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

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

\textsuperscript{a} Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, UK
\textsuperscript{b} Friedrich Baur Institute at the Department of Neurology, University Hospital, LMU Munich, Germany
\textsuperscript{c} John Walton Muscular Dystrophy Research Centre, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
\textsuperscript{d} Department of Pediatric Neurology, Developmental Neurology and Social Pediatrics, University of Essen, Germany
\textsuperscript{e} Dubowitz Neuromuscular Centre, University College London Great Ormond Street Institute of Child Health, London, UK
\textsuperscript{f} 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
\textsuperscript{g} Institute of Human Genetics, University of Greifswald (retired), Greifswald, Germany

\textbf{ARTICLE INFO}

\textbf{Article History:}
Received 14 October 2019
Revised 18 November 2019
Accepted 26 November 2019
Available online xxx

\textbf{Keywords:}
Emery-Dreifuss muscular dystrophy
Nuclear envelope
Nuclear envelope transmembrane protein primer library
Orphan disease

\textbf{ABSTRACT}

\textbf{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.

\textbf{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.

\textbf{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.

\textbf{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.

\textbf{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...
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 variability 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.

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, nespin 1, nespin 2, SUN1, and Tmem241 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 KHL31 track with microtubules on the nuclear surface [16] while NETS/Samp1 contributes actin and centrosome interactions [17]. NETs Tmem38A, WFS1, NET39/PLPP7 and, again, Tmem214 and NETS/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 Dubowirz Neumuscular 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+S and IRAS ID 118,436 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 CNMDBL63 CT-2925/CT-1402, and for this particular study was obtained from the West of Scotland Research Ethics Service (WoSRES) with REC reference 15/WS/0069 and IRAS project ID 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, Norcoss, 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 SureSelectXT 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 SureSelectQXT 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.org).
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 envelope 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\textsuperscript{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 candidate genes after primer library sequencing (lower panel); (d) Functional involvement of candidate genes in the extracellular matrix (ECM), plasmamembrane (PM), cytoskeleton, nuclear envelope (NE), nucleus, endoplasmatic reticulum, and cytoplasm. 

Please cite this article as: P. Meinke et al., A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism, EBioMedicine (2019), https://doi.org/10.1016/j.ebiom.2019.11.048
Roman numerals indicate categories due to linkage analysis; (I) 252 exome sequencing candidates, and (II) 25 genes from similar muscular dystrophies, (III) the 252 exome sequencing candidates, and (IV) 16 functional candidates, mostly muscle-specific, non-nuclear envelope proteins (Fig. 2a; Supplemental Table S4). Sequencing was performed on 56 additional clinically diagnosed EDMD patients unrelated to each other, obtaining an average 3427.992 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.R414H, p.R249Q, p.G535fs*5; 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 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, ANOS1, 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 ANOS1, DYSF and POMT1 the respective patients had two mutations, consistent with the reported inheritance (autosomal recessive for MD-20/ANOS1 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 primary library cohort based on the assumption that causative genes will be independently mutated in multiple patients. The top candidates were INTS1, ANK2, XRIP1 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 USP344 (heterozygous mutations in exome sequenced family 2) and XRIP1 (mutated in families 2 and 4), each with mutations in an additional five patients. Some patients had additional mutations in already associated genes, but all 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...
Table 1
Solved Patients.

| ID | Sex | Comments/Clinical issues | CK in | Mutation | Age of Onset | Age at Examination | Muscle wasting | Contractures | Heart involvement | Inheritance |
|----|-----|--------------------------|-------|----------|--------------|-------------------|----------------|--------------|-----------------|-------------|
|    |     |                          |       |          |              |                   |                |              |                 |             |
| MD-1 M |   | EDMD phenotype with contracts, 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) 0.00003 tolerated (0.96) |              |                   |                |                |                 |             |
| MD-4 F |   | LGMD, contracts | 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) 0.00003 deleterious (0) |              |                   |                |                |                 |             |
| MD-5 F |   | LGMD, contracts | 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) 0.00003 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 M |   | contracures, non-consanguineous, sporadic | normal | CAPN3 c.145C>T p.R49C ar, LGMD (CM076055) 0.00002 | --- | 50 | not doc | yes | no | sporadic |
|     |     |                          |       | CAPN3 c.1821-1825del p.R608Kfs*23 novel | --- |              |                |              |                 |             |
| MD-11 M |   | EDMD phenotype with contracts, | 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)
| ID  | Sex | Disorder                        | Gene     | Change     | Description          | Type     | Age  | Gender | Breed | Known | Behavior |
|-----|-----|---------------------------------|----------|------------|----------------------|----------|------|--------|-------|-------|----------|
| MD-19 | F   | EDMD phenotype, pacer           | LMNA     | c.1606delG  | p.E536Kfs*12         | novel    | 200  | yes    | no    | yes   | unsporadic |
| MD-37 | M   | contractures and cardiac conduction defect, non-consanguineous sporadic | LMNA     | c.746G>A    | p.R249Q       | n.a.     | 370  | yes    | yes   | yes   | sporadic |
| MD-13 | M   | mild LGMD, myalgia              | DMD      | c.9G>A      | p.W3*         | n.a.     | 2200 | yes    | no    | no    | unknown   |
| MD-17 | M   | moderate muscle wasting, hIBMFTD-Paget | VCP     | c.476G>A    | p.R159H       | tolerated | 700  | yes    | no    | no    | AD        |
|      |     |                                 | USP34    | c.2963T>C   | p.L988P       | novel    |      | yes    | no    | no    | AD        |
| MD-12 | M   | myalgia, proximal weakness, arrhythmia | COL6A2  | c.2795C>T   | p.P932L       | tolerated | 700  | yes    | no    | yes   | AD        |
|      |     |                                 | VCP      | c.17A>T     | p.D6V         | novel    |      | yes    | no    | no    | AD        |
|      |     |                                 | SYNE2    | c.2669C>A   | p.T890K       | tolerated |      | yes    | no    | yes   | AD        |
|      |     |                                 | SYNE2    | c.2647-2A>T |             | 0.00003  |      | yes    | no    | yes   | AD        |
|      |     |                                 | XIRP1    | c.4648A>T   | p.I1550F      | novel    |      | yes    | no    | yes   | AD        |
|      |     |                                 | XIRP1    | c.3612G>T   | p.W1204C      | novel    |      | yes    | no    | yes   | AD        |
| MD-6 | F   | EDMD                             | GBE1     | c.691+2T>   | splice donor  | AD, GSD4 (CS100318) | 600  | yes    | yes   | no    | AD        |

(continued)
| Case | Age | Phenotype | Genes | Allele | Protein | Status | PD | D | A | Know |
|------|-----|-----------|-------|--------|---------|--------|----|---|---|------|
| MD-22 | F | Distal myopathy | GBE1 | c.1382T>C | p.V461A | Novel | Tolerated (0.25) |
| MD-22 | F | Distal myopathy | TTN | c.107635C>T | p.Q35879* |
| MD-22 | F | Distal myopathy | TTN | c.22207C>T | p.Q7343* |
| MD-25 | F | Distal myopathy | USP34 | c.7411C>T | p.H2471Y |
| MD-34 | F | Contractures and mild cardiomyopathy | GBE1 | c.839G>A | p.G280D |
| MD-34 | F | Contractures and mild cardiomyopathy | DYSF | c.5698-5699del | p.S1900Q5=14 |
| MD-34 | F | Contractures and mild cardiomyopathy | TMEM34 | c.934C>T | p.R312W |
| MD-18 | EDMD phenotype | TTN | c.40787-2A>G | Novel |
| MD-18 | EDMD phenotype | TTN | c.9047del | p.M3016* |
| MD-18 | EDMD phenotype | TTN | c.72409T>C | p.S24137P |
| MD-18 | EDMD phenotype | SYNE1 | c.16843G>A | p.E5615K |

Please cite this article as: P. Meinke et al., A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism, EBioMedicine (2019), https://doi.org/10.1016/j.ebiom.2019.11.048
| Patient | Condition | Mutation | Gene | Codon Change | Protein Change | p-value | Evidence |
|---------|-----------|----------|------|--------------|---------------|---------|----------|
| MD-8    | EDMD-like rigid-spine syndrome, Bethlem | TTN | c.10495ZA>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.6597S | 0.00001 | deleterious (0.02) |
| MD-20   | M | COL6A1 | c.2166dup | p.1723Hfs*7 | novel | --- |
|         | M | COL6A1 | c.1784delA | p.E595Gfs*7 | 0.00000 | --- |
|         | M | COL6A1 | c.1786A>G | p.I596V | 0.00000 | tolerated (0.57) |
|         | M | WFS1 | c.1675G>A | p.A559T | 0.001 | tolerated |
| MD-21   | M | ANOS | c.191dup | p.N64Kfs*15 | ar, LGMD (C1101059) | 0.001 | --- |
|         | M | ANOS | c.1391C>A | p.A464D | ar, LGMD (CM137896) | n.a. | deleterious (0.02) |
| MD-23   | M | DYFS | c.3065G>A | p.R1022Q | ar, LGMD (CM09628) | 0.014 | --- |
|         | M | DYFS | c.3992G>T | p.R1331L | ar, Dysferlinopathy (CM103814) | 0.016 | tolerated (0.51) |
|         | M | POMT1 | c.1545C>G | p.S515R | 0.0006 | tolerated (0.27) |
|         | M | POMT1 | c.1838G>A | p.R613H | 0.0006 | tolerated (0.11) |
|         | M | INTS1 | c.2395T>C | p.V770A | novel | --- |
| MD-43   | F | CAV3 | c.136G>A | p.A46T | AD, LGMD (CM012082) | n.a. | deleterious (0) |

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).

Please cite this article as: P. Meinke et al., A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism, EBioMedicine (2019), https://doi.org/10.1016/j.ebiom.2019.11.048
### Table 2: Patients with Candidate Genes.

| ID   | Sex | Comments/Clinical issues | CK in (U/I) | Gene | cDNA | Protein | known disease causing | MAF | SIFT | Age of Onset | Age at Examination | Muscle wasting | Controversies | Heart involvement | Inheritance |
|------|-----|----------------------------|-------------|------|------|---------|----------------------|-----|------|-------------|-------------------|----------------|---------------|------------------|-------------|
| MD-2 | M   | EDMD phenotype            | 1500/2000   | DYSF | c.1369G>A | p.E457K | ar, MMD (CM07414B) | 0.008 | ---  | 2           | 16                | yes           | yes           | no               | ar          |
|      |     |                            |             | Trim32| c.1802A>G | p.H601R |                      |      |      |             |                   |               |               |                  |             |
| MD-3 | F   | distal LGMD               | 200-800     | PLEC | c.10562C>T | p.T3521M |                      | 0.0003 | deleterious (0) |             |                   | yes           | no            | no               | unknown     |
|      |     |                            |             | COL6A2| c.1769C>T | p.T590M | AD, BTHLM (CM1310895) | 0.0004 | deleterious (0.02) |             |                   | yes           | no            | no               | unknown     |
|      |     |                            |             | ATP12A| c.9+1G>A  |          |                      | 0.001 | ---  |             |                   | yes           | no            | no               |             |
|      |     |                            |             | XRIP1 | c.3763C>T | p.P1255S |                      | 0.002 | deleterious (0) |             |                   | yes           | no            | no               |             |
|      |     |                            |             | USP34 | c.19G>T   | p.D7Y   |                      |      |      |             |                   |               |               | deleterious (0.01) |             |
| MD-7 | F   | EDMD phenotype            | 350         | FKTN  | c.559G>A  | p.G187S |                      | 0.0002 | deleterious (0.02) |             |                   | yes           | yes          | no               | ar          |
|      |     |                            |             | KCNJ12| c.109G>A  | p.H37N  |                      |      |      |             |                   | yes           | no            | tolerated (0.11) |             |
|      |     |                            |             | ANK2  | c.11791G>A| p.E3931K |                      | 0.003 | deleterious (0.03) |             |                   | yes           | no            | no               |             |
| MD-9 | F   | Affected father with pacemaker | 350         | COL6A3| c.3508G>A | p.G3170R |                      | 0.0002 | ---  |             |                   | yes           | no            | not documented | AD          |
|      |     |                            |             | COL6A3| c.1024G>A | p.V342M |                      | 0.001 | ---  |             |                   | yes           | no            | no               |             |
|      |     |                            |             | Tmem214| c.536G>A  | p.R179H  |                      | 0.0002 | deleterious (0.04) |             |                   | yes           | no            | no               |             |
|      |     |                            |             | SYNE3 | c.2024G>A | p.R675Q |                      | 0.0004 | tolerated (1) |             |                   | yes           | no            | no               |             |
| MD-10| M   | EDMD phenotype, Father with pacemaker | 1000        | COL6A2| c.2102C>A  | p.T701N |                      |      |      |             |                   | yes           | yes          | no               | AD          |
|      |     |                            |             | AGRN  | c.1123G>T | p.A375S |                      | 0.005 | tolerated (0.44) |             |                   | yes           | yes          | no               |             |
|      |     |                            |             | PLEC  | c.5638G>A  | p.A1880T |                      | 0.0005 | tolerated |             |                   | yes           | yes          | no               |             |

(continued)
| MD-14 | M | EMD phenotype mild | 2000 | AKAP6 | c.2663C>A | p.T888N | 0.00002 tolerated (0.15) |
| MD-15 | M | EDMD phenotype mild | 400 | DYSF | c.3191_3196dup | p.A1064_E1065dup | LGMD (C105954)? 0.039 --- |
| MD-16 | M | Hemiatrophia totalis, Parry-Romberg | 245 | PLEC | c.12601G>A | p.E4201K | 0.001 deleterious (0.01) |
| MD-24 | M | Bethlem phenotype, recently confirmed STIM1 mutation | | LEMD3 | c.263G>T | p.688V | 0.000000 tolerated (0.19) |
| MD-26 | F | Scapuloperoneal syndrome | 200 | RYR3 | c.1508G>C | p.6503A | 0.000004 tolerated (0.3) |
| | | | | ANK2 | c.6228G>T | p.K2076N | 0.001 deleterious (0.05) |
| | | | | WFS1 | c.1294C>G | p.I432V | 0.004 deleterious (0.01) |
| | | | | FKR | c.456C>G | p.S152R | 0.0005 tolerated (0.51) |
| | | | | AGRN | c.4966C>T | p.R1656W | 0.001 deleterious (0) |
| | | | | PLEC | c.2648G>A | p.R883H | 0.001 tolerated (0.22) |
| | | | | WFS1 | c.2611G>A | p.V871M | 0.008 tolerated (0.16) |
| | | | | AKAP6 | c.3335G>A | p.G1112E | 0.001 deleterious (0.01) |

Please cite this article as: P. Meinke et al., A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism, EBioMedicine (2019), https://doi.org/10.1016/j.ebiom.2019.11.048
| ID  | Gender | Disease | Age | Family History | Mutation | Germline | pDB Score | Association | Gene | Description |
|-----|--------|---------|-----|----------------|-----------|----------|-----------|------------|------|-------------|
| MD-27 | F | distal myopathy | 300 | no | COL6A3 c.8009C>T p.A2670V | 0.0002 | no | 15 | 25 | yes | no | AD |
|     |      |         |     |     | SYNE1 c.19729C>T p.R6577W | 0.00007 | deleterio | (0) |             |       |             |      |
|     |      |         |     |     | TTN c.73705G>C p.V24569I | 0.0002 | no |     |         |       |             |      |
|     |      |         |     |     | RYR3 c.7249A>G p.I2417V | 0.001 | tolerated | (0.06) |             |       |             |      |
| MD-28 | M | distal myopathy: Nermaline Myopathy, regional ichthyosis | 200 | no | CAPN3 c.1578A>G p.T560A | 0.00002 | deleterio | (0) |             |       |             |      |
|     |      |         |     |     | PLEC c.1715C>T p.A572V | 0.00002 | tolerated | (0.12) |             |       |             |      |
|     |      |         |     |     | Lmo7 c.2797G>A p.A933T | 0.0001 | deleterio | (0.03) |             |       |             |      |
|     |      |         |     |     | KCNJ12 c.127C>T p.R43C | 0.00004 | deleterio | (0) |             |       |             |      |
|     |      |         |     |     | KCNJ12 c.758dupT p.F255Lfs* | 0.001 | --- |         |             |       |             |      |
| MD-29 | F | LGMD | up to 10.00 | no | SYNE1 c.18137C>T p.T6046M | 0.00002 | deleterio | (0) |             |       |             |      |
|     |      |         |     |     | LEMO3 c.689G>C p.R230T | novel | tolerated | (0.11) |             |       |             |      |
|     |      |         |     |     | INTS1 c.820C>T p.R274C | 0.001 | tolerated | (0.07) |             |       |             |      |
|     |      |         |     |     | TTN c.39624A>C p.K13208N | n.a. | --- |         |             |       |             |      |
| MD-30 | M | distal myopathy | 300 | no | DYSF c.1877C>T p.M626T | 0.001 | tolerated | (1) |             |       |             |      |
|     |      |         |     |     | DYSF c.3191_319 6dup p.A1064_E1065dup | LMGD (C105954)? | 0.039 | --- |         |       |             |      |
|     |      |         |     |     | PLEC c.424C>T p.R142W | 0.0003 | deleterio | (0.01) |             |       |             |      |
|     |      |         |     |     | MAGI1 c.977C>A p.T326K | novel | deleterio | (0) |             |       |             |      |
| MD-31 F ad MFM | SYNE2 | c.12659A>C | p.Q4220P | 0.000004 | tolerated (0.17) |
|---|---|---|---|---|---|
| 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) |
| DYSF | c.3191_319dup | p.A1064_E1065dup | LGMD (CI105954)? | 0.039 | --- |
| 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 | --- |
| 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 | DYSF | c.509C>A | p.A170E | 0.004 | tolerated (0.97) |
| COL6A3 | c.4156G>A | p.E1386K | AD, BTHLM (CM050230) | 0.002 | --- |
| AGRN | c.1528G>A | p.S510S | 0.008 | --- |

(continued)
| MD-36 | sporadic contractures and cardiac conduction defect, non-consanguineous, sporadic | SYNE2 | c.17191C>T | p.R5731C | 0.00004 | tolerated (0.15) | 20s | yes | no | 66 | yes | yes | 500-1000 |
|-------|-------------------------------------------------|-------|-------------|-------------|----------|----------------|-----|-----|----|-----|-----|-----|----------|
| MD-38 | contractures and cardiac conduction defect, non-consanguineous, sporadic | lPCAT3 | c.805C>T | p.R269C | 0.000004 | deleterious [0] | con genital | 56 | no | no | 30 | yes | yes | norm |
|       |                                                | XIRP2 | c.3442G>A | p.V1148M | 0.001 | deleterious [0.01] |      |     |    |     |     |     |          |
| MD-39 | contractures, non-consanguineous, sporadic | DMD | c.1252A>T | p.T418S | 0.002 | --- | early childhood | 21 | no | no | unknown |
|       |                                                | DYSF | c.3967C>G | p.Q1323E | 0.001 | tolerated (0.31) |      |     |    |     |     |     |          |
|       |                                                | FKTN | c.373G>A | p.G125S | 0.037 | --- |              |     |     |    |     |     |     |          |
|       |                                                | COL6A3 | c.4510C>T | p.R1504W | 0.001 | --- |              |     |     |    |     |     |     |          |
|       |                                                | PDE4DIP | c.5486C>T | p.S1829F | 0.0001 | deleterious [0.03] |      |     |    |     |     |     |          |
|       |                                                | PDE4DIP | c.4063C>T | p.R1355* | 0.0001 | --- |              |     |     |    |     |     |     |          |
|       |                                                | ANK2 | c.4744C>T | p.R1582W | 0.002 | deleterious [0.01] |      |     |    |     |     |     |          |
|       |                                                | Col6A3 | c.3419C>T | p.T1140M | 0.0004 | --- |              |     |     |    |     |     |     |          |
|       |                                                | SUN1 | c.278A>C | p.Q93P | 0.004 | --- |              |     |     |    |     |     |     |          |
|       |                                                | TTN | c.93768_93769del | p.K31257L | novel | --- |              |     |     |    |     |     |     |          |
|       |                                                | TTN | c.107840T>A | p.I35947N | 0.000004 | --- |              |     |     |    |     |     |     |          |
|       |                                                | ATP12A | c.349A>G | p.I117V | 0.00001 | tolerated (0.1) |      |     |    |     |     |     |          |
|       |                                                | ANK2 | c.11231C>A | p.T3744N | 0.001 | tolerated (0.1) |      |     |    |     |     |     |          |

(continued)
| MD-42 | contractures and cardiac arrhythmia, non-consanguineous, sporadic | DYSF | c.3065G>A | p.R1022Q | ar, LGMD (CM090628) | 0.014 | adult 42 no do | yes spo | yes r adi c |
|-------|---------------------------------------------------------------|------|-----------|-------------|-------------------|-------|-------------|-------|----------------|
|       |                                                              | COL6A2 | c.1552C>T | p.P518S    | tolerated (0.2)   |       |             |       |                |
|       |                                                              | COL6A3 | c.4727G>A | p.R1576Q   | tolerated (0.21)  |       |             |       |                |
|       |                                                              | WDR43 | c.366T>G  | p.S122R    | tolerated (0.21)  |       |             |       |                |
|       |                                                              | DYSF | c.3992G>T | p.R1331L   | Dysferlinopathy (CM103814) | 0.016 | tolerated (0.51) |       |                |
|       |                                                              | TRIM32 | c.558G>C  | p.Q186H    | deleted (0.04)    |       |             |       |                |
|       |                                                              | SUN1  | c.335C>T  | p.T112M    | tolerated (0.16)  |       |             |       |                |
|       |                                                              | BVES  | c.275A>G  | p.D92G     | novel deleted (0.01) |       |             |       |                |
|       |                                                              | SGC6  | c.792G>G  | p.I264M    | deleted 3 mths 5 ye yes no un |       |             |       |                |
| MD-45 | contractures and possible mild cardiac involvement, non-consanguineous, sporadic |          |          |            |                   |       |             |       |                |
|       |                                                              | COL6A2 | c.347G>A  | p.S116N    | AD, BTHLM (CM050211)? | 0.031 | ---         |       |                |
|       |                                                              | ITGA7  | c.824G>A  | p.R275H    | tolerated (0.86)   |       |             |       |                |
|       |                                                              | TMEM201 | c.52G>A   | p.G185     | novel tolerated (0.86) |       |             |       |                |
|       |                                                              | COL6A2 | c.3166G>A | p.E106K    | AD, BTHLM (CM050217)? | 0.002 | tolerated (1) |       |                |
|       |                                                              | RYR2   | c.1939C>T | p.R647C    | deleted (0.01)     |       |             |       |                |
|       |                                                              | POMT1  | c.1233C>A | p.D411E    | ---                |       |             |       |                |
|       |                                                              | RYR2   | c.2444C>T | p.P815L    | deleted (0.0)      |       |             |       |                |
|       |                                                              | USP34  | c.3938G>T | p.G1313V   | novel              |       |             |       |                |
|       |                                                              | ANK2   | c.9854T>C | p.I3285T   | deleted (0.05)     |       |             |       |                |
|       |                                                              | SGC6   | c.371T>C  | p.I124T    | LGMD               | 0.00006 | deleted |       |                |
| MD-47 |                                                                  |          |          |            |                   |       |             |       |                |
|       |                                                              |          |          |            |                   |       |             |       |                |
|       |                                                              |          |          |            |                   |       |             |       |                |
|       |                                                              |          |          |            |                   |       |             |       |                |

Please cite this article as: P. Meinke et al., A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism, EBioMedicine (2019), https://doi.org/10.1016/j.ebiom.2019.11.048
|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 49 | SGCG | c.320C>T | p.S107L | 0.000008 | deleterious (0) |
|   | COL6A1 | c.347G>A | p.S116N | AD, BTHLM (CM050211)? | 0.031 | --- |
|   | RYR3 | c.6698G>A | p.R2233Q | | 0.001 | tolerated (0.06) |
|   | RYR3 | c.9254C>G | p.P3085R | | 0.0004 | tolerated (0.27) |
|   | RYR2 | c.3407C>T | p.A1136V | | 0.007 | tolerated (1) |
| MD-50 | DYSF | c.509C>A | p.A170E | Dysferlinopathy (CM053208) | 0.004 | tolerated (0.97) |
|   | COL6A1 | c.347G>A | p.S116N | AD, BTHLM (CM050211)? | 0.031 | --- |
|   | COL6A2 | c.2558G>A | p.R853Q | AD, BTHLM (CM050225)? | 0.004 | tolerated (0.18) |
|   | AGRN | c.5201G>A | p.R1734H | | 0.004 | deleterious (0) |
| MD-51 | TMEM39A | c.778G>A | p.D260N | 0.002 | tolerated (0.62) |
|   | XIRP1 | c.802_805del | p.A269Vfs*6 | novel | --- |
| MD-52 | POMT1 | c.1233C>A | p.D411E | | 0.035 | --- |
|   | COL6A1 | c.347G>A | p.S116N | AD, BTHLM (CM050211)? | 0.031 | --- |
| MD-53 | DMO | c.3595G>A | p.E1199K | novel | --- |
|   | SYNE2 | c.1823A>G | p.E608G | | 0.000004 | tolerated (0.08) |
|   | TMEM43 | c.351dup | p.H118Afs*11 | | 0.00003 | --- |
| MD-54 | SYNE1 | c.17347G>C | p.E5783Q | | 0.000004 | tolerated (0.21) |
|   | FKTN | c.116C>T | p.R56C | | 0.01 | --- |
|   | AGRN | c.1528G>A | p.G510S | | 0.008 | --- |

(continued)
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 mechanosignalling 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 multistage approach. This approach pinpointed candidates in part based on the preferential tissue focus of pathology and in part on the subcellular 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, DISF, 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 caveolin2. 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 gkm cell-less [38]. FLH1 is linked to signal transduction and splice variant FH1B targets specifically to the nuclear envelope [39]. Moreover, FH1B 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, DISF, 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. Meinek reports personal fees from Greenovation Biopharma-
cueticals, outside the submitted work; Dr. Harris reports grants from 
Muscular Dystrophy-UK, during the conduct of the study; Dr. Koelbel 
reports personal fees from Genzyme, personal fees from Sarepta, per-
sonal fees from Biogen, personal fees from AveXis, outside the sub-
mited work; 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 
directed X-linked EMDM 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 neu-
romuscular disease in Northern Ireland. Neuromuscular Disord 1996;6:1:69–73.
[2] Norwood FL, Harling C, Chinnery PF, Eagle M, Bushby K, Straub V. Prevalence of 
gene defects and supports gene misregulation as its pathomechanism, EBioMedicine (2019),https://doi.org/10.1016/j.ebiom.2019.11.048

Please cite this article as: P. Meinek et al., A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism, EbioMedicine (2019), https://doi.org/10.1016/j.ebiom.2019.11.048

ARTICLE IN PRESS

JID: EBIOM
[m5G;December 5, 2019;19:55]
P. Meinek et al. / EbioMedicine 00 (2019) 1–20
20

Please cite this article as: P. Meinek et al., A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism, EbioMedicine (2019), https://doi.org/10.1016/j.ebiom.2019.11.048