Cryptic amyloidogenic elements in the 3`-UTR of neurofilament genes trigger axonal neuropathy

Adriana P. Rebelo[1], Alexander J. Abrams[1], Ellen Cottenie[2,3], Alejandro Horga[2,3], Michael Gonzalez[1], Dana M. Bis[1], Avencia Sanchez-Mejias[1], Milena Pinto[1], Elena Buglo[1], Kasey Markel[5], Jeffrey Prince[5], Matilde Laura[2,3], Henry Houlden[2,3,4], Julian Blake[2,3], Cathy Woodward[2,3], Mary Sweeney[2,3], Janice Holton[2,3], Michael Hanna[2,3], Julia E. Dallman[5], Michaela Auer-Grumbach[6], Mary M. Reilly[2,3*], Stephan Zuchner[1*]

1. Dr. John T Macdonald Department of Human Genetics and John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL 33136, USA

2. MRC Centre for Neuromuscular Diseases, 3. Department of Molecular Neurosciences, 4. Neurogenetics Laboratory, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK

5. Department of Biology, University of Miami, Miami, FL 33146, USA

6. Department of Orthopaedics, Medical University Vienna, Vienna, Austria

* These authors contributed equally

Corresponding Author

Dr. Stephan Zuchner, University of Miami Miller School of Medicine, Biomedical Research Building (BRB), Room 523, LC: M-860, 1501 NW 10th Avenue, Miami, FL 33136; USA; Email: szuchner@med.miami.edu

Abstract

Abnormal protein aggregation is observed in an expanding number of neurodegenerative diseases. We describe a mechanism for intracellular toxic protein aggregation induced by an unusual mutation event in axonal neuropathy families. These families carry distinct frameshift variants in *NEFH,* neurofilament heavy, leading to a loss of the terminating codon and an extended translation of 40 amino acids in the 3'-UTR. *In silico* aggregation prediction suggested the terminal 20 residues of the altered NEFH to be amyloidogenic, which we confirmed experimentally by serial deletion analysis. Expression of this amyloidogenic motif fused to NEFH caused prominent and toxic protein aggregates in transfected cells and disrupted motor neurons in zebrafish. We identified a similar aggregation inducing mechanism in *NEFL* (neurofilament light) and *FUS* (fused in sarcoma), in which mutations are known to cause aggregation in CMT and amyotrophic lateral sclerosis (ALS) respectively. In summary, we present a protein aggregation-triggering mechanism, which should be taken into consideration when evaluating stop-loss variants.

Introduction

Abnormal accumulations of protein aggregates are associated with a wide range of diseases mainly affecting the nervous system[1]. The origins of these aggregates are divers; however, they share similar structures and overlapping mechanisms of cellular toxicity in different diseases. Protein aggregates usually adopt high-ordered beta sheet quaternary structures forming insoluble fibrils termed amyloids[2]. In Huntington's disease (MIM: 143100), intranuclear inclusions and cytoplasmic aggregates are caused by polyglutamine expansion of huntingtin protein[3]. Parkinson disease (MIM: 168600) is characterized by inclusions known as Lewy bodies present in the cytoplasm of neurons[4]. In Alzheimer (MIM: 104300), aggregates can occur both extracellular as neuritic plaques composed of Aβ peptide, and intracellular as neurofibrillary tangles of hyperphosphorylated TAU protein[5]. In Amyotrophic lateral sclerosis (ALS [MIM: 105400]) aggregation of ubiquitinated proteins, including FUS, TDP-43 and OPTN, occurs in degenerating motor neurons[6]. Interestingly, these aggregated ubiquitinated proteins co-occur with perikaryal inclusions of neurofilament (NF)[7]. Abnormal NF aggregation and misassembly has also been reported in other motor neuron diseases such as giant axonal neuropathy (GAN [MIM: 256850])[8] and Charcot-Marie-Tooth disease (CMT [MIM: 607736])[9]; the latter being the most common inherited motor neuron disease. However, the causes for NF accumulation appear to be heterogeneous and not fully understood.

NFs are a class of intermediate filaments exclusively expressed in neurons, representing major components of the cytoskeleton responsible for regulating axonal diameter and growth[10]. NFs are composed of light (NEFL), medium (NEFM), and heavy (NEFH) molecular weight neurofilament proteins that form heteropolymers in order to assemble into 10 nm filaments[11]. Mutations in NF encoding genes have been reported in multiple neurodegenerative diseases. In ALS, NEFH expansion and contraction of the KSP (Lys-Ser-Pro; KSP) repeat motif has been shown to be a risk factor; yet, its clinical contribution to sporadic ALS remains to be determined[12; 13]. Mutations in *NEFL* (MIM: 162280) have been associated

with both axonal and demyelinated CMT (CMT2E [MIM: 607684] and CMT1F [MIM: 607734] respectively) presenting a wide spectrum of clinical phenotypes.[14 15; 16]. The frequency of the *NEFL* mutations in CMT is about 2% and mutations can occur in conserved domains of the protein, including the head (e.g. p.Pro8Arg [c.23C>G, GenBank: NM_006158.4]), coil (e.g. p.Gln334Pro [c.1001A>C, GenBank: NM_006158.4]) and tail (e.g. p.Glu527del [c.1579_1581delGAG, GenBank: NM_006158.4]) domains[16]. Most *NEFL* mutations have been identified as heterozygous, although homozygous mutations have also been reported[17]. Most mutations lead to abnormal NF aggregates and disruption of NF axonal transport and assembly in transfected cells[15]. Mutations in the heat shock *HSBP1* (MIM: 602195), which causes a subtype of CMT, also results in disruption and aggregation of NEFL, thus pointing to the importance of chaperone proteins for neurofilament integrity[18-20].

In the present study, we report previously unrecognized cryptic amyloidogenic elements in the 3'-UTR of *NEFH* (MIM: 162230), which lead to aggregation and neuronal degeneration in model systems and affected individuals. We demonstrate that frameshift variants in *NEFH* in CMT2 families resulted in stop-loss and translation of a cryptic amyloidogenic element in the 3'-UTR. Expression of the mutant *NEFH* exhibited prominent abnormal protein aggregates, disruption of the neurofilament network and altered cell dynamics. Interestingly, we obtained the same aggregation induced phenomenon by triggering the translation of the 3'-UTR of *NEFL* and *FUS* [MIM: 137070]. Our *in vivo* and *in vitro* results show that translation of cryptic amyloidogenic elements in the 3'-UTR of neurofilaments cause axonopathy and could be of broader impact for neurodegenerative diseases.

Material and Methods

Families

The families were identified as part of our ongoing genetic studies in CMT. The families were ascertained in Austria and Great Britain. Participants were recruited, enrolled, and sampled according to the

institutional review board protocols of the University of London and Vienna. A complete description of the study was provided to the subjects, and written informed consent was obtained. Whole blood was collected from all participants by venipuncture. Affection status was determined by consensus of physicians and clinical staff experienced in clinical CMT research, and was based on medical records and in-person evaluation.

Plasmid constructs and immunocytochemistry

The gene encoding the human NEFH was synthesized by Genscript. Site-directed mutagenesis was used to generate the affected individual's variant, c.3010_3011delGA. The GFP tag was introduced at the N-terminus of the gene by cloning into pcDNA3.1/NT-GFP-Topo (Invitrogen). *NEFL-Myc* cloned into pCMV6-entry vector was obtained from OriGene technologies. Neuro-2a cells were grown in complete DMEM media (Gibco) to 75% confluence and transfected with Lipofectamine 2000 (Invitrogen) following manufacturer's protocol. After 24hr, cells were fixed with paraformaldehyde for 20 min, permeabilized with cold methanol for 5 min and stained with anti-myc antibody (Cell Signaling) and anti-tubulin (Invitrogen). Cells were mounted onto microscope slides and imaged with a confocal microscope, Zeiss LSM710, using a 60X objective lens.

Co-immunoprecipitation and Western Blot

Immunoprecipitation was performed using the Pierce™ Crosslink Magnetic IP/Co-IP Kit and manufacturer's instructions were followed. Briefly, 500 ug of total cell lysates were incubated with 5 ug of either anti-GFP ChIP grade antibody (Abcam) or control Rabitt IgG (Santa Cruz) at 4°C overnight. The lysate/antibody mixture was incubated with protein A/G magnetic beads (Thermo Scientific) at room temperature for 1 hr on a rotator mixer. Beads were collected using a magnetic stand. Beads were washed twice with manufacturer's IP Lysis/wash buffer. Samples were boiled with a lane marker

containing reducing agent at 100°C for 5 minutes. Protein samples were analyzed by SDS-PAGE followed by western blot using appropriate antibodies. The following antibodies, diluted 1:1000, were used: mouse monoclonal anti-GFP (Santa Cruz), mouse monoclonal anti-Myc (Cell Signaling) and rabbit polyclonal anti-kinesin (Abcam).

IncuCyte analysis

Neuro-2a cells transfected with *GFP-WT-NEFH* and *GFP-FS-NEFH* were plated in a 24 well plate and 24 h after transfection, cells were incubated in an IncuCyte live-cell imager system (Essen Instruments) for 2 days at 37 °C and 5% CO_2. Replicates of 12 wells were used for each group (wild-type and mutant). Time lapse phase-contrast and GFP images were taken every 3 h for 24 h. A total of 36 images were acquired per well for each time point. The IncuCyte Zoom software was used to calculate cell size, confluence and eccentricity (roundness).

Transmission electron microscopy and analysis

Neuro-2a cells were cultured in 25 cm^2 flasks (Corning) and transfected with plasmids as previously described. After 2 days, cells were fixed for four days in 2.5% glutaraldehyde in Millonig's phosphate buffer, post-fixed in 1% O_SO_4 and uranyl acetate followed by dehydration in an ethanol series and embedded in Spurs. Ultrathin, 85 nm sections were made using a Leica microtome and were stained with cold lead citrate for 10 minutes.[114] Images were acquired with a Joel JEM-1400 transmission electron microscope with a digital Gatan camera. All images were processed and analyzed using Fiji (image J and photoshop CS5).

Zebrafish studies

Experiments were carried out using wildtype or transgenic tg(*Olig2:DsRed*) in a mixed AB/TL background. Adults were kept on a 14 hour light/10 hour dark cycle at 28°C. Embryos were obtained from natural crosses after removing a divider at first light. mRNAs were synthesized from the *GFP-WT-NEFH* and *GFP-FS-NEFH* (pcDNA3.1/NT-GFP-Topo) plasmids after linearization with NotI, using mMessage mMachine T7 Ultra (Ambion) and 400pg of RNA were microinjected into one cell stage embryos. Embryos were reared in petri dishes in a 28°C incubator with the same light dark cycle. Motor neuron outgrowth was assayed at 48 hpf in embryos obtained from tg(*Olig2:DsRed*) crosses. Live fish were anesthetized with tricaine methanesulfonate (Sigma), placed against a shelf of 1.5% agarose, and imaged using a Leica confocal microscope with a 20x lens. 1 µm Z-stacks were taken between segments 6 and 15 and the lengths of the first 4 caudal anterior primary axons were measured using Simple Neurite Tracer in Fiji. Images in the figure were displayed with LUT:edges in order to enhance contrast. GFP fluorescence intensity was assayed in microinjected wildtype embryos at 24 hpf by similar methods. Maximum intensity z-stacks were compiled and the trunk of the fish excluding the yolk was traced from the DIC image. Fluorescence intensity represents mean gray values between 0-255, which were calculated with the analyze measure function in Fiji. For Western blots, 20-50 embryos were manually dechorionated at 24 hpf and were processed through a batch deyolking method[21]. Deyolked and washed embryos were homogenized in 50 µl of RIPA with protease inhibitor using Fisherbrand disposable pestle system followed by sonication and freeze thaw at -80. The supernatant was collected after centrifugation and protein content was measured. NuPage sample buffer (Invitrogen) and reducing agent were added, samples were boiled, and approximately 12 µg of total fish proteins were loaded into SDS gel. For detection, anti-GFP tag antibody (Abcam) and anti-tubulin (Santa Cruz sc-9104) were used at 1/1000 concentration, and SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) was used following incubation with respective HRP conjugated secondary antibodies. For RT-PCR validation, RNA was extracted from approximately 20 dechorianted embryos at 24 hpf in TRIzol

(Invitrogen). The cDNA with generated with SuperScript III (Invitrogen) using random hexamers and PCR was performed with GFP-NEFH transcript primers: 5'-TTTTACCAGACAACCATTACCTG-3' and 5'-GGCTAGCGCGTAGTGGAG-3' or control (slc25a46) 5'- GCCACTGGGTGACGACTC-3' 5'-GAAGCGGAAGAAGTCGTTTG-3'. All experiments were conducted in accordance with the University of Miami Institutional Animal Care and Use Committee guidelines.

Sanger and exome sequencing

We performed exome sequencing in 269 index individuals with autosomal dominant CMT. 16 CMT families were from United Kingdom and 48 families were from Austria. The remaining families were from several other countries from Europe, North and South America. The SureSelect Human All Exon 50MB kit (Agilent, Santa Clara, CA, USA) was used for in-solution enrichment and the HiSeq2500 instrument (Illumina San Diego, CA, USA) was used to produce 100bp paired end sequence reads. BWA (24), Picard and GATK (25) software were used to align sequence reads and call variants. This data were imported into GEM.app (16) for further analysis. Variants were filtered for variants that segregated in an autosomal dominant fashion and met the "strict" criteria, which requires that variants are rare (NHLBI ESP6500 MAF<0.05%), present in less than three families within GEM.app (~4,300 exomes), conserved (GERP score > 2 or PhastCons score > 0.6), and have sufficient quality scores (Genotype Quality > 75). Mutations in known CMT-associated genes were absent in both Austrian and UK families. *NEFH* variant calls identified were validated using conventional Sanger sequencing. In addition, we had access to a control sample of 5200 from the GEM.app/GENESIS database. These samples broadly include 1824 neuromuscular disorders (of which 470 where peripheral neuropathies), 286 cardiomyopathies, 188 dementias, 509 deafness. For more details refer to Gonazlez et al[22].

The accession number for the *NEFH* sequence reported in this paper is GenBank: NM_021076.3.

Results

Identification of *NEFH* frameshift variants in CMT2 families.

We performed whole exome sequencing on three affected individuals belonging to different generations of a British family (UK1) diagnosed with autosomal-dominant CMT2 (Figure 1A). Exome data was analyzed with a strict filtering approach for segregation of non-synonymous heterozygous variants using the Genomes Management Application (GEM.app) software[23; 24]. A heterozygous frameshift variant in *NEFH* was identified as a top candidate for the disease from a list containing six additional variants (Table S1). *NEFH* was selected because neurofilament abnormalities have been previously reported in neurodegenerative diseases, including ALS[25]. In addition, mutations in *NEFL*, another major neurofilament component, also cause CMT[16]. The variant co-segregated with the phenotype across three generations in this family (Figure 1A). This variant, (c.3010_3011delGA, p.Asp1004Glnfs*58, chr22:29,886,637) affects the last coding exon and shifts translation into an alternative open reading frame (ORF) resulting in continued translation of an additional 40 amino acids beyond the stop codon in the original ORF (Figure 2A). The mutant protein retains its major functional domains, including the head, rod and tail domains (Figure 1C). We then screened an additional 322 CMT families with whole exome sequencing. We identified an Austrian CMT2 family (AT1) with four affected individuals carrying a nearby heterozygous frameshift variant in *NEFH* (c.3017_3020dup, p.Pro1008Alafs*56, chr22:29,886,645). This insertion interestingly also results in a stop-loss mutation and translation of the identical ORF as observed in family UK1 (Figure 2A). Co-segregation of the distinct variants with the phenotype was confirmed by Sanger sequencing in both families (Figure 1B). We were unable to sequence DNA from the deceased mother in family AT1, who was likely a carrier of the pathogenic variant. In approximately 5,000 additional exomes[22], with a wide range of clinical phenotypes, including other neuropathies, from our own collection we did not observe these mutations and also did not observe other frameshift variants in the last exon of *NEFH*.

Clinical features of family UK1 harboring p. Asp1004Glnfs*58 ***NEFH*** **variant.**

The proband of UK1 (IV:1; Figure 1; Table 1) had a normal birth and early development. He walked independently at age 18 months. Between age 2 and 4 years, he started having frequent falls with a tendency to turn over on his ankles, and was noted to have high arches and muscle wasting of the legs. Over the years he noticed progressive lower limb weakness, sensory loss in the feet, and occasional cramps. He started using support for walking in his early 20s and began using a wheelchair in his mid-20s. He had surgery for pyloric stenosis in his infancy; medical history was otherwise unremarkable.

Neurological examination at age 16 revealed minimal distal wasting in the four limbs, mild weakness of first dorsal interosseous and ankle dorsiflexion (Medical Research Council grade 4+), and reduced pinprick and vibration sense to the ankles. Subsequent examinations revealed proximal weakness and mildly increased tone in the lower limbs. At age 22, he walked with a slightly waddling and stiff gait and had a positive Gower's manoeuvre. Examination at age 24 revealed wasting of the intrinsic hand muscles and below the knees, and weakness of first dorsal interosseous (4), abductor pollicis brevis (4+), hip flexion (4+), knee flexion and extension (4+), ankle dorsiflexion (4), and ankle plantar flexion (2). Reflexes were present in the upper limbs and absent in the lower limbs. Plantar responses were flexor. Pinprick sensation was reduced in the fingers and to the proximal half of the calves. Vibration sense was reduced to the knees with a Rydel-Seiffer tuning fork. Proprioception was normal.

Nerve conduction studies (Table 1) were consistent with a motor and sensory axonal neuropathy predominantly affecting the lower limbs. EMG revealed proximal and distal chronic neurogenic changes in a non-length-dependent pattern, with additional myopathic features in proximal muscles. Central motor conduction times to the lower limbs were mildly prolonged bilaterally. Creatine kinase (CK) levels were raised at 721-1288 IU/L (38–204). Plasma lactate was elevated at 1.84 mmol/L (0.5–1.65) in one out of three measurements. CSF analysis was normal.

A muscle biopsy of vastus lateralis showed increased variation in fiber size, mild increase in connective tissue, frequent round and angular atrophic fibers of all types, nuclear bag fibers, group atrophy, and grouping of type I and IIA fibers. Split fibers and occasional regenerating fibers were also noted. Several fibres contained internal nuclei and three contained rimmed vacuoles. Many fibers had a disturbed architecture, but no core-targetoid fibers were observed. There were no cytochrome *c* oxidase-negative or ragged-red fibers. Immunohistochemistry revealed a patchy increase in myotilin staining. Respiratory chain enzyme assays of muscle homogenate revealed reduced activity of all complexes (complex I: 0.077 [reference range: 0.104-0.268]; complex II+III: 0.032 [0.040-0.204]; and complex IV: 0.008 [0.014-0.034]). A repeated assay revealed complex I and IV activities at the lower limit of the reference range (complex I: 0.106; complex IV: 0.015). Sequence analysis of *PMP22* (MIM: 601097), *MPZ* (MIM: 159400), *GJB1* (MIM: 304040), *MFN2* (MIM: 608507), *GDAP1* (MIM: 606598), *NEFL*, *BSCL2* (MIM:606158), *TRPV4* (MIM: 605427), *HSPB1*, and *HSPB8* (MIM: 608014) was negative. 17p11.2 rearrangements were excluded.

The father of the proband (III:1; Figure 1A; Table 1) had used shoes with ankle support since age 15. In his 40s, he started noticing occasional tripping episodes and cramps but was able to run until age 48. He started using support for walking at age 53. Neurological examination at age 57 showed mild distal wasting in the upper and lower limbs, and weakness of first dorsal interosseous (4), ankle dorsiflexion (4), and ankle plantar flexion (4). Reflexes were present except at the ankles. Plantar responses were flexor. Pinprick sensation was reduced to mid palms and to just above the ankles. Vibration sense was reduced to the knees with a Rydel-Seiffer tuning fork. Proprioception was normal.

The grandmother (II:1; Figure 1A; Table 1) had problems standing on her toes and a flat footed walk since her early 20s. In her 50s, she started using support for walking and developed problems with fine hand movements. She became wheelchair-bound in her 70s, and developed hearing loss in her mid-70s. Neurological examination at age 79 revealed muscle wasting of intrinsic hand muscles and from the

mid-thighs down. There was weakness of first dorsal interosseous (3), abductor pollicis brevis (3), hip flexion (4), knee extension (4), knee flexion (4+), ankle dorsiflexion (0), and ankle plantar flexion (4+). Reflexes were all absent. Pinprick was reduced to just below the knees. Vibration sense was reduced to left elbow, right shoulder and costal margins. Proprioception was normal.

Neurophysiological studies of individuals III:2 and II:1 (Table 2) revealed similar findings to that observed in the proband except for the absence of myopathic changes. Nerve conduction studies were consistent with a motor and sensory axonal neuropathy, which was severe in II:1. EMG revealed proximal and distal chronic neurogenic changes in a non-length-dependent pattern.

Our findings suggest an apparent anticipation pattern in this family, since the proband had an earlier age of onset with more severe manifestations compared to the previous generations. Increased severity of the disease over generations has been reported in other CMT families[26; 27]; however, the mechanism responsible for this phenomenon is unclear. It is possible that a modifier gene can affect the severity of the clinical phenotype; however, it is very challenging to identify a potential modifier gene from our exome data with that small sample size.

Clinical features of family harboring p.Pro1008Alafs*56 *NEFH* variant.

The proband of family AT1 (III:1; Figure 1B) was first brought to neurological attention at the age of 38 years when he noticed weakness in the lower limbs. He initially complained problems in climbing stairs, and running became impossible. There were no problems in the hands and no sensory disturbances. At examination there was mild atrophy in the lower limbs but this was more pronounced distally as was muscle weakness. Tendon reflexes were reduced to absent. There was mild *pes cavus* bilaterally. Nerve conduction velocity (NCV) studies revealed normal values in the upper limbs but slowing of the motor peroneal and tibial nerves bilaterally as well as low amplitudes pointing to an axonal neuropathy. NCV of the sural nerve was normal but the amplitude was reduced as well. Electromyography of the tibialis

anterior muscle revealed prominent chronic neurogenic disturbances with spontaneous activity. The disease was considerably progressive. Re-examination after 10 years revealed prominent proximal and distal weakness in the lower limbs. The individual walked with a very unsafe gait. Mild weakness and wasting then became also evident both in distal and proximal muscles (Table 1). CK-levels have been elevated from the beginning. Mutations in the following genes were excluded by direct sequencing: *PMP22, MPZ, NEFL, LITAF* (MIM: 603795), *GDAP1, YARS* (MIM: 603623), *MFN2, HSPB1, HSPB8, DNM2* (MIM: 602378 [ex13-16]), *FBLN5* (MIM:604580), *HINT1* (MIM: 601314).

Three sibs were reported to be similarly affected, but in them a combination of a neurogenic and myopathic disorder was identified. NCS revealed a mixed axonal-demyelinating neuropathy in all individuals, but detailed values were not available (Table 2). Notably, individual III:4 was initially diagnosed with polymyositis and treated appropriately. Later on, an inherited neuropathy was suspected and mutations in several genes known to cause peripheral neuropathies (*LMNA* [MIM: 150330], *MPZ, HSPB1, HSPB8, NEFL, MFN2*) were excluded for mutations. Neurological evaluation of the parents was not performed during this study. The mother (II:1) had died early whereas the father (II:2) was reported to be neurologically normal up to an advanced age.

Identification of cryptic amyloidogenic elements in *NEFH* and *NEFL* 3'-UTRs.

Because neurofilaments have a considerable tendency to aggregate in neurodegenerative diseases, we decided to investigate the intrinsic aggregation propensity of the extension of amino acids present in *NEFH* in the two CMT2 families. We used the web-based aggregation prediction tool TANGO[28] to analyze aggregation-prone segments based on the physico-chemical principles of beta-sheet formation. According to this algorithm a segment is predicted to aggregate when it contains at least five consecutive residues with a TANGO score above 5%[28]. Analysis of the 3'-UTR extension of amino acids present in the NEFH mutant p.Asp1004Glnfs*58, showed a hot spot for aggregation in eight stretches of

amino acids (QFSLFLSL) with a combined score of 250 (Figure 2B). Aggregation propensity was also analyzed with other prediction tools including AGGRESCAN[29], FoldAmyloid[30] and PASTA 2.0[31] that use different algorithms to predict aggregation. All tested tools detected aggregation and an amyloidogenic region in an overlapping stretch of amino acids of the mutant extension (Figure 2C and Figure S2). We refer to this stretch of amino acids predicted to induce aggregation as a cryptic amyloidogenic element (CAE). Prediction analysis of the other two open reading frames of the NEFH-3'UTR resulted in no aggregation for ORF1 and high aggregation scores for ORF2 (Figure 2B). This data suggests that any frameshift variants resulting in loss of the *NEFH* stop codon and translation of cryptic amyloidogenic elements in 3'-UTR would have the potential to cause protein aggregation.

Next we investigated *in silico* the aggregation propensity of the *NEFL* 3'-UTR region, since NEFL aggregation has been reported for ALS[6] and CMT[18]. Interestingly, only ORF1 of the 3'-UTR was positive for aggregation prediction with a high TANGO score of 443 comprising the motif ISLIISGII (Figure 2E). Positive scores for ORF1 were also obtained with AGGRESCAN, FoldAmyloid, and PASTA2.0 (Figure 2F and Figure S2). Thus, our *in silico* results for *NEFL* suggest that only missense mutations affecting the STOP codon resulting in translation of 3'-UTR ORF1 have the potential to cause aggregation. Importantly, because the 3'-UTR CAE of *NEFL* is in the first reading-frame, we should also consider the possibility of leaky translational read-through caused by reduced accuracy of translation termination. Therefore, NEFL aggregation would be a concern particularly for individuals treated with stop codon read-through inducing drugs, such as gentamicin G418[32].

Prominent aggregation in cultured Neuro-2a cells expressing neurofilaments encoding 3'-UTR CAEs.

Following our *in silico* analysis, we investigated the potential of the 3'-UTR CAEs identified in affected individuals to cause aggregation in Neuro-2a cells. Constructs were created encoding GFP-tagged wild type protein (*GFP-WT-NEFH*) and frameshift mutant protein encoding 40 additional amino acids of ORF3

(*GFP-FS-NEFH*), as identified in the UK1 family (p.Asp1004Glnfs*58). *GFP-WT-NEFH* transfection led to evenly distributed expression in the cytoplasm of Neuro-2a cells (Figure 3A). Expression of *GFP-FS-NEFH* revealed prominent abnormal perinuclear aggregation after 24hrs post transfection (Figure 3A). Quantification shows that over 75% of cells transfected with *GFP-FS-NEFH* contained aggregates compared to less than 1% in *GFP-WT-NEFH* cells (Figure 3A). Cells transfected with wild-type NEFH retained their typical Neuro-2a cell morphology with 'axon-like' projections extending from the cell body. By contrast, *GFP-FS-NEFH* expressing cells were round-shaped and their axon-like projections were significantly reduced (Figure 3A).

In order to experimentally identify which of the additional 40 amino acids are responsible for the aggregation we created a series of truncated constructs harboring stop codons (STOP1-STOP4) at different positions throughout the extension of amino acids (Figure S3). Cells transfected with constructs STOP1, STOP2, and STOP3 did not form aggregation in cells (Figure S3 and Figure 3B) and their subcellular distribution pattern was identical to *GFP-WT-NEFH*. We observed aggregates in a small proportion of cells (20%) transfected with the longest construct, *GFP*-STOP4-*NEFH* (Figure S3 and Figure 3B); however, aggregation was still more severe in the full-length extension described above (*GFP-FS-NEFH*). To challenge the predicted CAE, we cloned the most distal 22 amino acids predicted to cause aggregation (SSRIRVTQFSLFLSLCKKKLLR) directly *in frame* with the C-terminus end of the *GFP-WT-NEFH* to create the *GFP-NEFH*-CAE construct. As expected, cells transfected with the *GFP-NEFH*-CAE construct caused prominent aggregation at the same level observed in cells transfected with the *GFP-FS-NEFH* (Figure 3B). These results prove that the most distal 22 amyloidogenic amino acids are sufficient and necessary for the formation of aggregates.

We also tested the ability of the 3'-UTR CAE present in the *NEFL* to cause aggregation in cells. Neuro-2a cells were transfected with constructs encoding GFP-tagged NEFL (*GFP-WT-NEFL*), NEFL without Stop codon fused *in-frame* with the NEFL-3'-UTR containing the predicted CAE (*GFP-NEFL-*

ORF1), and NEFL fused with the ORF3 3'-UTR, which was not predicted to contain CAE (*GFP-NEFL*-ORF3). Although a few cells transfected with GFP-WT-NEFL presented aggregates due to the self-assembly nature of NEFL, a large portion of cells adopted neurofilament-like structures, an indication of neurofilament assembly, in about 45% of transfected cells (Figure 3C). These filamentous structures can also be explained by the ability of NEFL to self-assemble, in contrast to NEFH, which is an obligate heteropolymer and it requires interaction with either NEFL or NEFM in order to assemble into neurofilament structures[19]. However, all cells transfected with *GFP-NEFL*-ORF1 formed prominent aggregation and adopted a rounded shape without forming filamentous structures (Figure 3C). In contrast, cells transfected with *GFP-NEFL*-ORF3 resulted in neurofilaments structures comparable with the *GFP-WT*-NEFL. Those results suggest that specific translation of 3'-UTR CAE present in the first open reading-frame is required for the formation of aggregates.

Evaluation of cells expressing the *NEFH*-3'-UTR CAE.

Because we observed a dramatic difference in cell shape and viability in cells expressing *GFP-FS-NEFH* compared to wild-type NEFH, we decided to quantify those cellular features. Western blots were performed to show that the levels of protein expression in Neuro-2a cells transfected with *GFP-WT-NEFH* and *GFP-FS-NEFH* were similar (Figure 4A). Therefore, differences in aggregation cannot be due to protein expression levels. To analyze these cells we used the Incucyte imaging system with time-lapse images taken every 3 hours between 24-48hrs post transfection. We observed that the average green (GFP) object area (μm^2) per cell was smaller in cells expressing *GFP-FS-NEFH* (~ 120 μm^2) compared to *GFP-WT-NEFH* (>250 μm^2), suggesting that *GFP-FS-NEFH* cells were about 2X smaller (Figure 4B). The percentage of confluence of *GFP-WT-NEFH* positive cells increased with time from 2% to 6.5%, while the confluence of *GFP-FS-NEFH* remained below 2% (Figure 4C). This indicates decreased cell viability after *GFP-FS-NEFH* transfection. In addition, 48hrs post transfection, cells expressing *GFP-FS-NEFH* started to

detach from plates, whereas cells transfected with the *GFP-WT-NEFH* were still attached and viable several days after transfection. The average green object eccentricity, a parameter that measures object roundness from 0 to 1, where 0 represents a perfect circle, confirmed that *GFP-FS-NEFH* cells have a more rounded shape (Figure 4D). We concluded that the abnormal aggregation of mutant NEFH might cause a toxic gain-of-function effect leading to loss of neuronal characteristics and loss of cell viability.

To further characterize the NEFH aggregation structures, cells were stained with thioflavin T, a dye commonly used to stain amyloid fibrils with beta-sheet structures in individual tissues[23; 33]. Confocal fluorescent imaging showed strong thioflavin T staining of NEFH aggregates, suggesting a fibrillary amyloid-like type of structure (Figure 4F). In order to further understand the composition of the aggregates, we performed transmission electron microscopy in transfected Neuro-2a cells. In cells transiently expressing *GFP-FS-NEFH*, we observed a disordered arrays of filaments of approximately 10nm in diameter consistent with neurofilament size (Figure 4E). Similar disordered neurofilament inclusion structures were previously reported in the anterior horn cells in an individual with ALS harboring a SOD1 (MIM: 147450) mutation[34]. These disordered arrays of filaments were observed in 12/100 images of cells expressing *GFP-FS-NEFH* in no less than 5 different cells and were never observed in 0/40 *GFP-WT-NEFH* images or 0/20 images of untransfected cells. Although we cannot distinguish untransfected from transfected cells in the EM images, transfection efficiency was ~70% and we analyzed close to 20 rounded cells transfected with *GFP-FS-NEFH*, which is a feature of cells containing severe aggregates. Based on the frequency of this feature in cells expressing *GFP-FS-NEFH* and absence in cells expressing *GFP-WT-NEFH* or untransfected cells, we trust these structures to be the aggregate.

Expression of NEFH mutant protein disrupts the neurofilament network in cultured Neuro-2a cells.

In order to see the effect of the expression of the mutant NEFH on the neurofilament network, we co-transfected *GFP-NEFH* constructs with a plasmid encoding NEFL fused to a Myc-tag at the C-terminus

(NEFL-Myc). As expected, the *GFP-WT-NEFH* co-localized with NEFL-Myc protein and assembled into organized neurofilament like structures (Figure 5A). The GFP-FS-NEFL also co-localized with NEFL-Myc, but within the massive aggregates, suggesting arrest and co-aggregation of NEFL and consequently disruption of the neurofilament network (Figure 5A). Co-immunoprecipitation experiments were performed to confirm interaction between the mutant NEFH and NEFL. Cell lysates were immunoprecipitated with an anti-GFP antibody to pull-down NEFH. Western blot confirmed that NEFL-Myc was co-immunoprecipitated in cells transfected with either *GFP-WT-NEFH* or *GFP-FS-NEFH* (Figure 5B). The reverse co-immunoprecipitation with NEFL-Myc pull-down also confirmed the interaction. We further detected kinesin in the co-immunoprecipitate in cells transfected with either wild-type or mutant NEFH (Figure 5B). These results suggest that the mutant NEFH is trapping NEFL, kinesins and possibly other interacting proteins into the aggregates and consequently blocking these proteins from performing their proper functions. Hence, this stop-loss NEFH mutation found in individuals with CMT2 is most likely a toxic gain-of-function mutation. To determine whether the aggregates were affecting exclusively the neurofilament network or additional cytoskeleton components, cells were stained with a tubulin antibody. The microtubule network was normally distributed and assembled in cells transfected with either *GFP-WT-NEFH* or *GFP-FS-NEFH* indicating it was not affected by the aggregates (Figure 5C).

The neurofilament network has been shown to be important for the spatial subcellular distribution of mitochondria[35], therefore cells were stained with an antibody against the mitochondrial outer membrane protein TOM20 (MIM: 601848). Mitochondria were evenly distributed in cells expressing *GFP-WT-NEFH*; however, in cells expressing *GFP-FS-NEFH*, mitochondria accumulated adjacently to the NEFH aggregates (Figure 5D). Similarly, it has been shown that *NEFL* mutations linked to CMT cause altered mitochondrial distribution and co-localization with aggregates[36]. Our results support the importance of neurofilament integrity in proper mitochondrial distribution.

***NEFH* frameshift variant disrupts motor neuron development in zebrafish.**

In order to assess the effects of the *NEFH* frameshift variant *in vivo,* we injected RNA into one-cell stage zebrafish embryos. Equal amounts of RNA encoding either *GFP-WT-NEFH* or *GFP-FS-NEFH* were injected into transgenic Tg(*Olig2:DsRed*)[37] embryos at a dosage at which there was no apparent effect on body morphology (Figure 6A), but a measurable difference in motor neuron outgrowth (Figure 6B). We assessed the common path of the caudal anterior primary motor neurons at 48 hours post fertilization (hpf). We found that the *GFP-FS-NEFH* RNA injected embryos have significantly decreased axon lengths compared to both *GFP-WT-NEFH* and uninjected larvae, while there is no significant difference between the motor neurons lengths of the uninjected embryos and embryos injected with *GFP-WT-NEFH* RNA (Figure 6C). This supports the pathogenicity of *NEFH* mutations and indicates that the addition of the 3'-UTR CAE can function through a toxic gain-of-function mechanism.

To correlate the observed phenotype with the expression of the NEFH proteins, we assessed GFP fluorescence at 8 hpf (data not shown) and 24 hpf (Figures 6D and 6E), and quantitatively by western blot at 24 hpf (Figure 6F). Both methods shown markedly decreased expression of the mutant NEFH, which is likely due to the toxicity of the protein. We further confirmed that the decreased expression of the mutant NEFH occurs at the protein level by verifying the presence of microinjected RNA (Figure 6G). We speculate that the protein quality control mechanisms are efficient in larval zebrafish and that toxic misfolded proteins are rapidly degraded. Because motor neuron disease is a progressive degeneration, it is possible that aggregates can form during aging as protein turnover slows down. The effect of the *NEFH* frameshift variants even among the decreased protein levels further confirms the pathogenicity of the translation read through of the 3'-UTR of the NEFH protein.

Genome-wide analysis of human 3'-UTR CAE

In order to investigate the occurrence of additional potential candidate genes for aggregation caused by a cryptic amyloidogenic elements in the 3'-UTR, we performed a bioinformatics aggregation prediction analysis of all human 3'-UTR sequences. Human 3'-UTR sequences were acquired from the UTRef section of UTRdb, a curated collection of eukaryotic 5' and 3'-UTRs. The UTRef section contains 34,619 3'-UTR sequences from genes retrieved from the National Center for Biotechnology Information (NCBI) RefSeq transcripts[38]. These sequences were translated into the three forward reading-frames to simulate stop-loss mutations caused by either missense (frame 1) or frameshift (frames 2 and 3) mutations. After filtering out amino acid sequences with over 90% similarity and genes of uncertain function (LOC symbols), approximately 12,400 genes per reading-frame were annotated with the aggregation prediction programs, TANGO and PASTA. Next, sequences were filtered for highly stringent threshold aggregation scores, above 200 for TANGO and below -4 for PASTA. These score cutoffs were based on the aggregation prediction scores obtained for the *NEFH* 3'-UTR. Sequences that lack an alternative stop codon were filtered-out since they would likely be degraded by the non-stop decay mechanism. It has been reasoned that the stability of stop-loss mRNAs and/or proteins decreases as the distance between the mutated stop codon and the next alternative stop codon increases[39]. Therefore, only sequences containing an alternative stop codon within 50 amino acids were considered. After these filter criteria were applied, we obtained 4,861 genes, approximately 1,600 genes per reading-frame, containing a 3'-UTR sequence that has a high potential for aggregation if translated (Figure S4).

Although our results suggest that a large number of genes have the potential to cause aggregation as the result of a stop-loss mutation, several physiological factors that vary in different intracellular micro environments influence aggregation, such as, temperature, pH, pressure and protein concentration[40]. Moreover, the frequency of stop-loss caused by a missense mutation within the stop codon (frame 1) is very low, 0.027% (609 of 2,207,918 variants) as observed in the NHLBI GO Exome Sequencing Project. Finally, in order to cause significant protein aggregation disease, the protein must

be expressed in cells that can be negatively impacted by aggregations such as postmitotic neurons, and protein expression levels must overwhelm the cell's ability to clear aggregations.

Next, we filtered for disease-associated genes previously reported to aggregate that contain a predicted 3'-UTR-CAE in any frame. We obtained a list with the top 21 high-risk aggregation genes (Table S2). Although variants in these genes have already been shown to cause aggregation, stop-loss mutations caused by translation of the 3'-UTR-CAE have not been reported yet. Because the genes listed encode proteins prone to aggregation, they may be higher susceptible to aggregate by our proposed mechanism. We decided to experimentally validate three known ALS-associated genes from that list: *FUS, SOD1* and *TARDBP* (MIM: 605078). Neuro-2a cells transfected with *FUS* fused in frame with its predicted 3'-UTR-CAE frame 1 (*GFP-FUS-CAE*) (Figures 7A and 7B) show prominent aggregation in the cytoplasm and neuronal projections of transfected cells (Figure 7C). In comparison, wild-type FUS (*GFP-WT-FUS*) localized to the nucleus. Cells transfected with *SOD1* and *TARDBP* fused with their respective 3-UTR-CAE did not result in protein aggregation (Figure S5). These results show that aggregation induced by translation of the 3-UTR CAE is not a neurofilament exclusive phenomenon; however, it is important to validate the bioinformatics aggregation prediction of the 3-UTR-CAE since other intrinsic protein factors might interfere with protein structure.

Discussion

We describe a protein aggregation inducing mechanism triggered by expression of a cryptic amyloidogenic element, CAE, in the 3'-UTR of neurofilament genes. We report frameshift variants in *NEFH* in two multigenerational autosomal-dominant CMT2 families. Despite distinct stop-loss mutations in each family, both led to an alternative open reading frame and translation of an additional 40 amino acids of the 3'-UTR. Computer algorithms to predict aggregation identified a CAE, in the 3'-UTR comprising the last 20 amino acids of the mutant extension. Interestingly, analysis of the 3'-UTR of the

NEFL, a previously described CMT-associated gene and aggregation-prone protein in ALS, revealed a CAE in ORF1 of the 3'-UTR. Neuro-2a cells transfected with both *NEFH* and *NEFL* expressing their predicted amyloidogenic regions resulted in rounded cells with prominent perinuclear aggregates and lack of axon-like projections. A series of expression constructs for *NEFH* and *NEFL* demonstrated that aggregation is specifically caused by translation of the reading-frame predicted to contain an amyloidogenic sequence and it is independent of the length of the extension of the peptides. For NEFL, the presence of a CAE in-frame with ORF1 of the *NEFL*-3'-UTR could induce aggregation if a stop-loss missense mutation or translation termination read-through occurred. *NEFL* is terminated by the UGA codon, which has been demonstrated to be the least efficient termination codon during aminoglycoside-induced read-through[41]. Therefore, prolonged use of aminoglycoside antibiotics, such as gentamicin could potentially induce aggregation of NEFL. Interestingly, neurotoxic effects, including peripheral neuropathy have been associated with antibiotic use[42]. Further studies of NEFL aggregation induced by translational read-through activating drugs will be required.

Amyloid fibrils and neurofibrillary tangles are associated with severe neurodegenerative disease due to accumulation of toxic protein aggregates, including neurofilament aggregation in motor neuron diseases[9; 18]. Although variants in *NEFH* have been associated with increased susceptibility to ALS, *NEFH* does not represent a high-penetrance ALS associated-gene[12; 43]. Further, the nature of the mutations in our CMT individuals is distinct from the coding deletions or insertions of KLS repeats in *NEFH* previously described in ALS individuals[12; 43]. Our functional studies demonstrated that expression of NEFH containing additional amino acids encoded by the 3'-UTR is detrimental to transfected Neuro-2a cells, resulting in morphologically abnormal rounded cells with fewer projections, reduced viability, and prominent NEFH protein aggregates. In addition, cells expressing the CMT2 NEFH mutant (*GFP-FS-NEFH*) revealed co-aggregation with NEFL, disruption of the neurofilament network, sequentially altering cell morphology and mitochondrial distribution. Moreover, zebrafish embryos injected with mRNA encoding

the mutant *NEFH* resulted in significantly decreased lengths of motor neuron axons compared to those injected with wild-type. Our results are in line with previous studies demonstrating that mutations in *NEFL* and associated proteins, such as HSPB1 and KIF1A, cause neurofilament aggregation in motor neuron diseases including axonal CMT, distal motor neuropathy, and hereditary spastic paraplegia (MIM: 6110357)[18; 44]. These results combined make it evident that motor neurons are particularly susceptible to neurofilament dysfunction. Therefore, these overlapping, but distinct diseases caused by neurofilament abnormalities may share similar mechanism of pathogenesis. Our genetic, *in vitro,* and *in vivo* studies confirm previous hypotheses that aggregation of neurofilaments, and especially NEFH, is a powerful cytotoxic event that has detrimental effects on motor neurons. Because the stop-loss mutant NEFH studied retains all major domains intact, the mutant protein likely retains the ability to interact with other proteins. Yet, we show that these interacting proteins are trapped into the aggregates and thus removed from their normal function. Aggregation of NFs in the cell body may consequently impede transport along axons, since NFs are major components of the axonal cytoskeleton. The fact that NF proteins have long half-life and slow turnover rates probably contributes to aggregate accumulation over time *in vivo*. Therefore, studies aiming to help minimize or clear the aggregates could be beneficial to treat affected individuals.

Finally, our bioinformatics aggregation prediction analysis of the entire human 3'-UTR collection show that a large number of genes could potentially be affected if translated into a specific reading-frame where a CAE is present. We experimentally validated aggregation induced by translation of *FUS*-3'-UTR CAE confirming that this is not a neurofilament exclusive phenomenon. It is commonly implied that stop-loss mutations are associated to a clinical phenotype through a loss-of-function mechanism; however, our studies reveal the importance to investigate protein aggregation propensity due to cryptic amyloidogenic element in the 3'-UTR as a toxicity inducing mechanism.

Supplemental data

Supplemental data include 5 figures and 2 tables.

Acknowledgments

We deeply appreciate the commitment of the families that participated in this study. This work was supported by NIH (R01NS075764 and U54NS065712 SZ), the CMT Association, the Austrian Science Fund (FWF, P23223-B19 and FWF, P27634FW) and the Muscular Dystrophy Association. We also thank the Inherited Neuropathy Consortium for advice and general support. MMR is grateful to the Medical Research Council (MRC), MRC Centre grant (G0601943), and the National Institutes of Neurological Diseases and Stroke and office of Rare Diseases (U54NS065712) for their support. Part of this work was undertaken at University College London Hospitals/University College London, which received a proportion of funding from the Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme".

Web Resources

The URL for data presented herein are as follows:

OMIM, http://www.omim.org/

Gem.app/GENESIS, http://thegenesisprojectfoundation.org

UTRDb, http://utrdb.ba.itb.cnr.it/

Picard, http://picard.sourceforge.net

Tables

Table 1. Clinical characteristics of studied individuals

Table 2. Electrophysiological characteristics of studied individuals

Figure legends

Figure 1. *NEFH* Frameshift variants in CMT2 families. Asterisks indicate proband individuals. **(A)** Pedigree and Sanger Sequence traces of the CMT family carrying the *NEFH* variant c.3010_3011delGA (p.Asp1004Glnfs*58). Abbreviations: M, mutant allele c.3010_3011delGA; +, wild-type allele **(B)** Pedigree and Sanger Sequence traces of the CMT family carrying the *NEFH* variant c.3017_3020dup (p.Pro1008Alafs*56). Abbreviations: M, mutant allele c.3017_3020dup; +, wild-type allele **(C)** Diagram shows NEFH protein domains and mutations associated with diseases. Coding KSP deletions and insertions from reported ALS cases represented by triangles. CMT frameshift variants from UK1 and AT1 are pointed by the arrows.

Figure 2. Identification of cryptic amyloidogenic elements in the *NEFH* and *NEFL* 3'-UTRs. (A) Protein alignment of WT-NEFH and frameshift variants harbored by the CMT families using Clustal Omega multiple sequence alignment. Translation of the 3'-UTR open-reading frames (ORFs) is illustrated. **(B)** TANGO score of NEHF-3'UTR open-reading frames. **(C)** Consensus sequence of positive residues (asterisk) for all aggregation predictors tested for NEFH-3'UTR ORF3. **(D)** TANGO score of *NEFL*-3'-UTR open-reading frames. **(E)** NEFL-3'-UTR open reading frames sequences. **(F)** Consensus sequence of positive residues (asterisk) for all aggregation predictors tested for NEFL-3'UTR ORF1.

Figure 3. Protein aggregation in cultured Neuro-2a cells expressing neurofilaments encoding 3'UTR-CAEs. (A) Perinuclear aggregates in *GFP-FS-NEFH* transfected cells after 24 hr. Graphs show quantification of number of transfected cells containing NEFH aggregates and quantification of percentage of cells with neuronal projections from 6 independent experiments. **(B)** A small proportion

of cells showing protein aggregation in cells transfected with the truncated construct, *GFP-NEFH*-Stop4 and high levels of aggregation in cells transfected with *GFP-WT-NEFH*-CAE. **(C)** Prominent protein aggregation in cells transfected with *GFP-NEFL*-CAE. Graph shows quantification of number of cells expressing neurofilament-like structures.

Figure 4. Evaluation of cells expressing the NEFH-3'UTR CAE in cultured Neuro-2a cells. (A) Western blot shows comparable level of GFP in cells transfected with *GFP-WT-NEFH* and *GFP-FS-NEFH*. **(B-D)** Sample time-lapse phase contrast and GFP live-cell images obtained from IncuCyte imager system. **(B)** Calculation of the average green (GFP) object area (μM2) per cell at different time points. **(C)** Quantification of GFP positive cells confluence. **(D)** Quantification of average green object eccentricity, measures object roundness from 0 to 1, with a perfect circle having a value of 0. **(E**) Fibers of approximately 10nm in diameter in a neuo-2a cell transfected with *GFP-FS-NEFH*. Filaments are visible in the cytoplasm in longitudinal (blue box) and cross-sections (red box). Approximately 10% of cells transfected with *GFP-FS-NEFH* showed filaments. Scale bars = 200 nm **(F)** Aggregates from NEFH-FS-NEFH cells co-localize with thioflavin T staining.

Figure 5. Confocal microscopy showing NEFH subcellular co-localization. (A) Cells co-transfected with *GFP-WT-NEFH* and NEFL-Myc showing co-localization of NEFH and NEFL in neurofilament network structures. Cells co-transfected with *GFP-FS-NEFH* and NEFL-Myc showing co-localization in the aggregates. **(B)** Co-immunuoprecipitation assay: Cell lysates were immunoprecipitated with an antibody against either GFP or Myc and IgG as a negative control. Both GFP-WT-NEFH and *GFP-FS-NEFH* were co-immunoprecipitated with NEFL-myc. Kinesin was also present in the co-immunoprecipitate. **(C)** Cells transfected with either *GFP-WT-NEFH* or *GFP-FS-NEFH* show normal microtubule network distribution in cells stained with anti-tubulin. **(D)**. Subcellular localization of mitochondria: Cells transfected with *GFP-*

WT-NEFH show even cytoplasmic distribution of mitochondria, whereas, *GFP-FS-NEFH* cells show mitochondria accumulation next to aggregates.

Figure 6. Phenotypic analysis in zebrafish embryos injected with either wild-type or mutant RNAs. (A) Zebrafish embryos injected with RNA encoding either *GFP-WT-NEFH* or *GFP-FS-NEFH* do not show major morphological defects at 48 hpf. **(B)** Motor neurons labeled in the transgenic line, Tg(*Olig2:Dsred*). Embryos injected with *GFP-WT-NEFH* show normal motor neuron development, while of zebrafish injected with *GFP-FS-NEFH* show examples of stunted axons. Scale bar 100 µm **(C)** Quantification of the average axon length shows a decrease in *GFP-FS-NEFH* injected fishes compared to both *GFP-WT-NEFH* injected and uninjected larvae. Axon length is not significantly different between *GFP-WT-NEFH* injected fishes and uninjected control fish. The average axon length per fish was calculated from the first four myotomes. Data was compiled from three independent experiments and significance was determined by a one way ANOVA with Bonferroni post test. P-values are *0.024 and **0.014. **(D)** Confocal images depicting relative GFP fluorescence of the tagged NEFH proteins. Scale bar 100 µm **(E)** Semi-quantitative assessment of relative fluorescence intensity as mean gray values 0-255. **(F)** Western blot showing the presence of both wildtype and mutant NEFH proteins at 24 hpf. *GFP-FS-NEFH* injected larvae were also incubated in 50 µM chloroquine (Chq) in an attempt to increase expression of the mutant protein. **(G)** The presence of both *GFP-WT-NEFH* and *GFP-FS-NEFH* mRNAs were confirmed by rtPCR.

Figure 7: Protein aggregation in cultured Neuro-2a cells expressing FUS encoding the 3'-UTR-CAE. (A) TANGO score of *FUS*-3'-UTR open-reading frame 1. **(B)** Consensus sequence of positive residues (asterisk) for all aggregation predictors tested for *FUS*-3'-UTR ORF1. **(C)** GFP-WT-FUS localizes to the

nucleus, while GFP-FUS-CAE aggregates in the cytoplasm and neuronal-like projections. Magnified view of data (M) is shown in white box.

References

1. Ross, C.A., and Poirier, M.A. (2004). Protein aggregation and neurodegenerative disease. Nature medicine 10 Suppl, S10-17.
2. Aguzzi, A., and O'Connor, T. (2010). Protein aggregation diseases: pathogenicity and therapeutic perspectives. Nature reviews Drug discovery 9, 237-248.
3. Li, S.H., and Li, X.J. (1998). Aggregation of N-terminal huntingtin is dependent on the length of its glutamine repeats. Human molecular genetics 7, 777-782.
4. Gibb, W.R., and Lees, A.J. (1988). The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. Journal of neurology, neurosurgery, and psychiatry 51, 745-752.
5. LaFerla, F.M., and Oddo, S. (2005). Alzheimer's disease: Abeta, tau and synaptic dysfunction. Trends in molecular medicine 11, 170-176.
6. Blokhuis, A.M., Groen, E.J., Koppers, M., van den Berg, L.H., and Pasterkamp, R.J. (2013). Protein aggregation in amyotrophic lateral sclerosis. Acta neuropathologica 125, 777-794.
7. Lin, H., and Schlaepfer, W.W. (2006). Role of neurofilament aggregation in motor neuron disease. Annals of neurology 60, 399-406.
8. Johnson-Kerner, B.L., Ahmad, F.S., Diaz, A.G., Greene, J.P., Gray, S.J., Samulski, R.J., Chung, W.K., Van Coster, R., Maertens, P., Noggle, S.A., et al. (2015). Intermediate filament protein accumulation in motor neurons derived from giant axonal neuropathy iPSCs rescued by restoration of gigaxonin. Human molecular genetics 24, 1420-1431.
9. Llorens, J. (2013). Toxic neurofilamentous axonopathies -- accumulation of neurofilaments and axonal degeneration. Journal of internal medicine 273, 478-489.
10. Hoffman, P.N., Cleveland, D.W., Griffin, J.W., Landes, P.W., Cowan, N.J., and Price, D.L. (1987). Neurofilament gene expression: a major determinant of axonal caliber. Proceedings of the National Academy of Sciences of the United States of America 84, 3472-3476.
11. Carter, J., Gragerov, A., Konvicka, K., Elder, G., Weinstein, H., and Lazzarini, R.A. (1998). Neurofilament (NF) assembly; divergent characteristics of human and rodent NF-L subunits. The Journal of biological chemistry 273, 5101-5108.
12. Figlewicz, D.A., Krizus, A., Martinoli, M.G., Meininger, V., Dib, M., Rouleau, G.A., and Julien, J.P. (1994). Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis. Human molecular genetics 3, 1757-1761.
13. Al-Chalabi, A., Andersen, P.M., Nilsson, P., Chioza, B., Andersson, J.L., Russ, C., Shaw, C.E., Powell, J.F., and Leigh, P.N. (1999). Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. Human molecular genetics 8, 157-164.
14. Berciano, J., Garcia, A., Peeters, K., Gallardo, E., De Vriendt, E., Pelayo-Negro, A.L., Infante, J., and Jordanova, A. (2015). NEFL E396K mutation is associated with a novel dominant intermediate Charcot-Marie-Tooth disease phenotype. Journal of neurology 262, 1289-1300.
15. Brownlees, J., Ackerley, S., Grierson, A.J., Jacobsen, N.J., Shea, K., Anderton, B.H., Leigh, P.N., Shaw, C.E., and Miller, C.C. (2002). Charcot-Marie-Tooth disease neurofilament mutations disrupt neurofilament assembly and axonal transport. Human molecular genetics 11, 2837-2844.
16. Jordanova, A., De Jonghe, P., Boerkoel, C.F., Takashima, H., De Vriendt, E., Ceuterick, C., Martin, J.J., Butler, I.J., Mancias, P., Papasozomenos, S., et al. (2003). Mutations in the neurofilament light chain

gene (NEFL) cause early onset severe Charcot-Marie-Tooth disease. Brain : a journal of neurology 126, 590-597.
17. Yum, S.W., Zhang, J., Mo, K., Li, J., and Scherer, S.S. (2009). A novel recessive Nefl mutation causes a severe, early-onset axonal neuropathy. Annals of neurology 66, 759-770.
18. Zhai, J., Lin, H., Julien, J.P., and Schlaepfer, W.W. (2007). Disruption of neurofilament network with aggregation of light neurofilament protein: a common pathway leading to motor neuron degeneration due to Charcot-Marie-Tooth disease-linked mutations in NFL and HSPB1. Human molecular genetics 16, 3103-3116.
19. Evgrafov, O.V., Mersiyanova, I., Irobi, J., Van Den Bosch, L., Dierick, I., Leung, C.L., Schagina, O., Verpoorten, N., Van Impe, K., Fedotov, V., et al. (2004). Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy. Nature genetics 36, 602-606.
20. Ackerley, S., James, P.A., Kalli, A., French, S., Davies, K.E., and Talbot, K. (2006). A mutation in the small heat-shock protein HSPB1 leading to distal hereditary motor neuronopathy disrupts neurofilament assembly and the axonal transport of specific cellular cargoes. Human molecular genetics 15, 347-354.
21. Link, V., Shevchenko, A., and Heisenberg, C.P. (2006). Proteomics of early zebrafish embryos. BMC developmental biology 6, 1.
22. Gonzalez, M., Falk, M.J., Gai, X., Postrel, R., Schule, R., and Zuchner, S. (2015). Innovative genomic collaboration using the GENESIS (GEM.app) platform. Human mutation 36, 950-956.
23. Gonzalez, M., Falk, M.J., Gai, X., Postrel, R., Schule, R., and Zuchner, S. (2015). Innovative Genomic Collaboration Using the GENESIS (GEM.app) Platform. Human mutation.
24. Gonzalez, M.A., Lebrigio, R.F., Van Booven, D., Ulloa, R.H., Powell, E., Speziani, F., Tekin, M., Schule, R., and Zuchner, S. (2013). GEnomes Management Application (GEM.app): a new software tool for large-scale collaborative genome analysis. Human mutation 34, 842-846.
25. Liu, Q., Xie, F., Siedlak, S.L., Nunomura, A., Honda, K., Moreira, P.I., Zhua, X., Smith, M.A., and Perry, G. (2004). Neurofilament proteins in neurodegenerative diseases. Cellular and molecular life sciences : CMLS 61, 3057-3075.
26. Kovach, M.J., Campbell, K.C., Herman, K., Waggoner, B., Gelber, D., Hughes, L.F., and Kimonis, V.E. (2002). Anticipation in a unique family with Charcot-Marie-Tooth syndrome and deafness: delineation of the clinical features and review of the literature. American journal of medical genetics 108, 295-303.
27. Steiner, I., Gotkine, M., Steiner-Birmanns, B., Biran, I., Silverstein, S., Abeliovich, D., Argov, Z., and Wirguin, I. (2008). Increased severity over generations of Charcot-Marie-Tooth disease type 1A. Journal of neurology 255, 813-819.
28. Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nature biotechnology 22, 1302-1306.
29. de Groot, N.S., Castillo, V., Grana-Montes, R., and Ventura, S. (2012). AGGRESCAN: method, application, and perspectives for drug design. Methods Mol Biol 819, 199-220.
30. Garbuzynskiy, S.O., Lobanov, M.Y., and Galzitskaya, O.V. (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. Bioinformatics 26, 326-332.
31. Walsh, I., Seno, F., Tosatto, S.C., and Trovato, A. (2014). PASTA 2.0: an improved server for protein aggregation prediction. Nucleic acids research 42, W301-307.
32. Bidou, L., Allamand, V., Rousset, J.P., and Namy, O. (2012). Sense from nonsense: therapies for premature stop codon diseases. Trends in molecular medicine 18, 679-688.
33. Groenning, M. (2010). Binding mode of Thioflavin T and other molecular probes in the context of amyloid fibrils-current status. Journal of chemical biology 3, 1-18.

34. Kokubo, Y., Kuzuhara, S., Narita, Y., Kikugawa, K., Nakano, R., Inuzuka, T., Tsuji, S., Watanabe, M., Miyazaki, T., Murayama, S., et al. (1999). Accumulation of neurofilaments and SOD1-immunoreactive products in a patient with familial amyotrophic lateral sclerosis with I113T SOD1 mutation. Archives of neurology 56, 1506-1508.
35. Wagner, O.I., Lifshitz, J., Janmey, P.A., Linden, M., McIntosh, T.K., and Leterrier, J.F. (2003). Mechanisms of mitochondria-neurofilament interactions. The Journal of neuroscience : the official journal of the Society for Neuroscience 23, 9046-9058.
36. Perez-Olle, R., Lopez-Toledano, M.A., Goryunov, D., Cabrera-Poch, N., Stefanis, L., Brown, K., and Liem, R.K. (2005). Mutations in the neurofilament light gene linked to Charcot-Marie-Tooth disease cause defects in transport. Journal of neurochemistry 93, 861-874.
37. Kucenas, S., Snell, H., and Appel, B. (2008). nkx2.2a promotes specification and differentiation of a myelinating subset of oligodendrocyte lineage cells in zebrafish. Neuron glia biology 4, 71-81.
38. Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V.A., Horner, D.S., Pavesi, G., Picardi, E., et al. (2010). UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic acids research 38, D75-80.
39. Hamby, S.E., Thomas, N.S., Cooper, D.N., and Chuzhanova, N. (2011). A meta-analysis of single base-pair substitutions in translational termination codons ('nonstop' mutations) that cause human inherited disease. Human genomics 5, 241-264.
40. Huang, L., Liu, X., Cheng, B., and Huang, K. (2015). How our bodies fight amyloidosis: effects of physiological factors on pathogenic aggregation of amyloidogenic proteins. Archives of biochemistry and biophysics 568, 46-55.
41. Howard, M.T., Shirts, B.H., Petros, L.M., Flanigan, K.M., Gesteland, R.F., and Atkins, J.F. (2000). Sequence specificity of aminoglycoside-induced stop condon readthrough: potential implications for treatment of Duchenne muscular dystrophy. Annals of neurology 48, 164-169.
42. Duquesnoy, B., Grardel, B., Wemeau, J.L., Thevenon, A., Hubaud, P., Proye, C., Lefebvre, J., Decoulx, M., and Delcambre, B. (1988). [Osteomalacia symptoms in primary hyperparathyroidism]. Revue du rhumatisme et des maladies osteo-articulaires 55, 505-510.
43. Skvortsova, V., Shadrina, M., Slominsky, P., Levitsky, G., Kondratieva, E., Zherebtsova, A., Levitskaya, N., Alekhin, A., Serdyuk, A., and Limborska, S. (2004). Analysis of heavy neurofilament subunit gene polymorphism in Russian patients with sporadic motor neuron disease (MND). European journal of human genetics : EJHG 12, 241-244.
44. Wang, L., and Brown, A. (2010). A hereditary spastic paraplegia mutation in kinesin-1A/KIF5A disrupts neurofilament transport. Molecular neurodegeneration 5, 52.

Table 1. Clinical characteristics of studied individuals.

Family no.	**UK1**	**UK1**	**UK1**	**UK1**	**AT1**	**AT1**	**AT1**	**AT1**
Patient	**IV:1**	**III:2**	**II:1**	**I:1**	**III:1**	**III:2**	**III:3**	**III:4**
Sex	male	male	female	male	male	male	male	female
Current age	26 y	57 y	84 y	deceased 80 y	56 y	58 y	62 y	54 y
Age at first symptoms	2-4 y	15 y	early 20s	n/a	38 y	34 y	36-40 y	27 y
First symptoms	frequent falls, tendency to turn over on ankles	shoes with ankle support; able to run until late 40s	unable to stand on tiptoes, flat-footed walk	n/a	gait disturbance, problems with climbing stairs	weakness in lower limbs	pain and weakness in lower limbs	n/a
Age at exam	24 y	57 y	79 y	n/a	49 y	40 y	56 y	34 y
Muscle wasting								
Upper limbs	+	+	+	n/a	N	N	N	n/a
Lower limbs	++	++	+++	n/a	++	+	n/a	n/a
Muscle weakness								
Upper limbs	+	+	++	n/a	N	N	N	++
Lower limbs	+++	+	+++	n/a	++	+	+++	n/a
Pinprick sensation								
Upper limbs	+	+	N	n/a	N	N	N	n/a
Lower limbs	++	++	++	n/a	+	N	N	n/a
Vibration sense								
Upper limbs	N	N	++	n/a	n/a	N	n/a	n/a
Lower limbs	++	++	+++	n/a	+	++	++	n/a
Joint position sense								
Upper limbs	N	N	N	n/a	N	N	N	n/a
Lower limbs	N	N	N	n/a	N	N	N	n/a
Reflexes								
Upper limbs	+	+	abs	n/a	↓	abs	n/a	n/a
Lower limbs	abs	abs (a)	abs	n/a	↓/abs	↓/abs	abs	n/a
Plantar responses								
Right/left	f/f	f/f	f/–	n/a	f/f	n/a	n/a	n/a
Walk with support	yes (20s)	yes (50s)	yes (50s)	yes (50s)	no	no	no	yes
Wheelchair-bound	no	no	yes (70s)	yes (70s)	no	no	no	n/a

Other features	increased tone in lower limbs; waddling and stiff gait[1]	cramps in hands and lower limbs	hearing loss (70s)	hearing loss (50s)		cramps in lower limbs	mild weakness in hands	initial diagnosis of myositis
Creatine kinase	1288[2]	442[2]	150[3]	n/a	686	505	351	920
Brain/spinal cord MRI	N	N	n/a	n/a	n/a	n/a	n/a	n/a

Muscle tone: N: normal; +: increased. **Muscle wasting**: N: no muscle wasting; +: below wrist/ankle; ++: below elbow/knee; +++: above elbow/knee. **Muscle weakness**: N: normal; +: >4 distal muscle groups (FDIO, APB, ankle plantar flexion or extension or below); ++: <4 distal muscles; +++: proximal weakness (knee flexion/extension, elbow flexion/extension or above). **Plantar responses**: f: flexor; –: mute; e: extensor. **Reflexes**: ↓ diminished: d; +: present; ++: brisk; +++: brisk with extensor plantars; ±: present with reinforcement; abs: absent; abs (a): absent ankle jerks only. **Sensory examination**: N: normal; +: reduced below wrist/ankle; ++: reduced below elbow/knee; +++: reduced at or above elbow/knee.

N = normal; n/a = not available; 1 = on subsequent assessments; 2 = reference range 38-204 IU/L; 3 = reference range 26-140 IU/L.

Table 2. Electrophysiological characteristics of studied individuals.

Family no.	**UK1**		**UK1**		**UK1**		**AT1**		**AT1**	
Patient	**IV:1**		**III:2**		**II:1**		**III:1**		**III:2**	
Age at examination	24 y		55 y		72 y		49 y		n/a	
Nerve conduction studies	R	L	R	L	R	L	R	L	R	L
Radial nerve										
Sensory Amp	20 µv		**11 µv**		**abs**		n/a		n/a	
Sensory CV	56 m/s		60 m/s		–		n/a		n/a	
Median nerve										
DML	4.1 ms		4.0 ms		**4.8 ms**		3.6 ms		n/a	
Motor Amp	10.7 mV		6.7 mV		**1.7 mV**		9.8 mV		n/a	
Motor CV	**46 m/s**		54 m/s		**38 m/s**		59.2 m/s		n/a	
F wave latency	**34.1 ms**		**33.9 ms**		22.4 ms		n/a		n/a	
Sensory Amp	**6 µv**	**3 µv**	**3 µv**		**abs**		**9.5 µv**		n/a	
Sensory CV	56 m/s	59 m/s	51 m/s		–		50.9		n/a	
Ulnar nerve										
DML	3.2 ms	3.2 ms	2.8 ms		**3.7 ms**		2.6 ms		n/a	
Motor Amp	10.8 mV	10.4 mV	12.1 mV		7.7 mV		20.6 mV		n/a	
Motor CV	55 m/s	54 m/s	51 m/s		56 m/s		48.7 m/s		n/a	
F wave latency	31.0 ms		32.2 ms		**34.2 ms**		n/a		n/a	
Sensory Amp	**3 µv**		**1 µv**		**abs**		10.1 µv		n/a	
Sensory CV	52 m/s		52 m/s		–		52.1 m/s		n/a	
Peroneal nerve										
DML	**7.5 ms**		6.0 ms		–		4.0 ms	4.4 ms	n/a	
Motor Amp	**1.9 mV**		**0.5 mV**		**abs**		**4.2 mV**	**1.8 mV**	n/a	
Motor CV	**35 m/s**[1]		**41 m/s**		–		**43.2 m/s**	**32.3 m/s**	n/a	
F wave latency	n/a		n/a		–		n/a		n/a	
Sensory Amp	**abs**		**abs**[2]		**abs**		n/a		n/a	
Sensory CV	–		–		–		n/a		n/a	
Tibial nerve										
DML	**7.3 ms**		n/a		–		3.5 ms	4.3 ms	n/a	
Motor Amp	**1.2 mV**		n/a		**abs**		**3.0 mV**	**2.7 mV**	n/a	