	no	unknown

(continued)

6	phenotype with mild contractures, Polyglucosan bodies			C														know
				<i>SYNE2</i>	c.2647-2A>T				0.00003	---								n
				<i>WFS1</i>	c.1316T>G	p.F439C			0.00009	deleterious (0)								
MD-22	F	distal myopathy	400	<i>GBE1</i>	c.1382T>C	p.V461A			novel	tolerated (0.25)								
				<i>TTN</i>	c.107635C>T	p.Q35879*			0.00002	---								
				<i>TTN</i>	c.22027C>T	p.Q7343*			n.a.	---	16	34	yes	no	no			AD
				<i>PLPP7</i>	c.275T>A	p.M92K			novel	deleterious (0)								
				<i>USP34</i>	c.7411C>T	p.H2471Y			0.00003	deleterious (0.03)								
MD-25	F			<i>GBE1</i>	c.2017G>A	p.A673T			0.005	tolerated (0.11)	4 mths	17	yes	no	yes			unknown
				<i>TMEM38A</i>	c.739G>A	p.V247M			0.001	tolerated (0.67)								n
MD-34		contractures and mild cardiomyopathy, non-consanguineous, sporadic	240	<i>GBE1</i>	c.839G>A	p.G280D			0.004	deleterious (0)								
				<i>DYSF</i>	c.5698-5699del	p.S1900Qfs*14	Miyoshi myopathy (CD982604)		0.00004	---	childhood	39	not doc	yes	yes			sporadic
				<i>TMEM43</i>	c.934C>T	p.R312W			0.01	---								
MD-18	M	EDMD phenotype	n.d.	<i>TTN</i>	c.40787-2A>G				novel	---								
				<i>TTN</i>	c.9047del	p.M3016*			novel	---	unknown	unknown	yes	yes	yes			unknown
				<i>TTN</i>	c.72409T>C	p.S24137P			novel	---	known	known	yes	yes	yes			known
				<i>SYNE1</i>	c.16843G>A	p.E5615K			novel	tolerated (0.08)								
MD-44		contractures and cardiac	2700	<i>TTN</i>	c.107377+1G>A				0.00001	---	20	41	not doc	yes	yes			sporadic

(continued)

		conduction defect, non-consanguineous, sporadic		TTN c.104952A>C	p.E34984D		0.00002	---									
				TTN c.87529A>T	p.K29177*		novel	---									
				SYNE1 c.4562G>A	p.R1521Q		0.00004	tolerated (0.08)									
				SUN1 c.608C>T	p.A203V		0.002	tolerated (0.31)									
				TMEM201 c.1789G>A	p.G597S		0.00001	deleterious (0.02)									
MD-8	M	EDMD-like rigid-spine syndrome, Bethlehem	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	tolerated (0.57)									
				WFS1 c.1675G>A	p.A559T		0.001	tolerated									
								(0.8)									
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.	deleterious (0.02)									
MD-21	M		200	DYSF c.3065G>A	p.R1022Q	ar, LGMD (CM090628)	0.014	---		20	23	yes	no	no	known		
				DYSF c.3992G>T	p.R1331L	ar, Dysferlinopathy (CM103814)	0.016	tolerated (0.51)									
MD-23	M	axial weakness	500	POMT1 c.1545C>G	p.S515R		0.0006	tolerated (0.27)		10	20	yes	no	no	known		
				POMT1 c.1838G>A	p.R613H		0.0006	tolerated (0.11)									
				INTS1 c.2395T>C	p.V770A		novel	tolerated (0.32)									
MD-43	F	contracures, AD family history	500-800	CAV3 c.136G>A	p.A46T	AD, LGMD (CM012082)	n.a.	deleterious (0)	childhood	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/I	Mutation						Age of Onset	Age at Examination	Muscle wasting	Contractures	Heart involvement	Inheritance
				Gene	cDNA	Protein	known disease causing	MAF	SIFT						
MD-2	M	EDMD phenotype	1500-	<i>DYSF</i>	c.1369G>A	p.E457K	ar, MMD (CM074148)	0.008	---	2	16	yes	yes	no	ar
			2000	<i>Trim32</i>	c.1802A>G	p.H601R		novel	---						
MD-3	F	distal LGMD	200-800	<i>PLEC</i>	c.10562C>T	p.T3521M		0.00003	deleterious (0)	38	58	yes	no	no	known
				<i>COL6A2</i>	c.1769C>T	p.T590M	AD, BTHLM (CM1310895)	0.0004	deleterious (0.02)						
				<i>ATP12A</i>	c.9+1G>A			0.001	---						
				<i>XIRP1</i>	c.3763C>T	p.P1255S		0.002	deleterious (0)						
				<i>USP34</i>	c.19G>T	p.D7Y		novel	deleterious (0.01)						
MD-7	F	EDMD phenotype	350	<i>FKTN</i>	c.559G>A	p.G187S		0.00002	deleterious (0.02)	2	10	yes	yes	not doc	ar
				<i>KCNJ12</i>	c.109C>A	p.H37N		novel	tolerated (0.11)						
				<i>ANK2</i>	c.11791G>A	p.E3931K		0.003	deleterious (0.03)						
MD-9	F	Affected father with pacemaker	350	<i>COL6A3</i>	c.9508G>A	p.G3170R		0.0002	---	28	48	yes	no	not doc	AD
				<i>COL6A3</i>	c.1024G>A	p.V342M		0.001	---						
				<i>Tmem214</i>	c.536G>A	p.R179H		0.0002	deleterious (0.04)						
				<i>SYNE3</i>	c.2024G>A	p.R675Q		0.0004	tolerated (1)						
MD-10	M	EDMD phenotype, Father pacemaker	1000	<i>COL6A2</i>	c.2102C>A	p.T701N		novel	deleterious (0.01)	5	46	yes	yes	no	AD
				<i>AGRN</i>	c.1123G>T	p.A375S		0.005	tolerated (0.44)						
				<i>PLEC</i>	c.5638G>A	p.A1880T		0.0005	tolerated						

(continued)

							(0.11)													
				<i>AKAP6</i>	c.2663C>A	p.T888N		0.00002	tolerated (0.15)											
				<i>ATP12A</i>	c.1897G>A	p.G633S		0.0002	deleterious (0)											
				<i>USP34</i>	c.4387A>G	p.S1463G		novel	tolerated (0.14)											
MD-14	M	EMD phenotype mild	2000	<i>DYSF</i>	c.3191_3196dup	p.A1064_E1065dup	LGMD (CI105954)?	0.039	---											
				<i>PLEC</i>	c.12601G>A	p.E4201K		0.001	deleterious (0.01)											
				<i>LEMD3</i>	c.263G>T	p.G88V		0.000000	tolerated (0.19)											
				<i>RYR3</i>	c.1508G>C	p.G503A		0.000004	tolerated (0.3)											
				<i>ANK2</i>	c.6228G>T	p.K2076N		0.001	deleterious (0.05)											
MD-15	M	EDMD phenotype mild	400	<i>WFS1</i>	c.1294C>G	p.L432V		0.004	deleterious (0.01)	4	6	yes	no	no	no	no	unknown			
MD-16	M	Hemiatrophia totalis, Parry-Romberg	245	<i>FKRP</i>	c.456C>G	p.S152R		0.0005	tolerated (0.51)	12	43	yes	no	no	no	no	unknown			
MD-24		Bethlem phenotype, recently confirmed <i>STIM1</i> mutation		<i>AGRN</i>	c.4966C>T	p.R1656W		0.001	deleterious (0)			yes	yes				unknown			
MD-26	F	SPS scapulo-peroneal syndrome	200	<i>PLEC</i>	c.2648G>A	p.R883H		0.001	tolerated (0.22)											
				<i>WFS1</i>	c.2611G>A	p.V871M		0.008	tolerated (0.16)											
				<i>AKAP6</i>	c.3335G>A	p.G1112E		0.001	deleterious (0.01)											

(continued)

				<i>FHOD1</i>	c.2714G>A	p.R905Q		0.001	tolerated (0.06)							
MD-27	F	distal myopathy	300	<i>COL6A3</i>	c.8009C>T	p.A2670V		0.0002	---	15	25	yes	no	no	AD	
				<i>SYNE1</i>	c.19729C>T	p.R6577W		0.00007	deleterious (0)							
				<i>TTN</i>	c.73705G>C	p.V24569L		0.0002	---							
				<i>RYR3</i>	c.7249A>G	p.I2417V		0.001	tolerated (0.06)							
MD-28	M	distal myopathy: Nemaline- Myopathy, regional ichthyosis	200	<i>CAPN3</i>	c.1678A>G	p.T560A		0.00002	deleterious (0)	6	48	yes	no	no	unknown	
				<i>PLEC</i>	c.1715C>T	p.A572V		0.00002	tolerated (0.12)							
				<i>Lmo7</i>	c.2797G>A	p.A933T		0.0001	deleterious (0.03)							
				<i>KCNJ12</i>	c.127C>T	p.R43C		0.00004	deleterious (0)							
				<i>KCNJ12</i>	c.758dupT	p.F255Lfs*		0.001	---							
						19										
MD-29	F	LGMD	up to 10.00 0	<i>SYNE1</i>	c.18137C>T	p.T6046M		0.00002	deleterious (0)	33	47	yes	no	no	unknown	
				<i>LEMD3</i>	c.689G>C	p.R230T		novel	tolerated (0.11)							
				<i>INTS1</i>	c.820C>T	p.R274C		0.001	tolerated (0.07)							
				<i>TTN</i>	c.39624A>C	p.K13208 N		n.a.	---							
MD-30	M	distal myopathy	300	<i>DYSF</i>	c.1877C>T	p.M626T		0.001	tolerated (1)	57	73	yes	no	no	AD	
				<i>DYSF</i>	c.3191_3196dup	p.A1064_E 1065dup	LGMD (CI105954)?	0.039	---							
				<i>PLEC</i>	c.424C>T	p.R142W		0.0003	deleterious (0.01)							
				<i>MAGI1</i>	c.977C>A	p.T326K		novel	deleterious (0)							

(continued)

MD-31	F	ad MFM	500	SYNE2	c.12659A>C	p.Q4220P		0.000004	tolerated (0.17)	32	46	yes	no	no	AD
				ATP12A	c.1663A>G	p.T555A		0.0002	tolerated (0.47)						
				XIRP1	c.1055G>A	p.R352Q		0.0002	tolerated (0.62)						
				INTS1	c.4282C>T	p.P1399L		0.000007	tolerated (0.18)						
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	tolerated (0.22)						
				PLEC	c.4697C>T	p.S1566L		0.00005	deleterious (0.04)						
				TMEM38A	c.778G>A	p.D260N		0.002	tolerated (0.62)						
				PLPP7	c.755G>C	p.R252P		novel	tolerated (0.1)						
				INTS1	c.5707C>T	p.P1874L		0.000006	deleterious (0)						
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	tolerated (0.23)						
				WDR43	c.1075A>G	p.I359V		0.001	tolerated (0.18)						
				ANK2	c.11791G>A	p.E3931K		0.003	deleterious (0.03)						
MD-35		contractures and mild cardiomyopathy, non-consanguineous,	4000	DYSF	c.509C>A	p.A170E	ar, Dysferlinopathy (CM053208)	0.004	tolerated (0.97)	20	27	no	t	do	c
				COL6A3	c.4156G>A	p.E1386K	AD, BTHLM (CM050230)	0.002	---						
				AGRN	c.1528G>A	p.G510S		0.008	---						

(continued)

		sporadic																
MD-36		contractures and cardiac conduction defect, non-consanguineous, sporadic	500-1000	<i>SYNE2</i>	c.17191C>T	p.R5731C			0.00004	tolerated (0.15)	20s	66	no	yes	yes		sporadic	
MD-38		contractures and cardiac conduction defect, non-consanguineous, sporadic	normal	<i>LPCAT3</i>	c.805C>T	p.R269C			0.000004	deleterious (0)	congenital	56	no	yes	yes		sporadic	
				<i>XIRP1</i>	c.3442G>A	p.V1148M			0.001	deleterious (0.01)			no	yes	yes		sporadic	
MD-39		contractures, non-consanguineous, sporadic	1300	<i>DMD</i>	c.1252A>T	p.T418S			0.002	---	early childhood	21	no	yes	no		unknown	
				<i>DYSF</i>	c.3967C>G	p.Q1323E			0.001	tolerated (0.31)			no	yes	no		unknown	
				<i>FKTN</i>	c.373G>A	p.G125S	Walker-Warburg		0.037	---							unknown	
							syndrome?											
				<i>COL6A3</i>	c.4510C>T	p.R1504W			0.001	---								
				<i>PDE4DIP</i>	c.5486C>T	p.S1829F			0.0001	deleterious (0.03)								
				<i>PDE4DIP</i>	c.4063C>T	p.R1355*			0.0001	---								
				<i>ANK2</i>	c.4744C>T	p.R1582W			0.002	deleterious (0.01)								
MD-40		contractures, non-consanguineous, sporadic	normal	<i>Col6A3</i>	c.3419C>T	p.T1140M			0.0004	---								
				<i>SUN1</i>	c.278A>C	p.Q93P			0.004	---								
				<i>TTN</i>	c.93768_93769dup	p.K31257L fs*6			novel	---								
				<i>TTN</i>	c.107840T>A	p.I35947N			0.000004	---	5	46	yes	yes	no		sporadic	
				<i>ATP12A</i>	c.349A>G	p.I117V			0.00001	tolerated (0.1)								
				<i>ANK2</i>	c.11231C>A	p.T3744N			0.001	tolerated (0.1)								

(continued)

			<i>INTS1</i>	c.4969C>T	p.R1657C		0.00007	deleterious (0)								
MD-42	contractures and cardiac arrhythmia, non-consanguineous, sporadic	3000	<i>DYSF</i>	c.3065G>A	p.R1022Q	ar, LGMD (CM090628)	0.014	---	adult	42	no	t	ye	s	yes	sporadic
			<i>COL6A2</i>	c.1552C>T	p.P518S		0.003	tolerated (0.2)								
			<i>COL6A3</i>	c.4727G>A	p.R1576Q		0.004	---								
			<i>WDR43</i>	c.366T>G	p.S122R		0.0003	tolerated (0.21)								
MD-45	contractures and possible mild cardiac involvement, non-consanguineous, sporadic	1000	<i>DYSF</i>	c.3992G>T	p.R1331L	Dysferlinopathy (CM103814)	0.016	tolerated (0.51)	20s	46	no	t	ye	s	yes	sporadic
			<i>TRIM32</i>	c.558G>C	p.Q186H		0.001	deleterious (0.04)								
			<i>SUN1</i>	c.335C>T	p.T112M		0.0002	tolerated (0.16)								
			<i>BVES</i>	c.275A>G	p.D92G		novel	deleterious (0.01)								
MD-46			<i>SGCG</i>	c.792C>G	p.I264M		0.00006	deleterious (0.02)	3 mths	5	ye	ye	no	un		
46								us (0.02)			s	s				known
			<i>COL6A1</i>	c.347G>A	p.S116N	AD, BTHLM (CM050211)?	0.031	---								
			<i>ITGA7</i>	c.824G>A	p.R275H		0.033	---								
			<i>TMEM201</i>	c.52G>A	p.G18S		novel	tolerated (0.86)								
MD-47			<i>COL6A2</i>	c.316G>A	p.E106K	AD, BTHLM (CM050217)?	0.002	tolerated (1)	1	28	ye	s	ye	s	yes	unknown
			<i>RYR2</i>	c.1939C>T	p.R647C		0.0001	deleterious (0.01)								
MD-48			<i>POMT1</i>	c.1233C>A	p.D411E		0.035	---	5	21	ye	s	ye	s	no	unknown
			<i>RYR2</i>	c.2444C>T	p.P815L		0.00003	deleterious (0)								
			<i>USP34</i>	c.3938G>T	p.G1313V		novel	deleterious (0)								
			<i>ANK2</i>	c.9854T>C	p.I3285T		0.009	deleterious (0.05)								
MD-			<i>SGCA</i>	c.371T>C	p.I124T	LGMD	0.00006	deleterious (0)	2	9	ye	ye	no	AD		

(continued)

49						us (0)					s	s		
			<i>SGCG</i>	c.320C>T	p.S107L		0.000008							
			<i>COL6A1</i>	c.347G>A	p.S116N	AD, BTHLM (CM050211)?	0.031							
			<i>RYR3</i>	c.6698G>A	p.R2233Q		0.001							
			<i>RYR3</i>	c.9254C>G	p.P3085R		0.0004							
			<i>RYR2</i>	c.3407C>T	p.A1136V		0.007							
MD-50			<i>DYSF</i>	c.509C>A	p.A170E	Dysferlinopathy (CM053208)	0.004							
			<i>COL6A1</i>	c.347G>A	p.S116N	AD, BTHLM (CM050211)?	0.031							
			<i>COL6A2</i>	c.2558G>A	p.R853Q	AD, BTHLM (CM050225)?	0.004							
			<i>AGRN</i>	c.5201G>A	p.R1734H		0.004							
			<i>TMEM38A</i>	c.778G>A	p.D260N		0.002							
			<i>XIRP1</i>	c.802_805dup	p.A269Vfs*6		novel							
MD-51			<i>POMT1</i>	c.1233C>A	p.D411E		0.035							
			<i>COL6A1</i>	c.347G>A	p.S116N	AD, BTHLM (CM050211)?	0.031							
MD-52			<i>DMD</i>	c.3595G>A	p.E1199K		novel							
			<i>SYNE2</i>	c.1823A>G	p.E608G		0.000004							
			<i>TMEM43</i>	c.351dup	p.H118Afs*11		0.00003							
MD-53			<i>SYNE1</i>	c.17347G>C	p.E5783Q		0.000004							
MD-54			<i>FKTN</i>	c.116C>T	p.R56C		0.01							
			<i>AGRN</i>	c.1528G>A	p.G510S		0.008							

(continued)

				<i>PLEC</i>	c.8497_851 6dup	p.L2839*		novel	---										wn
				<i>PLEC</i>	c.11708G>C	p.R3903P		0.000004	deleterious (0.01)										
				<i>PLEC</i>	c.8917G>A	p.D2973N		0.0004	deleterious (0.01)										
				<i>SYNE1</i>	c.25403G>A	p.R8468H		0.0002	tolerated (0.61)										
				<i>TMEM214</i>	c.787G>A	p.G263S		0.001	deleterious (0.02)										
MD-55				<i>AGRN</i>	c.5353G>A	p.D1785N		0.006	tolerated (0.51)	8	16	yes	yes	no					unknown
MD-56	M	FSHD1	400 U/L	<i>TMEM38A</i>	c.778_780del	p.D260del		0.005	---	15	74	yes	no	no					unknown
				<i>KLHL31</i>	c.1817G>A	p.G606D		novel	deleterious (0)										
				<i>RYR3</i>	c.5356G>A	p.G1786S		0.001	tolerated (0.67)										

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).

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% - considerably 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 [16,29]. 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).

3.4. Confirmation of novel EDMD alleles

Thus far only the three *LMNA* mutations, the *CAV3* and one of the *CAPN3* (MD-43) mutations have been fully 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 largely non-overlapping sets of genes to the nuclear periphery to enhance their repression and many of the genes targeted are antagonistic to myogenesis or from alternate (non-muscle) differentiation pathways [18]. 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

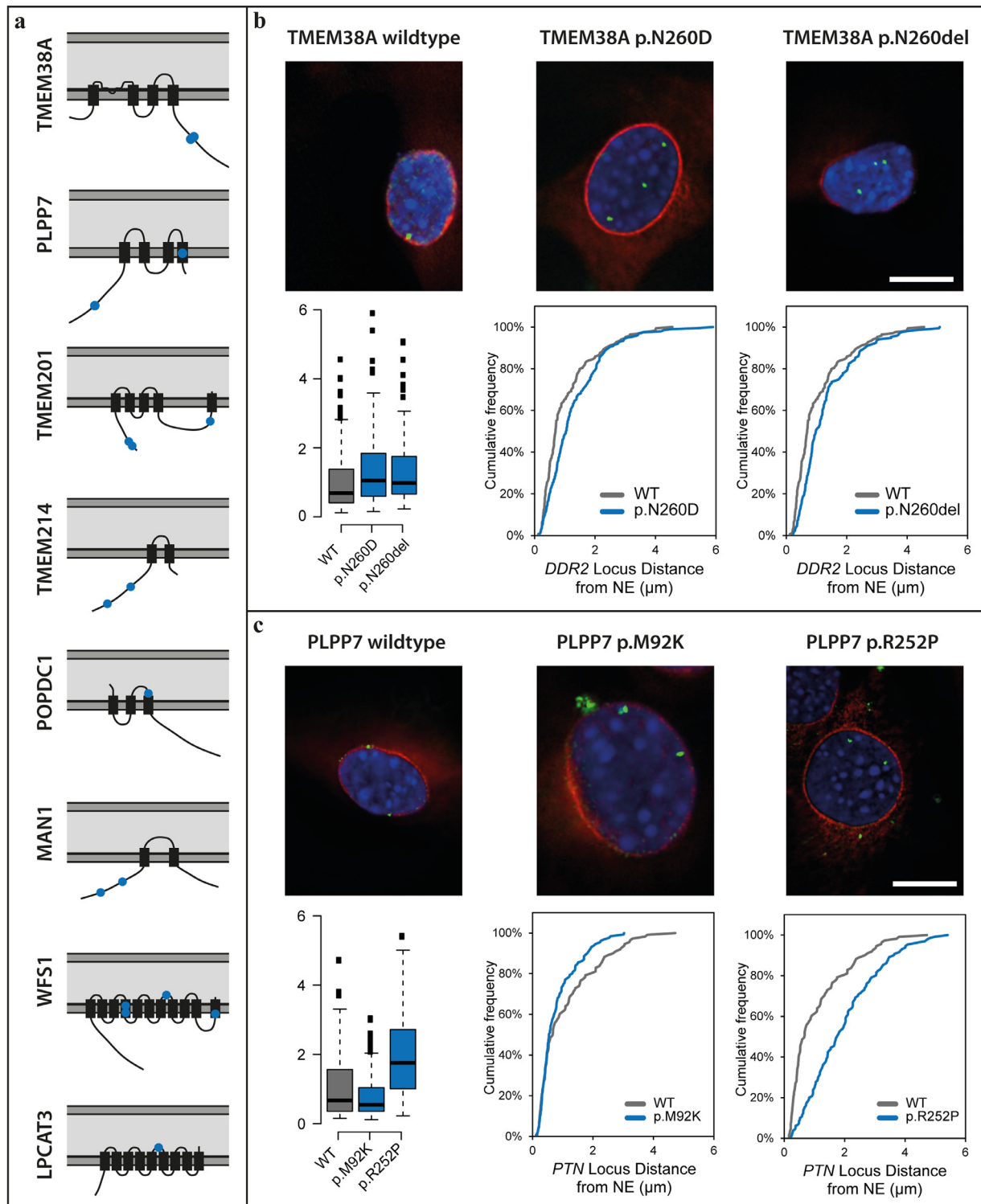


Fig. 3. Mutations in muscle gene-repositioning NETs affect their ability to recruit genes to the nuclear envelope (NE). (a) Schematic presentation of the topology of further muscle NETs and their mutations identified by the primer library sequencing. The lipid bilayers of the nuclear envelope are shown 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 localisation 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 localisation 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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). Testing effects of both PLPP7 and TMEM38A mutations on nuclear morphology [30] did show that only TMEM38A p.N260D had a slight increase in numbers of abnormally shaped nuclei (Supplementary Fig. 1).

4. Discussion

Failure of high throughput genomic approaches to identify new disease alleles can at least in some cases be overcome by our multi-stage approach. This approach pinpointed candidates in part based on the preferential tissue focus of pathology and in part on the sub-cellular localisation 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 [16]. 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 [31].

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 [32]. 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 mislocalised in isolated differentiating EDMD muscle cells or muscle biopsy sections [33].

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 [34–37]. 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 signalling 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 [16,41,42], 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 [18]. 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. Interestingly, Tmem214 regulated genes exhibited considerable overlap with WFS1, Tmem38A, and NET39 regulated genes, suggesting it functions in mechanosignal transduction, while the others each had principally unique gene targets. The links of candidate alleles to gene misregulation are the more compelling in this context because the different sets of genes regulated by WFS1, Tmem38A, and NET39 – all important in myogenesis – thus support 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 for the gene regulation pathomechanism as this chain could enable 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 may be disease modifiers and so might eventually be used as predictors of severity or other aspects of presentation such as cardiac involvement 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 analysis 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. A further benefit of this primer library is the possibility to add additional genes at a later time point. This might become necessary when a better understanding of EDMD identifies more causative or modifying genes involved in the disease pathology.

Funding sources

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 no. 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. The funders played no role in study design, data collection, data analysis, interpretation, and writing of the report.

Declaration of Competing Interest

Dr. Meinke reports personal fees from Greenovation Biopharmaceuticals, outside the submitted work; Dr. Harris reports grants from Muscular Dystrophy-UK, during the conduct of the study; Dr. Koelbel reports personal fees from Santhera, personal fees from Sarepta, personal fees from Biogen, personal fees from Avexis, outside the submitted work; Dr. Schara reports personal fees for consulting from Alexion, Santhera, PTC, Ipsen, Biogen, Sarepta, Dynacure, Novartis, Avexis and Sanofi; Dr. Muntoni reports personal fees from Pfizer, personal fees from Roche, personal fees from Biogen, personal fees from Avexis, personal fees from Sarepta, grants from Sarepta, outside the submitted work; Dr. Schoser reports Advisory boards: Amicus Therapeutics, Audentes Therapeutics, Nexien Biopharm, Lupin therapeutics, Sanofi Genzyme. Contracted research: Amicus therapeutics, Greenovation Biopharm, Sanofi Genzyme. Honoraria: Kedrion. Travel expenses: Sanofi Genzyme. Dr. Kerr, Dr. Czapiewski, Dr. de las Heras, Mr. Dixon, Dr. Straub, Dr. Wehnert and Dr. Schirmer have nothing to disclose.

Acknowledgements

We dedicate this study in honour of Alan E.H. Emery who not only described X-linked EDMD as a distinct disease in 1966 but also helped instigate the longest standing regular Laminopathy meeting in 2006. We further want to thank all patients and their families for their valuable contribution.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2019.11.048.

References

- [1] Hughes MI, Hicks EM, Nevin NC, Patterson VH. The prevalence of inherited neuromuscular disease in Northern Ireland. *Neuromuscul Disord* 1996;6(1):69–73.
- [2] Norwood FL, Harling C, Chinnery PF, Eagle M, Bushby K, Straub V. Prevalence of genetic muscle disease in Northern England: in-depth analysis of a muscle clinic population. *Brain* 2009;132(11):3175–86.
- [3] Emery AE, Dreifuss FE. Unusual type of benign x-linked muscular dystrophy. *J Neurol Neurosurg Psychiatry* 1966;29(4):338–42.
- [4] Emery AE. Emery-Dreifuss syndrome. *J Med Genet* 1989;26(10):637–41 [Review].
- [5] Knoblauch H, Geier C, Adams S, Budde B, Rudolph A, Zacharias U, et al. Contractures and hypertrophic cardiomyopathy in a novel FHL1 mutation. *Ann Neurol* 2010;67(1):136–40.
- [6] Bonne G, Mercuri E, Muchir A, Urtizberea A, Becane HM, Recan D, 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* 2000;48(2):170–80.
- [7] Bione S, Maestrini E, Rivella S, Mancini M, Regis S, Romeo G, et al. Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat Genet* 1994;8(4):323–7.
- [8] Bonne G, Di Barletta MR, Varnous S, Becane HM, Hammouda EH, Merlini L, et al. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nat Genet* 1999;21(3):285–8.
- [9] Zhang Q, Bethmann C, Worth NF, Davies JD, Wasner C, Feuer A, 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* 2016;25(23):2816–33.
- [10] Meinke P, Mattioli E, Haque F, Antoku S, Columbaro M, Straatman KR, et al. Muscular dystrophy-associated SUN1 and SUN2 variants disrupt nuclear-cytoskeletal connections and myonuclear organization. *PLoS Genet* 2014;10(9):e1004605.
- [11] Gueneau L, Bertrand AT, Jais JP, Salih MA, Stojkovic T, Wehnert M, et al. Mutations of the FHL1 gene cause Emery-Dreifuss muscular dystrophy. *Am J Hum Genet* 2009;85(3):338–53.
- [12] Muntoni F, Bonne G, Goldfarb LG, Mercuri E, Piercy RJ, Burke M, et al. Disease severity in dominant Emery Dreifuss is increased by mutations in both emerin and desmin proteins. *Brain* 2006;129(5):1260–8.
- [13] Meinke P. Molecular genetics of Emery–Dreifuss muscular dystrophy. Chichester: John Wiley & Sons, Ltd; 2018.
- [14] Meinke P, Schirmer EC. The increasing relevance of nuclear envelope myopathies. *Curr Opin Neurol* 2016;29(5):651–61.
- [15] Wilkie GS, Schirmer EC. Guilt by association: the nuclear envelope proteome and disease. *Mol Cell Proteomics* 2006;5(10):1865–75.
- [16] Wilkie GS, Korfali N, Swanson SK, Malik P, Srsen V, Batrakou DG, et al. Several novel nuclear envelope transmembrane proteins identified in skeletal muscle have cytoskeletal associations. *Mol Cell Proteomics* 2011;10(11):M110 003129.
- [17] Osorio DS, Gomes ER. Connecting the nucleus to the cytoskeleton for nuclear positioning and cell migration. *Adv Exp Med Biol* 2014;773:505–20.
- [18] Robson MI, de las Heras JI, Czapiewski R, Le Thanh P, Booth DG, Kelly DA, et al. Tissue-Specific gene repositioning by muscle nuclear membrane proteins enhances repression of critical developmental genes during myogenesis. *Mol Cell* 2016;62(6):834–47.
- [19] Zuleger N, Boyle S, Kelly DA, de las Heras JI, Lazou V, Korfali N, et al. Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. *Genome Biol* 2013 15;14(2):R14.
- [20] Zuleger N, Boyle S, Kelly DA, de las Heras JI, Lazou V, Korfali N, et al. Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. *Genome Biol* 2013;14(2):2013–4.
- [21] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20(9):1297–303.
- [22] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43(5):491–8.
- [23] Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;43 11 0 1–33.
- [24] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. 2019:531210.
- [25] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29(1):15–21.
- [26] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25(16):2078–9.
- [27] Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;42:W187–91.
- [28] Flanigan KM, von Niederhausern A, Dunn DM, Alder J, Mendell JR, Weiss RB. Rapid direct sequence analysis of the dystrophin gene. *Am J Hum Genet* 2003;72(4):931–9.
- [29] Korfali N, Wilkie GS, Swanson SK, Srsen V, de las Heras J, Batrakou DG, et al. The nuclear envelope proteome differs notably between tissues. *Nucleus* 2012;3(6):552–64.
- [30] Filippi-Chiela EC, Oliveira MM, Jurkovski B, Callegari-Jacques SM, da Silva VD, Lenz G. Nuclear morphometric analysis (NMA): screening of senescence, apoptosis and nuclear irregularities. *PLoS ONE* 2012;7(8):e42522.
- [31] 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* 2015;43(6):3114–27.
- [32] Liu J, Lee KK, Segura-Totten M, Neufeld E, Wilson KL, Gruenbaum Y. MAN1 and emerin have overlapping function(s) essential for chromosome segregation and cell division in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 2003;100(8):4598–603.
- [33] Mattioli E, Columbaro M, Jafferli MH, Schena E, Hallberg E, Lattanzi G. Samp1 mislocalization in Emery-Dreifuss muscular dystrophy. *Cells* 2018;7(10).
- [34] Broers JL, Kuijpers HJ, Ostlund C, Worman HJ, Endert J, Ramaekers FC. Both lamin A and lamin C mutations cause lamina instability as well as loss of internal nuclear lamin organization. *Exp Cell Res* 2005;304(2):582–92.
- [35] Lammerding J, Hsiao J, Schulze PC, Kozlov S, Stewart CL, Lee RT. Abnormal nuclear shape and impaired mechanotransduction in emerin-deficient cells. *J Cell Biol* 2005;170(5):781–91.
- [36] Rowat AC, Lammerding J, Ipsen JH. Mechanical properties of the cell nucleus and the effect of emerin deficiency. *Biophys J* 2006;91(12):4649–64.
- [37] Hakelien AM, Delbarre E, Gaustad KG, 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* 2008;314(8):1869–80.
- [38] Holaska JM, Wilson KL. 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* 2007;46(30):8897–908.
- [39] Ziat E, Mamchaoui K, Beuvin M, Nelson I, Azibani F, Spuler S, 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* 2016;3(4):497–510.
- [40] Windpassinger C, Schoser B, Straub V, Hochmeister S, Noor A, Lohberger B, et al. An X-linked myopathy with postural muscle atrophy and generalized hypertrophy, termed XMPMA, is caused by mutations in FHL1. *Am J Hum Genet* 2008;82(1):88–99.
- [41] Buch C, Lindberg R, Figueroa R, Gudise S, Onischenko E, Hallberg E. An integral protein of the inner nuclear membrane localizes to the mitotic spindle in mammalian cells. *J Cell Sci* 2009;122(12):2100–7.
- [42] Tao S, Yamazaki D, Komazaki S, Zhao C, Iida T, Kakizawa S, et al. Facilitated hyperpolarization signaling in vascular smooth muscle-overexpressing TRIC-A channels. *J Biol Chem* 2013;288(22):15581–9.