

## ARTICLE


<https://doi.org/10.1038/s41467-022-30015-1>

OPEN

# Heterozygous frameshift variants in *HNRNPA2B1* cause early-onset oculopharyngeal muscular dystrophy

Hong Joo Kim <sup>1,39</sup>, Payam Mohassel <sup>2,39</sup>, Sandra Donkervoort<sup>2</sup>, Lin Guo<sup>3,4</sup>, Kevin O'Donovan<sup>1</sup>, Maura Coughlin<sup>1</sup>, Xaviere Lornage<sup>5</sup>, Nicola Foulds<sup>6</sup>, Simon R. Hammans<sup>7</sup>, A. Reghan Foley<sup>2</sup>, Charlotte M. Fare <sup>3</sup>, Alice F. Ford<sup>3</sup>, Masashi Ogasawara<sup>8,9</sup>, Aki Sato<sup>10</sup>, Aritoshi Iida <sup>9</sup>, Pinki Munot<sup>11</sup>, Gautam Ambegaonkar<sup>12</sup>, Rahul Phadke<sup>13</sup>, Dominic G. O'Donovan<sup>14</sup>, Rebecca Buchert <sup>15</sup>, Mona Grimm<sup>15</sup>, Ana Töpfl<sup>16</sup>, Irina T. Zaharieva<sup>11</sup>, Lauren Brady<sup>17</sup>, Ying Hu<sup>2</sup>, Thomas E. Lloyd <sup>18</sup>, Andrea Klein<sup>19,20</sup>, Maja Steinlin<sup>19</sup>, Alice Kuster<sup>21</sup>, Sandra Mercier<sup>22,23</sup>, Pascale Marcorelles<sup>24</sup>, Yann Péréon<sup>25</sup>, Emmanuelle Fleurence <sup>26</sup>, Adnan Manzur<sup>11</sup>, Sarah Ennis <sup>27</sup>, Rosanna Upstill-Goddard <sup>27</sup>, Luca Bello <sup>28</sup>, Cinzia Bertolin<sup>29</sup>, Elena Pegoraro<sup>28</sup>, Leonardo Salvati <sup>30</sup>, Courtney E. French <sup>31</sup>, Andriy Shatillo <sup>32</sup>, F. Lucy Raymond <sup>33</sup>, Tobias B. Haack<sup>15</sup>, Susana Quijano-Roy<sup>34</sup>, Johann Böhm<sup>5</sup>, Isabelle Nelson<sup>35</sup>, Tanya Stojkovic<sup>36</sup>, Teresinha Evangelista<sup>37</sup>, Volker Straub <sup>16</sup>, Norma B. Romero<sup>36,37</sup>, Jocelyn Laporte <sup>5</sup>, Francesco Muntoni<sup>11</sup>, Ichizo Nishino <sup>8,9</sup>, Mark A. Tarnopolsky <sup>17</sup>, James Shorter<sup>3</sup>, Carsten G. Bönnemann <sup>2</sup>✉ & J. Paul Taylor <sup>1,38</sup>✉

Missense variants in RNA-binding proteins (RBPs) underlie a spectrum of disease phenotypes, including amyotrophic lateral sclerosis, frontotemporal dementia, and inclusion body myopathy. Here, we present ten independent families with a severe, progressive muscular dystrophy, reminiscent of oculopharyngeal muscular dystrophy (OPMD) but of much earlier onset, caused by heterozygous frameshift variants in the RBP hnRNPA2/B1. All disease-causing frameshift mutations abolish the native stop codon and extend the reading frame, creating novel transcripts that escape nonsense-mediated decay and are translated to produce hnRNPA2/B1 protein with the same neomorphic C-terminal sequence. In contrast to previously reported disease-causing missense variants in *HNRNPA2B1*, these frameshift variants do not increase the propensity of hnRNPA2 protein to fibrillize. Rather, the frameshift variants have reduced affinity for the nuclear import receptor karyopherin  $\beta 2$ , resulting in cytoplasmic accumulation of hnRNPA2 protein in cells and in animal models that recapitulate the human pathology. Thus, we expand the phenotypes associated with *HNRNPA2B1* to include an early-onset form of OPMD caused by frameshift variants that alter its nucleocytoplasmic transport dynamics.

A full list of author affiliations appears at the end of the paper.

**R**NA-binding proteins (RBPs) play central roles in regulating RNA metabolism, including transcription, RNA splicing, polyadenylation, stabilization, localization, and translation<sup>1</sup>. In eukaryotic cells, RBPs are frequently associated with RNA in ribonucleoprotein (RNP) granules, which are complex assemblies that arise via liquid–liquid phase separation (LLPS)<sup>2</sup>. Many RBPs contain intrinsically disordered low-complexity domains (LCDs) that support LLPS through weak, multivalent interactions. The ability of RBPs to facilitate LLPS and partition RNAs into condensates is an important mechanism by which they govern RNA metabolism<sup>3</sup>.

Pathogenic missense variants in RBPs such as TDP-43<sup>4,5</sup>, hnRNPA1<sup>6–9</sup>, hnRNPA2/B1<sup>6,10</sup>, hnRNPD1<sup>11–14</sup>, FUS<sup>15,16</sup>, and TIA1<sup>17,18</sup> cause a spectrum of diseases with pleomorphic phenotypic manifestations including amyotrophic lateral sclerosis (ALS)/motor neuron disease, frontotemporal dementia (FTD), inclusion body myopathy (IBM), distal myopathy, and Paget's disease of bone (PDB). Patients may also present with complex, combined phenotypes impacting muscle, brain, and bone, a syndrome that has been termed multisystem proteinopathy (MSP)<sup>19</sup>. Remarkably, most pathogenic variants in RBPs are located in intrinsically disordered LCDs<sup>4–6,15–18</sup>. Studies using purified proteins and cellular expression models have demonstrated that LCD variants can alter LLPS and ultimately result in aggregation and fibrillization of the mutant protein<sup>6,9</sup>. We have proposed that pathological phase separations of this sort can drive neuronal dysfunction or demise either by altering the function of biomolecular condensates, by producing toxic, aggregated protein, or both<sup>20,21</sup>.

We previously reported a family with the MSP phenotype manifesting as degeneration of muscle, bone, or brain, either alone or in combination in different family members, caused by a rare missense mutation in *HNRNPA2B1* [NM\_002137, p.(D290V)]<sup>6</sup>. This mutation, which is within the LCD of the protein, promotes assembly of hnRNPA2 into self-seeding fibrils by introducing a potent steric-zipper motif, thereby dysregulating polymerization and ultimately driving the formation of cytoplasmic inclusions<sup>6,22</sup>. A second heterozygous missense mutation in *HNRNPA2B1* [NM\_002137, p.(P298L)], also within the LCD and predicted to promote aggregation, was found to cause pure PDB without additional multisystemic features<sup>10</sup>. MSP-causing heterozygous missense mutations have been identified in the LCD of two additional RBPs, hnRNPA1<sup>6</sup> and TIA1<sup>18</sup>. Moreover, ALS/FTD-causing mutations in the RBP TDP-43 most often impact the LCD<sup>23</sup>, and ALS-causing mutations of the RBP FUS sometimes impact the LCD<sup>24</sup>. Similar to the observed consequences of missense mutations to the LCD of hnRNPA2/B1, many of the disease-causing mutations in the LCDs of hnRNPA1, hnRNPD1, TIA1, TDP-43, and FUS have been shown or are predicted to promote assembly of self-seeding fibrils, ultimately driving the formation of cytoplasmic inclusions<sup>6,18,25–28</sup>.

Modest expansion of a polyalanine repeat tract in *PABPN1*, which also encodes an RBP, underlies the majority of cases of oculopharyngeal muscular dystrophy (OPMD), a late-onset (typically 5th decade) disease characterized by ptosis and dysphagia, and in rare cases progressive limb weakness and later stage vertical ophthalmoparesis<sup>29</sup>. OPMD is clinically distinct from the pleomorphic syndrome MSP, and, unlike other RBP-associated diseases, has a homogeneous and recognizable phenotypic spectrum that predominantly manifests in specific skeletal muscles.

Here, we report and characterize 10 independent families with a severe, progressive, early-onset OPMD-like phenotype (eoOPMD) that is distinct from MSP and is caused by a novel class of heterozygous frameshift variants in the LCD of hnRNPA2/B1. In contrast to the missense variants that cause MSP or PDB, these frameshift variants do not promote

protein fibrillization. Rather, these frameshift variants alter the C-terminal portion of the nuclear localization sequence (NLS) of hnRNPA2 and promote cytoplasmic accumulation of the protein by impairing its interaction with the nuclear import receptor karyopherin  $\beta 2$  (Kap $\beta 2$ ). Furthermore, our findings demonstrate that two classes of disease-causing mutation have distinct consequences for the hnRNPA2 protein: the first (missense mutations in the LCD) increases the propensity toward fibrillization, whereas the second (frameshift mutations that partially impair nuclear import) increases the cytoplasmic concentration of the protein without changing its intrinsic propensity to fibrillize. Both classes of mutation culminate in muscle pathology, albeit with different clinical syndromes characterized by differing ages of onset and anatomic distribution of pathology.

## Results

***HNRNPA2B1* frameshift variants manifest with a distinct, early-onset progressive myopathy.** We identified 11 patients from 10 independent families with progressive early-onset myopathy with ophthalmoplegia, ptosis, and respiratory insufficiency of variable degrees (Table 1, Fig. 1a). Six patients presented with first recognition of symptoms before 2 years of age, ranging from respiratory insufficiency requiring tracheostomy at birth, to delayed motor milestones or isolated ptosis and ophthalmoplegia. All patients had axial weakness and progressive proximal and distal weakness, which was more pronounced in the lower extremities compared to the upper extremities (Fig. 1b). Respiratory involvement was variable, ranging from a severely decreased forced vital capacity (FVC) (21% predicted) to normal. Ptosis and progressive ophthalmoparesis were uniformly present in all patients and were noted as early as 6 months of age. Cognition was normal and patients did not have a history of seizures, cardiac involvement, or bone abnormalities. Serum creatine kinase levels were elevated in all patients, ranging from 3 to 35 times the upper limit of normal (Table 1). Detailed clinical and genetic information for each patient can be found in Supplementary Table 1 and Supplementary Information.

To elucidate the genetic origin of the muscle disease, we pursued exome sequencing and identified nine (one recurrent) distinct heterozygous frameshift variants in *HNRNPA2B1* (MIM:600124), which were absent in the Genome Aggregation Database (gnomAD)<sup>30</sup> (Fig. 1c and Supplementary Table 1). In family 4, the disease was inherited in a dominant fashion and the variant segregated with disease. The disease occurred sporadically in the remaining nine families (Fig. 1a) and the *HNRNPA2B1* variants were confirmed to be *de novo* (i.e., absent in unaffected parents) in seven families for which parental samples were available for testing.

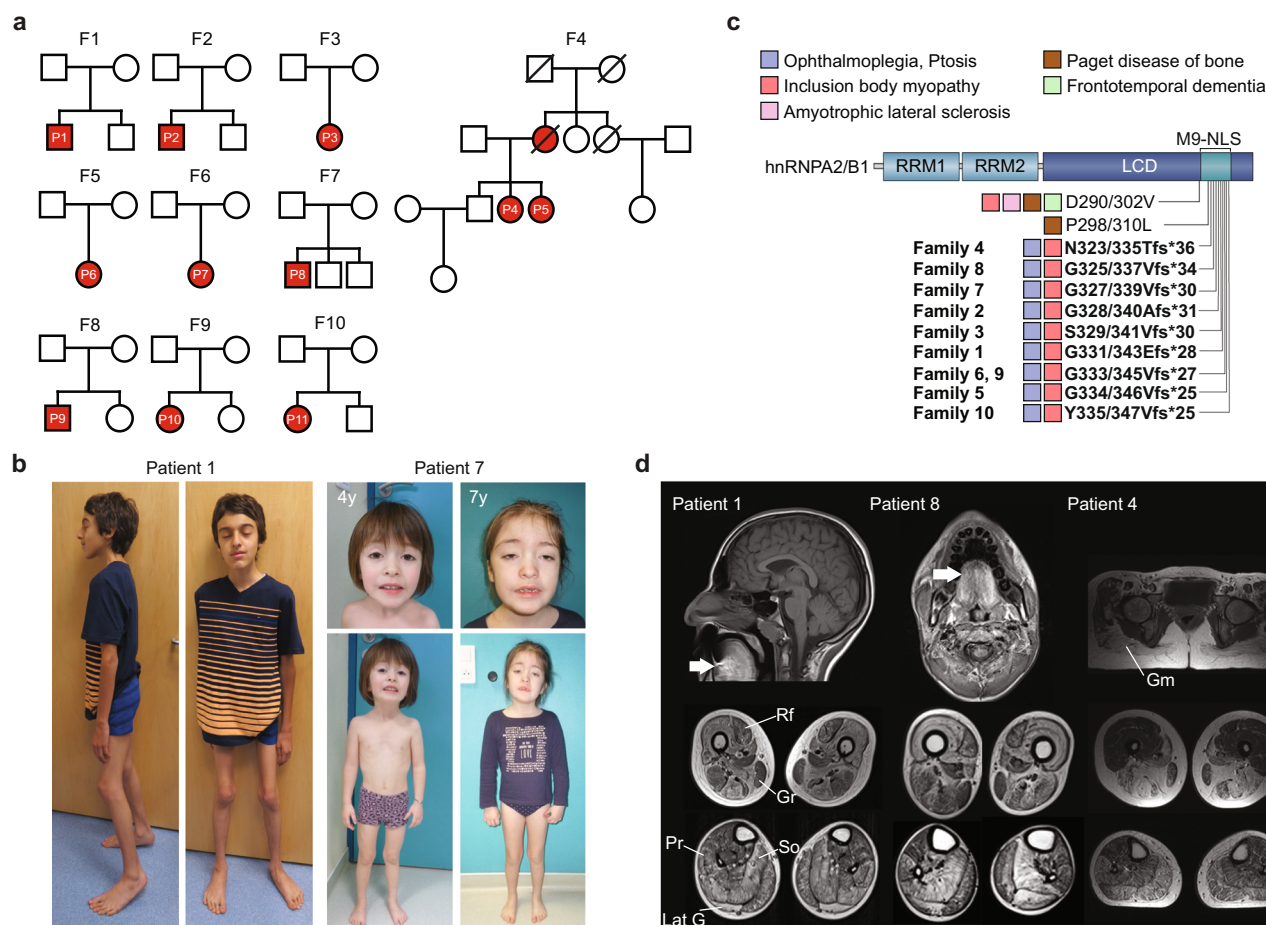
**Muscle imaging reveals myopathic changes with discrete foci of increased T1 signal.** Muscle MRI images of lower extremities showed muscle atrophy and a heterogeneous pattern of T1 signal hyperintensity, suggestive of structural changes in the muscle composition (Fig. 1d). Presence of patchy foci of T1 signal hyperintensity, suggestive of focal fatty replacement of muscle, was noted in mildly affected muscles, while more severely affected muscles showed a diffuse pattern of T1 hyperintensity. In the anterior thigh compartment, the rectus femoris muscle was relatively spared. Medial and posterior thigh compartments were selectively affected, albeit not uniformly, with the gracilis muscle appearing relatively spared. In the lower legs, the peroneus group, soleus, and lateral gastrocnemius muscles were selectively affected; however, the appearance of this pattern was variable among patients. In the head and neck MRI images, T1 hyperintensity in the tongue was notable (Fig. 1d, arrows).

**Table 1 Clinical features of patients with pathogenic variants in *HNRNPA2B1*.**

Family/ Patient	Ethnicity	Sex/age at last examination (y)	Pathogenic <i>HNRNPA2B1</i> variant <sup>a</sup>	Inheritance	Onset	Pattern of muscle weakness	CK level U/L
F1/P1	Italian/ Colombian	M/12	c.992delG, p.(G331Efs*28)	de novo	Swallowing difficulties (2 y)	Ptosis, ophthalmoplegia, tongue weakness, dysphonia, symmetric proximal and distal weakness in LE > UE; moderate progression, dysphagia, respiratory insufficiency	724
F2/P2	Northern European	M/17	c.981delA, p.(G328Afs*31)	de novo	Neonatal feeding difficulties	Ptosis, ophthalmoplegia, symmetric proximal and distal weakness in LE > UE; moderate progression, dysphagia	602
F3/P3	Ukrainian	F/17	c.984delC, p.(S329Vfs*30)	de novo	Muscle weakness (6 y)	Ptosis, ophthalmoplegia, dysphonia, symmetric proximal and distal weakness in LE > UE; dysphagia, moderate progression	994
F4/P4	English	F/40	c.966delA, p.(N323Tfs*36)	Dominant	Ptosis, ophthalmoplegia (18 y)	Ptosis, ophthalmoplegia, symmetric proximal and distal weakness in LE > UE; moderate progression, dysphagia, respiratory insufficiency	808
F4/P5	English	F/43	c.966delA, p.(N323Tfs*36)	Dominant	Ptosis, ophthalmoplegia, dysphagia (17 y)	Ptosis, ophthalmoplegia, symmetric proximal and distal weakness in LE > UE; moderate progression, dysphagia, respiratory insufficiency	1105
F5/P6	English	F/12	c.1001delG, p.(G334Vfs*25)	Presumed de novo	Respiratory insufficiency requiring tracheostomy (birth)	Ptosis, ophthalmoplegia, asymmetric proximal and distal weakness LE > UE; mild progression, dysphagia, respiratory insufficiency	770
F6/P7	French	F/9	c.996_997dupTG, p.(G333Vfs*27)	de novo	Ptosis, ophthalmoplegia (6 months)	Ptosis, ophthalmoplegia, dysphonia, axial weakness; respiratory insufficiency, dysphagia, loss of independent ambulation (8 years)	536
F7/P8	French	M/20	c.980_986del, p.(G327Vfs*30)	de novo	Delayed motor milestones (infancy)	Ptosis, ophthalmoplegia, dysphonia, axial weakness, symmetric proximal UE and LE weakness, respiratory insufficiency	2484
F8/P9	Italian	M/35	c.974delG, p.(G325Vfs*34)	de novo	Slower than peers (5 y)	Ptosis, ophthalmoplegia, tongue weakness, symmetric proximal and distal weakness LE > UE; moderate/severe progression, respiratory insufficiency, dysphagia	7067
F9/P10	Indian	F/7	c.996_997dupTG, p.(G333Vfs*27)	de novo	Respiratory illness with acute onset ptosis (18 months)	Ptosis, ophthalmoplegia, dysphonia, symmetric LE > UE proximal, distal and axial weakness; loss of independent ambulation (7 years), ankle contractures, dysphagia, respiratory insufficiency	424
F10/P11	Japanese	F/18	c.1001_1002dupGT, p.(Y335Vfs*25)	Presumed de novo	Slow runner (childhood)	Ptosis, ophthalmoplegia, symmetric LE > UE proximal muscle weakness, respiratory insufficiency	501

M male, F female, y years, CK creatine kinase, LE lower extremity, UE upper extremity.

<sup>a</sup>Transcript ID: NM\_002137.



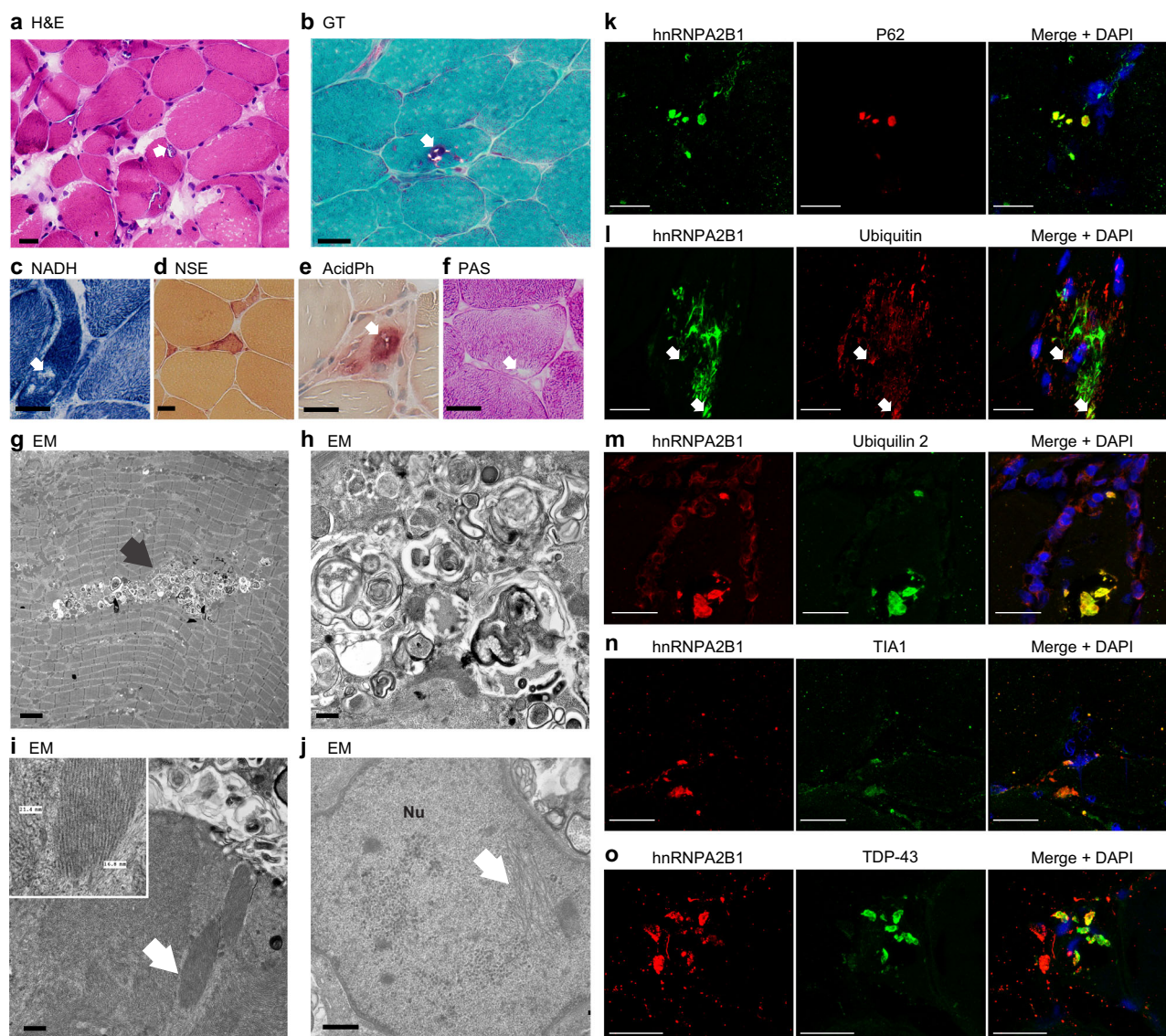
**Fig. 1 Clinical characteristics and genetic findings in ten independent families with *HNRNPA2B1* variants.** **a** Pedigrees of the 10 families. F family, P patient. Filled red indicates affected individuals with eOPMD phenotype and heterozygous *HNRNPA2B1* frameshift variant. **b** Patient 1 with c.992delG, p.(G331Efs\*28) variant, and patient 7 with c.996\_997dupTG, p.(G333Vfs\*27) showing prominent ptosis and muscle atrophy. Note the progression of ptosis in patient 7 from 4 years of age to 7 years of age. **c** hnRNPA2/B1 domain structure with conserved regions. Variants identified from previous studies and in this study (in bold) are indicated in both hnRNPA2 and hnRNPA1 isoforms with their associated phenotypic features. RRM RNA-recognition motif, LCD low complexity domain, M9-NLS M9-nuclear localization signal. **d** T1 MRI images of the head and lower extremities. Head and neck MRI highlights T1 hyperintensity in the tongue (arrows indicate "bright tongue sign"). In the hips, gluteus maximus (Gm) is selectively affected. In the lower extremities, there is selective involvement of the posterior and medial thigh compartment with relative sparing of the rectus femoris (Rf) and gracilis (Gr) muscle. Lower leg images show relative sparing of the tibialis anterior and selective involvement of the peroneus group (Pr), soleus (So), and lateral gastrocnemius (Lat G).

***HNRNPA2B1* frameshift variants manifest pathologically with rimmed vacuoles and intracytoplasmic aggregates.** Muscle biopsies of the patients were consistent with a chronic degenerative myopathy, characterized by myofiber atrophy, fiber size variability, increased internalized nuclei, and presence of subsarcolemmal and cytoplasmic rimmed vacuoles without inflammatory infiltrates (Fig. 2a, b). Light microscopy revealed that the vacuoles contained proteinaceous material (Fig. 2c–f), and ultrastructural analysis confirmed the presence of membrane-bound vacuoles containing debris in a "myelin-like" configuration often attributed to excess autophagy (Fig. 2g, h). Cytoplasmic, perinuclear, and on occasion intranuclear tubulofilamentous inclusions with ~15–20 nm thickness were also identified, indicating the accumulation of microfilamentous protein aggregates (Fig. 2i, j, arrows and inset). Immunofluorescence staining of muscle tissue highlighted scattered myofibers with hnRNPA2/B1-positive inclusions that partially colocalized with P62, ubiquitin 2, TIA1, and TDP-43, and to a lesser degree with ubiquitin (Fig. 2k–o), similar to the end-stage fibrillar accumulation of RBPs in microfilamentous structures observed in related disorders<sup>6</sup>. These inclusions did not stain positive for Z-disk-associated

proteins that typically aggregate in myofibrillar myopathies (Supplementary Fig. 1) and were unvested within membrane layers.

**Frameshift variants cluster in the hnRNPA2B1 LCD and escape RNA quality control degradation.** *HNRNPA2B1* is expressed as two alternatively spliced isoforms, *HNRNPA2* (NM\_002137, 11 exons) and *HNRNPB1* (NM\_031243, 12 exons). The shorter isoform, *HNRNPA2*, which lacks an exon (exon 2 of *HNRNPB1*) and its associated 12 amino acids in the N-terminal region, is the main isoform, accounting for 90% of the protein in most tissues. All of the frameshift variants we identified in our cohort cluster in exon 10 of *HNRNPA2* and exon 11 of *HNRNPB1*, which encode the highly conserved LCD and are respectively the last coding exon in each isoform. These frameshifts all abolish the native stop codon and extend the reading frame (Fig. 1c). Irrespective of the point at which each frameshift occurred, all mutations result in the same frameshift with a common C-terminal sequence (VMVGGADTELLPICHGLHCINRRG) (Supplementary Fig. 2a, b). RT-PCR analysis in muscle tissue from patients 1 and 2 suggested that these frameshift



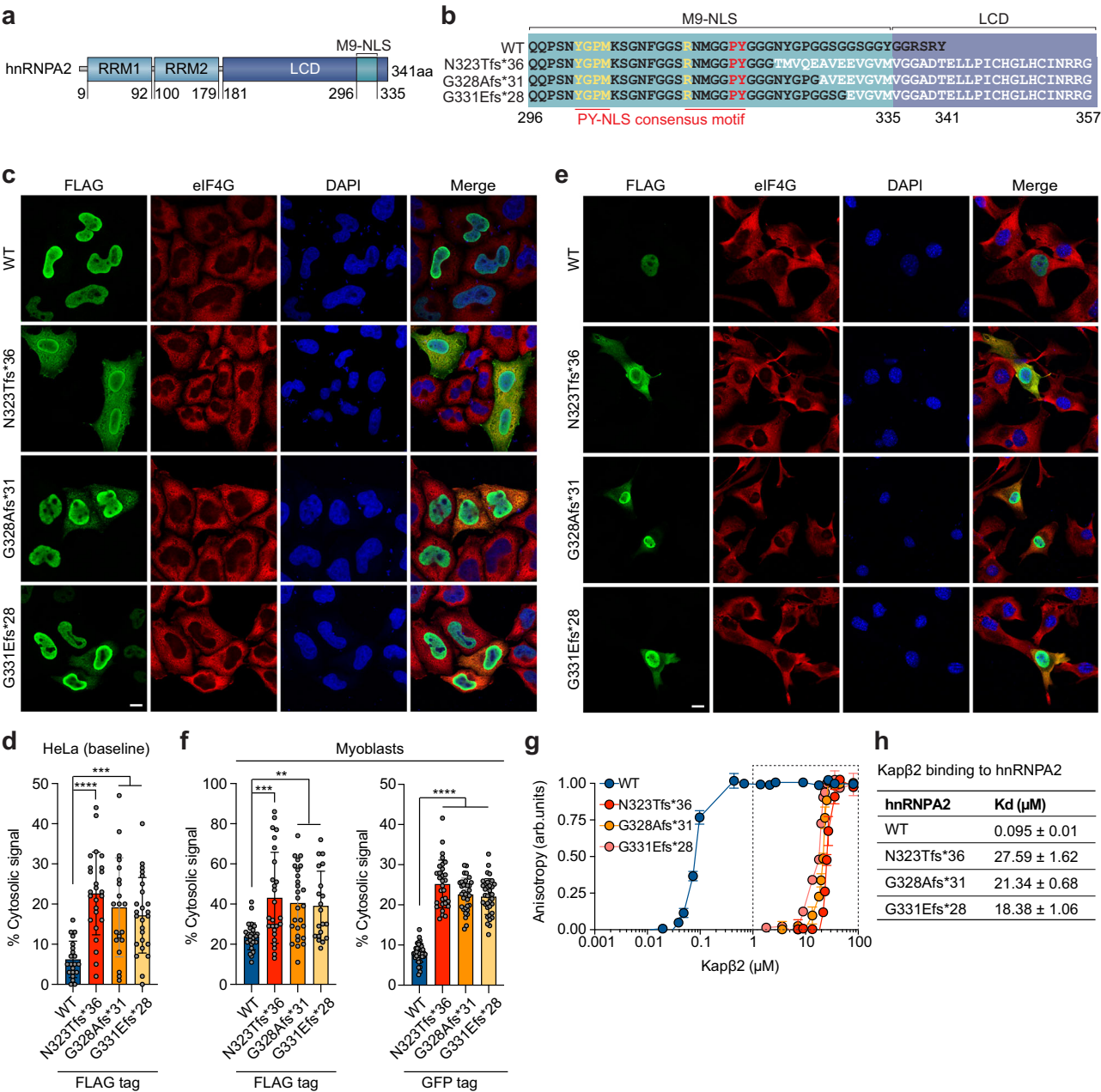


**Fig. 2 Histological findings of hnRNPA2B1 myopathy.** **a** Hematoxylin and eosin (H&E) image of right rectus femoris muscle biopsy obtained at 8 years of age from patient 1 with c.992delG, p.(G331Efs\*28) showing myofiber atrophy, fiber size variability, and internalized nuclei with several fibers with rimmed subsarcolemmal vacuoles (arrow). **b** Modified Gömöri trichrome (GT) staining highlights similar findings and rimmed vacuoles (arrow) (patient 1 muscle biopsy). **c–f** Vacuole contents (arrow) do not stain positive with NADH, suggesting they are devoid of mitochondria (**c**) and do not show increased non-specific esterase (NSE) activity (**d**). The vacuoles (arrow) do not contain glycoproteins based on periodic acid-Schiff (PAS) stain (**f**) but have increased acid phosphatase (AcidPh) activity (arrow) (**e**), suggesting they are lysosomal or autophagic in origin. NSE stain (**d**) also highlights a few angular atrophic fibers, suggestive of mild acute neurogenic atrophy. **c–f** are from patient 2 muscle biopsy. **g, h** Electron microscopy (EM) studies showing marked autophagic changes (black arrow) and vacuoles containing membranous myelin-like whorls. **i, j** Many myofibers contain areas with ~15–20-nm thick, tubulofilamentous inclusions (**i**, white arrow and inset), which on occasion were also seen within the nuclei (Nu) near the vacuoles (**j**, white arrow). **g–i** are from patient 2 muscle biopsy, **j** is from patient 8 muscle biopsy. **k–o** Immunofluorescence staining of muscle biopsy of patient 1 with c.992delG, p.(G331Efs\*28) showing cytoplasmic hnRNPA2/B1-positive inclusions that partially co-localize with P62 (**k**), ubiquilin 2 (**m**), TIA1 (**n**), and TDP-43 (**o**). Co-localization with ubiquitin (**l**) is restricted to very few perinuclear and cytoplasmic puncta (arrows). Scale bars: **a–f** 25  $\mu$ m; **g** 2  $\mu$ m; **h–j** 500 nm; **k–o** 20  $\mu$ m. The micrographs shown are representative images of a single diagnostic muscle biopsy in each indicated patient. No independent replicates were performed.

variants escape degradation by the RNA quality control system (Supplementary Fig. 2c).

**Frameshift variants alter the nucleocytoplasmic ratio of hnRNPA2 and enhance its recruitment to RNP granules.** Nucleocytoplasmic shuttling of hnRNPA2 is regulated by its 40-amino acid M9 sequence located within the C-terminal LCD (Fig. 3a, b). The M9 sequence serves as both an NLS and a nuclear

export signal (NES) and is recognized by the nuclear transport receptor Kap $\beta$ 2 (also known as transportin 1, TNPO1)<sup>31–33</sup>. M9-NLSs are structurally disordered and have an overall basic character with weakly conserved sequence motifs composed of an N-terminal hydrophobic or basic motif followed by a C-terminal R/H/KX<sub>2–5</sub>PY consensus sequence and its flanking region<sup>34</sup> (Fig. 3b). Due to the strict conservation of proline-tyrosine (PY) residues and the importance of these residues in the binding of Kap $\beta$ 2, M9-NLSs are now often defined as PY-NLSs, a more



**Fig. 3 Frameshift mutations impair nucleocytoplasmic trafficking of hnRNP A2 by disrupting interaction between hnRNP A2 and Kapβ2.** **a** Domain architecture of hnRNP A2 illustrating RNA recognition motifs 1 and 2 (RRM1 and RRM2), a low complexity domain (LCD), and an M9 nuclear localization signal (M9-NLS) within the LCD. **b** Amino acid sequences of relevant domains in WT hnRNP A2 and frameshift mutants. Consensus PY-NLS motifs within the M9-NLS are underlined in red. **c** Intracellular localization of indicated FLAG-tagged hnRNP A2 proteins under basal conditions in HeLa cells. eIF4G was used as a cytoplasmic and stress granule marker. Scale bar, 10 μm. **d** Quantification of hnRNP A2 cytosolic signal intensity in HeLa cells as shown in **(c)**. An interleaved scatter plot with individual data points is shown; error bars represent mean ± s.d. For WT, N323Tfs\*36, G328Afs\*31, and G331Efs\*28,  $n = 21$ , 25, 20, and 25 cells, respectively. \*\*\* $P = 0.0001$  (WT vs. G328Afs\*31), \*\*\* $P = 0.0006$  (WT vs. G331Efs\*28), and \*\*\*\* $P < 0.0001$  by one-way ANOVA with Dunnett's multiple comparisons test. **e** Intracellular localization of indicated FLAG-tagged hnRNP A2 proteins in C2C12 myoblasts. eIF4G was used as a cytoplasmic and stress granule marker. Scale bar, 10 μm. **f** Quantification of FLAG-tagged or GFP-tagged hnRNP A2 cytosolic signal intensity in C2C12 myoblasts. An interleaved scatter plot with individual data points is shown; error bars represent mean ± s.d. For FLAG-hnRNP A2,  $n = 26$ , 26, 27, and 20 cells for WT, N323Tfs\*36, G328Afs\*31, and G331Efs\*28, respectively. For GFP-hnRNP A2,  $n = 36$ , 30, 33, and 35 cells for WT, N323Tfs\*36, G328Afs\*31, and G331Efs\*28, respectively. \*\*\*\* $P < 0.0001$ , \*\*\* $P = 0.0003$  (WT vs. N323Tfs\*36), \*\* $P = 0.0017$  (WT vs. G328Afs\*31), and \*\* $P = 0.0094$  (WT vs. G331Efs\*28) by one-way ANOVA with Dunnett's multiple comparisons test. **g** Fluorescence polarization measurements between TAMRA-labeled M9-NLS of hnRNP A2 WT, N323Tfs\*36, G328Afs\*31, or G331Efs\*28 peptide (100 nM) with increasing concentrations of Kapβ2. Peptide sequences are shown in **(b)**. Values represent means ± s.e.m. ( $n = 3$  independent experiments). **h**,  $K_d$  values between TAMRA-labeled M9-NLS of hnRNP A2 peptides and Kapβ2 WT. Values were calculated as detailed in Methods and represent means ± s.d. from three independent experiments.



minimal class of NLS that ends on the PY motif<sup>34</sup> (Fig. 3b). Because hnRNPA1 is a much less abundant isoform that uses precisely the same mechanism of nuclear import and is expected to be impacted in the same way as hnRNPA2, it will not be discussed further.

All nine *HNRNPA2B1* frameshift variants identified in our cohort altered the M9-NLS amino acid sequence by shifting the reading frame by one base pair (Supplementary Fig. 2a, b), suggesting that nucleocytoplasmic trafficking might be impaired by these variants. To examine the effect of these C-terminal frameshift variants on nucleocytoplasmic trafficking of hnRNPA2, we expressed FLAG-tagged versions of wild-type (WT) and three different mutant hnRNPA2 proteins (N323Tfs\*36, G328Afs\*31, and G331Efs\*28) in HeLa cells. Whereas hnRNPA2 WT almost exclusively localized to nuclei, N323Tfs\*36, G328Afs\*31, and G331Efs\*28 mutants showed cytoplasmic accumulation at baseline, suggesting impairment of nucleocytoplasmic transport (Fig. 3c, d). The cytoplasmic accumulation of mutant proteins became more apparent when cells were subjected to oxidative stress, which induces assembly of stress granules (Supplementary Fig. 3a, b). We also evaluated the impact of the variants on hnRNPA2 localization in a disease-relevant cell type and found that, similar to HeLa cells, C2C12 myoblast cells (Fig. 3e, f) and differentiated myotubes (Supplementary Fig. 3c) expressing either FLAG-tagged or GFP-tagged hnRNPA2 showed increased cytoplasmic localization of mutant hnRNPA2 proteins compared to WT.

To more clearly demonstrate the distribution of hnRNPA2 variants in cells, we next measured the signal intensities of WT, the MSP-associated D290V mutant, and the eoOPMD-associated N323Tfs\*36 mutant hnRNPA2 in the nucleus and cytoplasm. Line scan intensity graphs indicated that both WT and D290V predominantly localized to the nucleus before stress (Supplementary Fig. 4a). Upon oxidative stress (500  $\mu$ M sodium arsenite), D290V, but not WT, showed association with stress granules, as demonstrated by colocalization with eIF4G punctae (Supplementary Fig. 4b). However, the majority of the hnRNPA2 signal was still found in the nucleus. In contrast, the N323Tfs\*36 variant showed substantial accumulation in the cytoplasm before stress (Supplementary Fig. 4a) and strong association with stress granules upon oxidative stress (Supplementary Fig. 4b). When not associated with stress granules, the cytoplasmic distribution of N323Tfs\*36 was similar to that of eIF4G, which was diffuse in the cytoplasm, demonstrating that N323Tfs\*36 protein remains diffuse rather than in aggregates.

We next introduced fluorescent C-terminal tags to WT and frameshift hnRNPA2 mutants to characterize their respective distributions and to determine whether the C termini of frameshift variants might be cleaved from the full-length protein. In C2C12 cells, C-terminally GFP-tagged hnRNPA2 proteins showed equivalent localization patterns to N-terminally GFP-tagged hnRNPA2 proteins, with WT protein in the nucleus and frameshift mutant proteins showing accumulation in the cytoplasm (Supplementary Fig. 4c). We observed no cytoplasmic aggregation of C-terminally-tagged frameshift mutant proteins, consistent with our observations using N-terminally tagged mutant proteins in HeLa and C2C12 cells (Fig. 3c–f). Furthermore, the C-terminally GFP-tagged hnRNPA2 proteins migrated on a gel with an observed molecular weight of ~62 kDa, which is close to the estimated molecular weight of GFP-tagged full-length protein (hnRNPA2 37 kDa + GFP 27 kDa), and no cleaved product was observed by Western blot analyses (Supplementary Fig. 4d). In addition, solubility assays revealed that all three frameshift variants were more RIPA-soluble than WT proteins (Supplementary Fig. 4e). Thus, we conclude that frameshift variants cause diffuse cytoplasmic accumulation of hnRNPA2

protein in cells and that frameshift mutant proteins are less intrinsically aggregation-prone compared with WT protein.

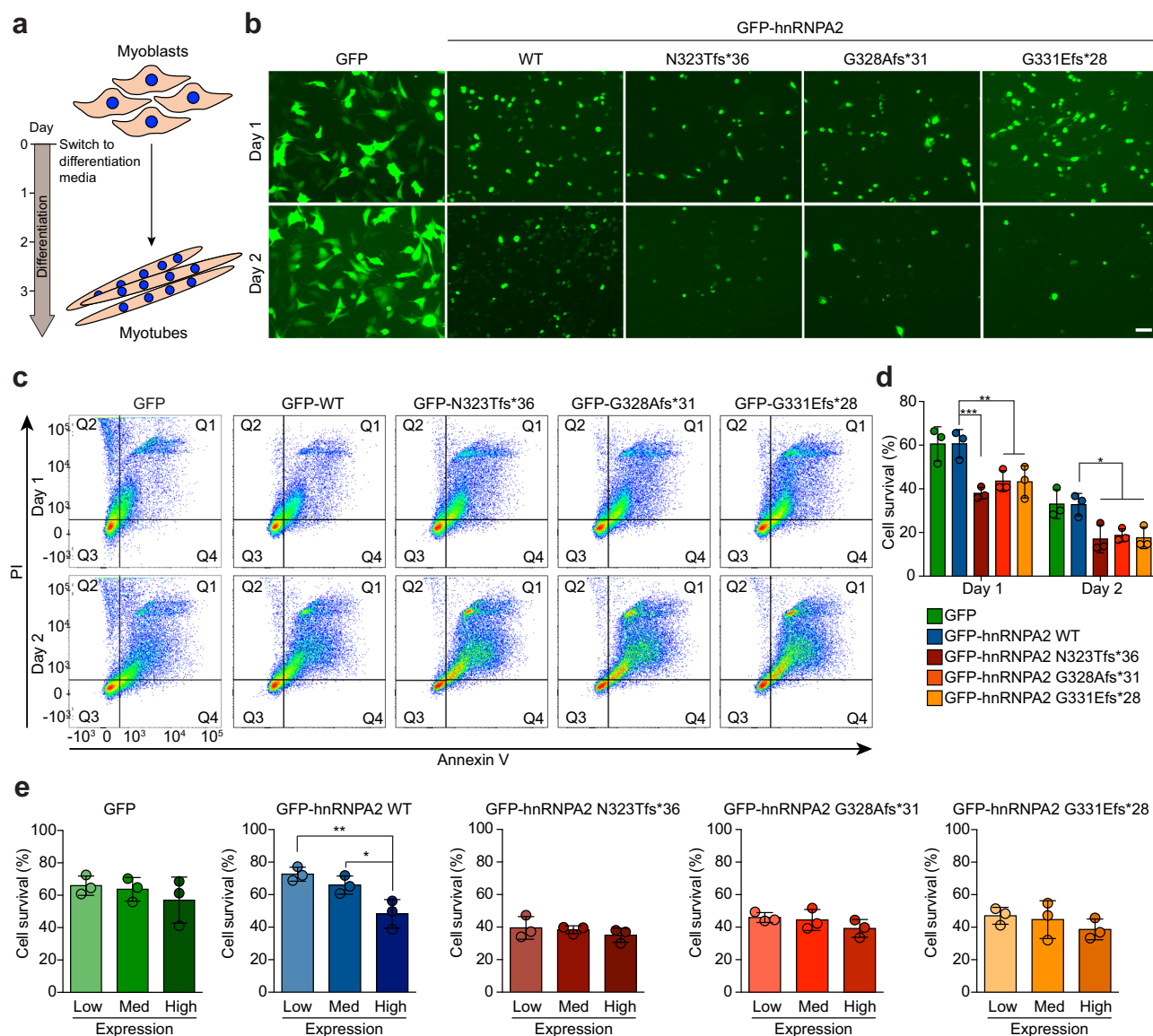
Distinct from findings in patient muscle biopsies, we did not observe hnRNPA2-positive protein aggregates in cells within the timeframe of our analyses. Thus, the hnRNPA2/B1-positive inclusions observed in patient muscle biopsies may not represent the immediate consequences of these expressed frameshift variants on protein aggregation and homeostasis, and may instead reflect a pathologic manifestation of the disease process over time.

### Frameshift variants impair the interaction between hnRNPA2 and its nuclear transport receptor Kap $\beta$ 2.

The conserved consensus motifs of the PY-NLS are located proximal to the region in which the *HNRNPA2B1* reading frame is altered by the frameshift variants (Fig. 3b and Supplementary Fig. 2b). Previous studies of hnRNPA1, a homologous protein of hnRNPA2/B1 that also harbors an M9-NLS, have shown that post-translational modifications of residues that are C-terminal to the conserved PY residues inhibit the interaction of hnRNPA1 with Kap $\beta$ 2 and impair nuclear import of hnRNPA1, indicating that C-terminal flanking regions can influence PY-NLS activity<sup>34,35</sup>. This prior observation in the closely related hnRNPA1 protein suggests that the cytoplasmic accumulation of frameshift mutant hnRNPA2 might reflect loss of Kap $\beta$ 2 interaction and subsequent impaired nuclear import. To test the impact of frameshift variants on Kap $\beta$ 2 binding, we synthesized WT and three frameshift mutant peptides (Fig. 3b) with a 5' tetramethylrhodamine (TAMRA) label and used fluorescence polarization assays to quantify their interaction with Kap $\beta$ 2<sup>36</sup>. The hnRNPA2 WT peptide bound Kap $\beta$ 2 with a  $K_d$  of 95 nM (Fig. 3g, h). All three frameshift mutant peptides showed a significant decrease in Kap $\beta$ 2 binding (Fig. 3g, h), suggesting that frameshift variants affect the ability of hnRNPA2 to bind Kap $\beta$ 2 and thus impair its subsequent nuclear import.

### Frameshift variants cause apoptotic cell death in differentiating cells.

To address the consequence of the frameshift mutants in a physiologically relevant cell type, we expressed WT or mutant versions of hnRNPA2 in C2C12 myoblasts and subsequently exposed the cells to differentiation factors to induce differentiation to myotubes (Fig. 4a). Whereas there was no obvious impact of exogenous expression of WT or mutant hnRNPA2 in myoblast cells, we observed that upon differentiation, the number of myotubes expressing mutant forms of hnRNPA2 decreased dramatically over time (Fig. 4b). Thus, to determine whether mutant forms of hnRNPA2 cause cell death in differentiating C2C12 cells, we monitored annexin V and propidium iodide (PI) staining to assess apoptotic cell death. Notably, cell death was significantly enhanced in the presence of the frameshift variants, with approximately 60% and 80% of frameshift mutant hnRNPA2-expressing cells becoming positive for annexin V and PI staining one and two days after differentiation, respectively (Fig. 4c, d, Supplementary Fig. 5a). In contrast, no increased cell death was observed in the overall population of cells expressing hnRNPA2 WT (Fig. 4c, d). However, when we specifically examined subpopulations of cells based on expression levels, we noted that the population of cells with the highest expression of hnRNPA2 WT did show increased cell death relative to controls (Fig. 4e). In cells expressing frameshift mutant forms of hnRNPA2, increased cell death was observed even in the population of cells with the lowest levels of mutant hnRNPA2 expression (Fig. 4e). We also examined the consequences of depleting endogenous hnRNPA2. Specifically, we found that depletion of wild-type endogenous hnRNPA2/B1 by expressing



**Fig. 4 Expression of frameshift variants causes increased cell toxicity during C2C12 differentiation.** **a** Schematic of C2C12 differentiation. Differentiation media consists of DMEM supplemented with 0.5% FBS, 1% penicillin/streptomycin, and 1% L-glutamate. **b** Representative images of C2C12 cells one (top) or two (bottom) days after change to differentiation media from three independent experiments. Scale bar, 20  $\mu$ m. **c** Representative scatter plots of GFP-positive C2C12 cells stained with propidium iodide (PI) and annexin V, one (top) or two (bottom) days after change to differentiation media.

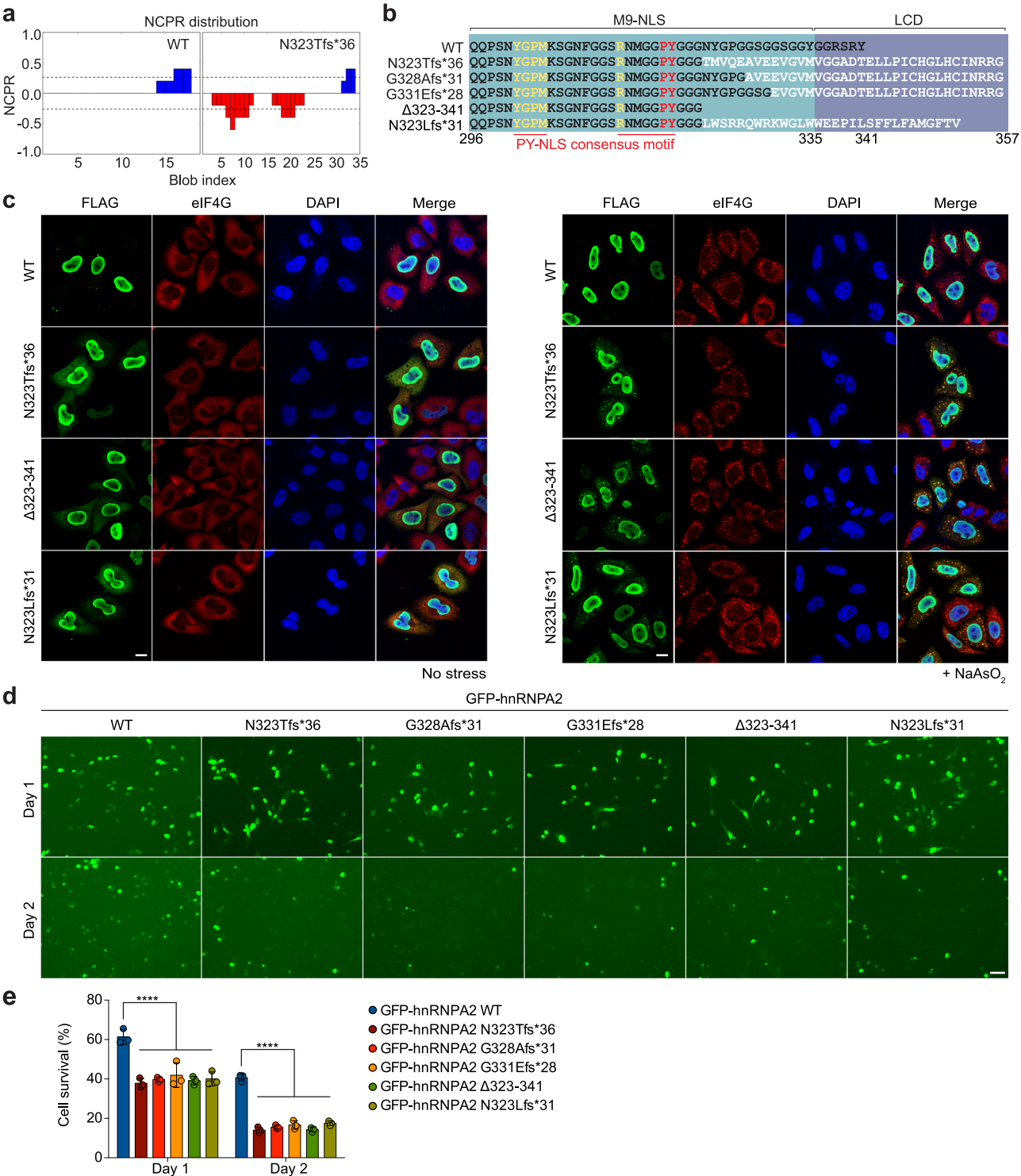
**d** Quantification of GFP-positive cells that were negative for both PI and annexin V. Values represent means  $\pm$  s.e.m. ( $n=3$  independent experiments). \*\*\* $P=0.0004$ , \*\* $P=0.0061$ , and \*\* $P=0.0052$  (day 1) and \* $P=0.0116$ , \* $P=0.0252$ , and \* $P=0.0147$  (day 2) for N323Tfs\*36, G328Afs\*31, and G331Efs\*28 mutants, respectively, by two-way ANOVA with Dunnett's multiple comparisons test. **e** Quantification of cell survival for all GFP-positive cells from day 1 post differentiation split into three categories. The range of GFP intensity was 500–140,000. Cells whose GFP intensity was below 45,000, between 45,000 and 90,000, or above 90,000 were grouped into low, medium, or high GFP-expressing cells, respectively. Values represent means  $\pm$  s.e.m. from  $n=3$  independent experiments. \* $P=0.0364$  and \*\* $P=0.0088$  by one-way ANOVA with Tukey's multiple comparisons test.

small interfering RNA against *HNRNPA2B1* did not cause cell toxicity (Supplementary Fig. 5b–d). These data suggest that despite the loss of Kap $\beta$ 2 binding activity by frameshift mutations, the mechanism of toxicity associated with these mutations is mediated by a gain of function rather than simply a loss of function.

**No evidence that the frameshift mutations introduce a specific toxic sequence.** We found it intriguing that all of the eoOPMD-associated variants we identified had a +1 frameshift resulting in a neomorphic amino acid sequence with a shared C-terminal end (Fig. 3b). This alteration introduces several negatively

charged amino acids and makes the C-terminal peptide more hydrophobic compared to the WT peptide (Fig. 5a, b). It is conceivable, therefore, that the precise sequence introduced by the frameshift mutation is responsible for the toxic gain of function. Thus, we tested whether the anomalous sequence was directly responsible for the phenotypes associated with frameshift variants of hnRNPA2. To this end, we generated two new constructs in which the C-terminal flanking sequence was either deleted ( $\Delta$ 323–341) or +2 frameshifted by a two-base-pair deletion (N323Lfs\*31) (Fig. 5b). Interestingly, both  $\Delta$ 323–341 and N323Lfs\*31 proteins accumulated in the cytoplasm (Fig. 5c) and caused cell toxicity in differentiated myoblasts at levels comparable to the eoOPMD-associated frameshift variants





**Fig. 5 A loss of wild-type C-terminal sequence, not a neomorphic sequence, is responsible for frameshift variant phenotypes.** **a** NCPR (net charge per residue; 5-aa window) of WT and N323Tfs\*36 peptides. Positively (blue) and negatively (red) charged amino acids are indicated. **b** Sequences in WT, +1 frameshift mutants in eoOPMD, Δ323-341, and +2 frameshift mutant (N323Lfs\*31). **c** Localization of FLAG-tagged hnRNP2 proteins without (left) or with (right) 0.5 mM NaAsO<sub>2</sub>. eIF4G was used as a cytoplasmic and stress granule marker. Representative images from three independent experiments are shown. Scale bar, 10 μm. **d** Representative images of C2C12 cells one (top) or two (bottom) days after change to differentiation media from three independent experiments. Scale bar, 20 μm. **e** Quantification of GFP-positive cells negative for both PI and annexin V. Values represent means ± s.e.m. ( $n = 3$  independent experiments). \*\*\*\* $P < 0.0001$  by two-way ANOVA with Dunnett's multiple comparisons test.

(Fig. 5d, e, Supplementary Fig. 5e). This result suggests that loss of the wild-type C-terminal flanking sequence, rather than addition of a specific, negatively charged C-terminal sequence, is responsible for the observed cellular phenotypes associated with *HNRNPA2B1* frameshift variants.

**Altered nucleocytoplasmic distribution of hnRNPA2 is responsible for the cytotoxicity associated with frameshift variants.** To further investigate the contribution of nucleocytoplasmic transport defects to the toxic mechanism of frameshift variants, we generated a series of additional PY-NLS mutants. These included the introduction of missense mutations (P318A and Y319A) to or deletions ( $\Delta$ PY and  $\Delta$ PY-NLS) of key residues in the NLS. We found that a single amino acid substitution (P318A or Y319A) or a deletion ( $\Delta$ PY) of the PY motif did not alter nuclear localization of hnRNPA2 (Fig. 6a, b)<sup>37</sup>. In contrast, the  $\Delta$ PY-NLS mutant (in which residues spanning the two conserved PY-NLS consensus motifs, **YGPMKSGNFGGSRNMGGPY** [aa 301–319], were deleted) showed cytoplasmic accumulation of the protein that was comparable to the N323Tfs\*36 variant (Fig. 6a, b). The  $\Delta$ 323–341 mutant (in which the C-terminal flanking sequence was deleted) also showed cytoplasmic accumulation of the protein that was comparable to the N323Tfs\*36 variant (Fig. 5b, c). Taken together, these results confirm the importance of PY-NLS flanking sequences in binding Kap $\beta$ 2 and are consistent with previous structural studies of hnRNPA1 that revealed that in addition to modest energetic contributions of the PY residues to binding Kap $\beta$ 2, important electrostatic interactions are also provided by the adjacent C-terminal flanking regions<sup>34,35</sup>. Importantly, the partial shift to cytoplasmic localization of hnRNPA2 was sufficient to drive toxicity. Indeed, expression of the artificial  $\Delta$ PY-NLS mutant in differentiated C2C12 cells, which causes modest cytoplasmic accumulation of this protein, resulted in cellular toxicity similar to the frameshift mutants (Fig. 6c, d).

To complement these additional PY-NLS mutants, we also designed a rescue experiment. Specifically, we added the canonical monopartite cNLS sequence from simian virus 40 (SV40) T-antigen<sup>38</sup> to the C terminus of N323Tfs\*36. Not only did appending the NLS sequence onto this mutant hnRNPA2 fully rescue nuclear localization of N323Tfs\*36 (Fig. 6a, b), but it also prevented cell toxicity (Fig. 6c, d). Thus, we conclude that the wild-type C-terminal flanking sequence is important for nuclear localization of hnRNPA2 by regulating its interaction with the nuclear transport receptor Kap $\beta$ 2, that this interaction is partly impaired by frameshift mutation, and that the altered distribution of hnRNPA2 leads to cell toxicity.

**Frameshift variants cause eye and muscle degeneration in a *Drosophila* model.** We next investigated the in vivo effects of missense and frameshift mutations in a *Drosophila* model system. To this end, we generated transgenic *Drosophila* expressing human hnRNPA2 WT, the MSP-associated mutant D290V<sup>6</sup>, or frameshift mutants (N323Tfs\*36, G328Afs\*31, and G331Efs\*28) via PhiC31 integrase-mediated site-specific insertion of a single copy of the human *HNRNPA2B1* gene. We established multiple fly lines per genotype and observed that all lines expressing the N323Tfs\*36 mutant, and some lines expressing G328Afs\*31 and G331Efs\*28, showed low or no protein expression with the eye tissue-targeting GMR-GAL4 driver, which may reflect cytotoxicity associated with expression (Supplementary Fig. 6a). From the remaining lines, we selected one G328Afs\*31 line and one G331Efs\*28 line for subsequent analysis and compared the consequences of transgene expression with flies expressing

similar levels of hnRNPA2 WT or D290V (Supplementary Fig. 6a).

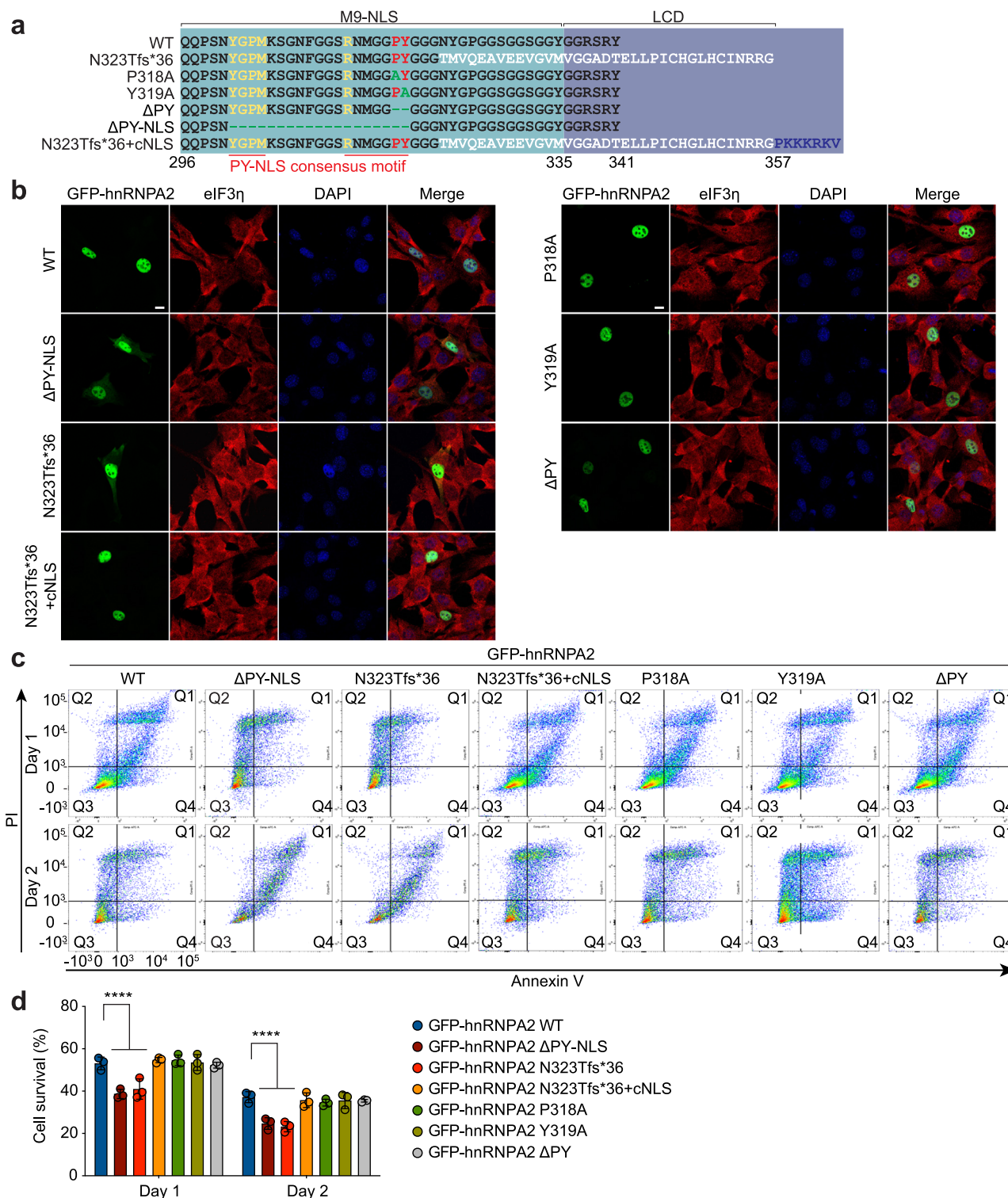
Expression of these transgenes in the fly eye revealed a mutation-dependent rough eye phenotype that was modest at 22 °C but enhanced at 25 °C, a temperature that causes a greater level of transgene expression (Fig. 7a)<sup>39</sup>. At 25 °C, flies expressing G328Afs\*31 showed pupal lethality, although one lethality “escaper” G328Afs\*31-expressing fly exhibited severe eye degeneration (Fig. 7a). When expression was driven by the muscle-specific driver MHC-GAL4, we observed robust expression of N323Tfs\*36, G328Afs\*31, and G331Efs\*28 mutant proteins, in contrast with the low expression induced by GMR-GAL4 (Supplementary Fig. 6b). Although a single copy of the *HNRNPA2B1* gene was inserted in all lines, hnRNPA2 protein levels were consistently modestly higher in the N323Tfs\*36, G328Afs\*31, and G331Efs\*28 flies compared with the hnRNPA2 WT and D290V flies, suggesting that the frameshift mutations may increase the stability of hnRNPA2 protein in fly muscles (Supplementary Fig. 6b).

A mutation-dependent phenotype was more obvious when the transgenes were expressed in skeletal muscle and assessed by wing position, as previously described<sup>40</sup>. Whereas flies expressing hnRNPA2 WT had normal wing position, flies expressing either hnRNPA2 with a missense mutation (D290V) or a frameshift mutation (N323Tfs\*36, G328Afs\*31, or G331Efs\*28) exhibited a wing position defect. In these animals, a mild mutation-dependent wing position defect was evident on the first day after hatching (eclosion) and was found to affect up to 100% of the animals by day 5 after eclosion (Fig. 7b).

Consistent with the normal wing position, expression of hnRNPA2 WT in skeletal muscle caused no histological defect in adult skeletal muscle (Fig. 7c). Tissue architecture and myofibril organization were normal, and hnRNPA2 protein was detected only in the nucleus (Fig. 7c). In contrast, expression of either the missense or frameshift mutants caused a significant disruption in cytoarchitecture as evidenced by defects in the organization of myofibrils and substantial redistribution of the protein from the nucleus into the cytoplasm (Fig. 7c). However, there was a considerable difference in the cytoplasmic distribution of the hnRNPA2 protein in animals expressing missense vs. frameshift mutants. The missense mutant (D290V) accumulated in dense sarcoplasmic punctae, consistent with our prior observation in both *Drosophila* and human cells that hnRNPA2 D290V accumulates in sarcoplasmic RNP granules<sup>6,41</sup>. In contrast, the frameshift mutants accumulated in the sarcoplasm with a diffuse distribution (Fig. 7c–e).

We previously showed that the missense (D290V) mutant hnRNPA2 protein shows a stark reduction in solubility relative to WT when expressed in either human HeLa cells or *Drosophila* skeletal muscle cells<sup>6</sup>. In contrast, frameshift (N323Tfs\*36, G328Afs\*31, and G331Efs\*28) mutants showed no reduction in solubility in *Drosophila* skeletal muscle cells (Fig. 7f, g) and indeed showed somewhat increased solubility when expressed in human cells (Supplementary Fig. 4e). These results suggest that the consequences of these two classes of mutations on intrinsic properties of hnRNPA2 protein may be very different.

To directly address the consequences of the different classes of mutation on the biophysical properties of hnRNPA2, we examined the fibrillization kinetics of purified proteins. These results revealed that frameshift variants exhibited decelerated fibrillization kinetics compared to WT protein, whereas hnRNPA2 D290V assembled into fibrils at a faster rate, as previously reported (Supplementary Fig. 7)<sup>6</sup>. Specifically, the N323Tfs\*36 mutant formed very few fibrils over 24 h, whereas the G328Afs\*31 mutant fibrillized only after a long lag phase



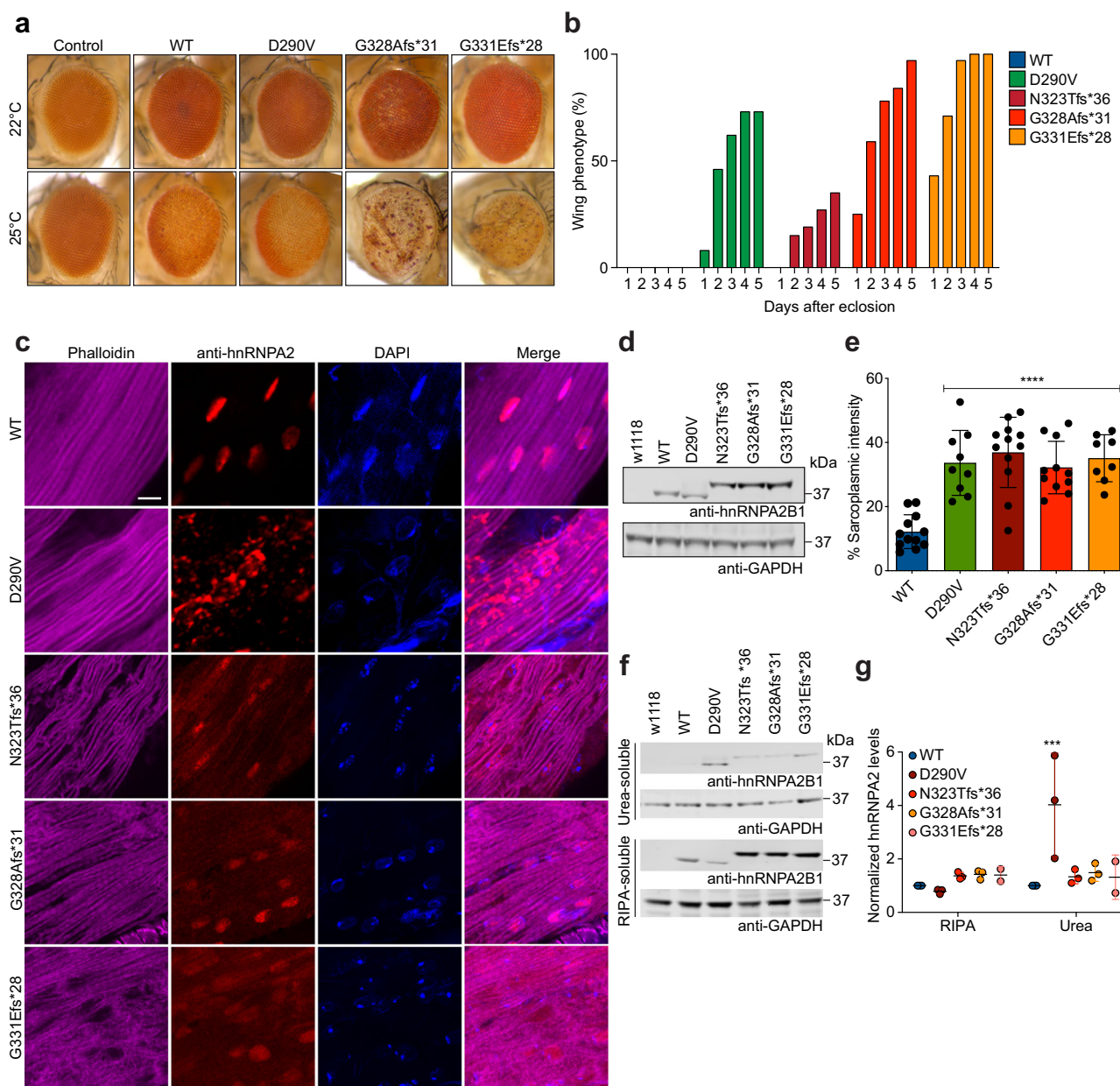
**Fig. 6** Altered nucleocytoplasmic distribution of hnRNP2 is responsible for the cytotoxicity associated with its frameshift variants. **a** The amino acid sequences of the relevant domains in WT, N323Tfs\*36, P318A, Y319A,  $\Delta$ PY,  $\Delta$ PY-NLS and N323Tfs\*36+cNLS. **b** Intracellular localization of FLAG-tagged hnRNP2 proteins. Representative images from three independent experiments are shown. Scale bar, 10  $\mu$ m. **c** Representative scatter plots of GFP-positive C2C12 cells stained with propidium iodide (PI) and annexin V, one (top) or two (bottom) days after change to differentiation media. **d** Quantification of GFP-positive cells that were negative for both PI and annexin V. Values represent means  $\pm$  s.e.m. ( $n = 3$  independent experiments). \*\*\*\* $P < 0.0001$  by two-way ANOVA with Dunnett's multiple comparisons test.

(Supplementary Fig. 7a, b). Thus, these findings suggest that the pathogenic mechanism of the frameshift variants differs from that of D290V and likely involves reduced efficiency of nuclear import rather than an increased propensity toward fibrillization.

## Discussion

Here we describe a distinct phenotype of early-onset myopathy caused by specific heterozygous *HNRNP2B1* frameshift mutations clinically manifesting with progressive muscle weakness,





**Fig. 7** Frameshift variants cause eye and muscle degeneration in a *Drosophila* model. **a** Expression of frameshift variants in *Drosophila* eye tissue using GMR-GAL4 causes a rough eye phenotype. **b** Expression of frameshift variants in *Drosophila* muscle using MHC-GAL4 causes an abnormal wing posture phenotype. The percentage of flies with abnormal wing posture is plotted for 5 days after eclosion.  $n = 26, 26, 26, 32,$  and  $35$  flies for WT, D290V, N323Tfs\*36, G328Afs\*31, and G331Efs\*28, respectively. **c** Adult flies expressing hnRNP2 transgene using MHC-GAL4 were dissected to expose the dorsal longitudinal indirect flight muscle and stained with rhodamine-phalloidin (purple), hnRNP2 (red), and DAPI (blue). hnRNP2 WT localized exclusively to nuclei, whereas hnRNP2 D290V accumulated extensively in cytoplasmic inclusions. Frameshift mutants showed both nuclear staining and diffuse sarcoplasmic accumulation. Scale bar,  $10\ \mu\text{m}$ . **d** hnRNP2 expression in thoraces of adult flies driven by MHC-GAL4 in transgenic flies. Representative immunoblots from three independent experiments are shown. **e** Ratio of sarcoplasmic intensity of hnRNP2 signal in indirect flight muscles. Error bars represent mean  $\pm$  s.d.  $n = 12, 9, 12, 11,$  and  $8$  fly muscle samples for WT, D290V, N323Tfs\*36, G328Afs\*31, and G331Efs\*28, respectively. \*\*\*\* $P < 0.0001$  by one-way ANOVA with Dunnett's multiple comparisons test. **f** Thoraces of adult flies expressing hnRNP2 transgene using MHC-GAL4 were dissected and sequential extractions were performed to examine the solubility profile of hnRNP2. w1118 flies are included as a non-transgenic control. Representative immunoblots from three independent experiments are shown. **g** Quantification of RIPA-soluble and -insoluble fraction of hnRNP2. Data represent mean  $\pm$  s.d.,  $n = 3$  (WT, D290V, N323Tfs\*36, G328Afs\*31) and  $2$  (G331Efs\*28) independent experiments. \*\*\* $P = 0.0002$  by two-way ANOVA with Dunnett's multiple comparisons test.

ptosis, ophthalmoplegia, dysphagia, and variable degrees of respiratory insufficiency. With the notable exception of the very early disease onset and rapid progression in our cohort, the overall clinical presentation was reminiscent of classical OPMD, such that the disease could be understood as a distinct early-onset

form of OPMD (eoOPMD). The uniform absence of bone, cognitive and motor neuron involvement in our patients distinguishes our cohort from the previously described MSP phenotype associated with a p.D290V *HNRNP2B1* variant<sup>6</sup>. In addition, ptosis, ophthalmoparesis, and dysphagia were absent in the



reported *HNRNPA2B1* missense (D290V) family, whereas these were consistent phenotypic features within our patient cohort, further differentiating these distinct *HNRNPA2B1*-related phenotypes.

Muscle pathologic features in our *HNRNPA2B1* cohort were overall consistent with a chronic, degenerative myopathy, with notable histological findings of rimmed autophagic vacuoles and cytoplasmic and intranuclear tubulofilamentous inclusions. These non-specific findings have been reported in several myopathies, including *HNRNPA2B1* missense variant MSP, inclusion body myositis<sup>42</sup>, MYH2-related inclusion body myopathy<sup>43</sup>, hnRNPA1-associated inclusion body myopathy<sup>44</sup>, TIA1-related myopathies<sup>45</sup>, hnRNPD-related myopathies<sup>11</sup>, and OPMD<sup>46</sup>, although the intranuclear tubulofilamentous inclusions we noted lacked OPMD-associated palisading morphology. The relevance of autophagic vacuoles and tubular filamentous inclusions to the pathophysiology of the specific myopathic phenotype observed in our patients remains unclear.

Since the identification of the first *HNRNPA2B1* MSP family, screening of MSP and large ALS patient cohorts has shown that *HNRNPA2B1* variants are a rare cause of sporadic and familial motor neuron disease<sup>47–49</sup>; however, its potential role in myopathies has remained unclear. Degenerative myopathy was observed in several family members carrying the missense D290V mutations<sup>6</sup>. More recently, a p.P298L missense *HNRNPA2B1* mutation was found to cause isolated familial PDB without multisystem involvement, highlighting the divergent phenotypes that may arise from the various *HNRNPA2B1* disease variants<sup>10</sup>. In contrast with previously reported variants, the *HNRNPA2B1* variants reported here are all specific frameshift mutations that tightly cluster in the highly conserved LCD of the protein and escape RNA quality control degradation (Supplementary Fig. 2). These frameshift *HNRNPA2B1* variants alter the C-terminal M9-NLS amino acid sequence of the protein, while sparing the immediately upstream PY residues. Further investigation of the sequence features that contribute to frameshift variant-associated phenotypes indicated that partial loss of Kap $\beta$ 2 binding, rather than gain of a neomorphic amino acid sequence in frameshift variants, is responsible for the cellular phenotypes observed (Fig. 5, Supplementary Fig. 5). Thus, we suggest that the C-terminal flanking sequence after the PY-NLS of hnRNPA2/B1 contributes to Kap $\beta$ 2 binding, similar to the molecular interaction of the closely related hnRNPA1 protein with Kap $\beta$ 2<sup>34,35</sup>. We propose further that loss of this C-terminal flanking sequence is responsible for the partial loss of interaction with Kap $\beta$ 2, which is the most proximal consequence of the mutations and initiates the slow pathological process. It remains to be seen whether the same eoOPMD phenotype is observed in rare individuals with pure truncations or alternative frameshifts in the *HNRNPA2B1* C-terminal flanking sequence. While it is possible that loss of hnRNPA2/B1 nuclear function may contribute to disease progression, the overall reduction in nuclear localization of hnRNPA2/B1 is modest. Moreover, we found that strong knockdown of endogenous hnRNPA2/B1 is well tolerated in differentiating myotubes, at least on a time scale of days. Rather, the data presented herein suggests that the disease mechanism promoted by frameshift mutations is toxic gain of function in the cytoplasm. It is important to note that Kap $\beta$ 2 not only regulates nuclear import of hnRNPA2/B1 (and other clients such as hnRNPA1 and FUS) but also regulates higher-order assembly of these RBPs via LLPS and fibrillization<sup>50–52</sup>. Thus it is possible that loss of Kap $\beta$ 2-mediated chaperoning of hnRNPA2/B1 in the cytoplasm may contribute to pathogenesis, and this will require further investigation.

The similarity in muscle pathological features of *HNRNPA2B1*-associated MSP and eoOPMD may appear paradoxical, as these

two classes of mutation have distinct consequences for the hnRNPA2 protein: the first (missense mutations in the LCD sequence) increases the propensity toward fibrillization, whereas the second (frameshift mutations that partially impair nuclear import) increases the cytoplasmic concentration. The long-term consequence of both classes of mutations is the accrual of amorphous aggregates that include the hnRNPA2 protein. Thus, our working model is that there are two distinct, but not mutually exclusive mechanisms that can lead to pathological proteinaceous deposits<sup>20,21</sup>. First, pathological aggregates can arise from mutations that reduce the energy threshold for fibrillization. In this category are mutations that alter key residues in the LCDs of RBPs (e.g., D290V in hnRNPA2/B1, A315T in TDP-43). Second, pathological aggregates can also arise through the evolution of poorly dynamic RNP granules. RNP granules are dynamic liquids that, when perturbed, can undergo a liquid-to-solid phase transition that culminates in pathological aggregates similar to those that arise through primary nucleation. Such perturbations include increased concentration of constituent proteins (e.g., RBPs) and impairment of regulatory factors (e.g., VCP). Disease-causing mutations in this second category include (1) NLS mutations that increase the cytosolic concentration of RBPs (e.g., the frameshift mutations in hnRNPA2/B1 reported here, and a series of mutations in FUS that are clustered at the NLS), (2) hexanucleotide expansion in *C9ORF72* that produces dipeptide repeats that insinuate into biomolecular condensates and alter their material properties<sup>53,54</sup>, and (3) mutations in VCP that impair its ability to disassemble RNP granules<sup>55</sup>. Although the nature and direct molecular consequences of different classes of mutation are clearly different (e.g., D290V vs. the eoOPMD frameshift variants in *HNRNPA2B1*), mutations in both categories can ultimately lead to pathological phase transition and deposition of protein inclusions over time. Indeed, in human pathological studies of ALS/FTD brain samples, which at late stages are also characterized by presence of TDP-43 cytoplasmic aggregates, loss of nuclear TDP-43 staining precedes the development of cytoplasmic inclusions by several years<sup>56</sup>. Thus, we suggest that the hnRNPA2/B1 aggregates in our patient biopsies are the end-stage epiphenomenon of the disease process over a long period of time and are preceded by abnormal nucleocytoplasmic transport dynamics of frameshift variant hnRNPA2/B1 and its accumulation in the cytoplasm. This may also explain why a plethora of RBPs in RNP granules lead to co-deposition of additional RBPs (e.g., TIA1 and TDP-43) that appear as pathological features.

All the *HNRNPA2B1* frameshift variants reported here were found in heterozygosity and thus appear to act in a dominant manner. This observation raises the possibility of loss of function (i.e., haploinsufficiency), gain of function, or dominant-negative pathomechanisms. As noted above, missense mutations in the LCD of hnRNPA2 may drive primary nucleation of hnRNPA2 fibrils and culminate in the accrual of pathological aggregates, reflecting a straightforward gain of toxic function. In contrast, frameshift variants in hnRNPA2 partially impair nuclear import—a partial loss of function. However, the consequence of this impaired nuclear import is persistently increased concentration of hnRNPA2 in the cytoplasm. Substantial evidence has now accrued that increased concentration of an RNP granule constituent is one of several mechanisms that impairs the dynamics of RNP granules, culminating in liquid-to-solid phase transition and evolution toward pathological aggregates that also include hnRNPA2 protein<sup>20,21</sup>. Thus, whereas the primary consequence of the frameshift mutations is loss of function (impaired binding to Kap $\beta$ 2), we suggest that the resulting increased cytoplasmic concentration ultimately results in the same consequences as the

missense mutations—a toxic gain of function in the cytoplasm. This model is further supported by the results of knockdown experiments showing that depletion of endogenous *HNRNPA2B1* by siRNA does not cause toxicity in differentiated C2C12 cells, arguing against a simple loss of function mechanism (Supplementary Fig. 5b–d).

It is noteworthy that similar C-terminal frameshift variants in *HNRNPA1* (\*321Eext\*6, \*321Qext\*6, and G304Nfs\*3) were identified in a broad spectrum of patients with hereditary motor neuropathy, ALS and myopathy<sup>7</sup>. The amino acid sequences altered by these frameshift mutations in hnRNPA1 are located far downstream of the PY-NLS, and the impact of these frameshift mutations on nucleocytoplasmic transport of hnRNPA1 is not as clear as observed for the *HNRNPA2B1* frameshift variants. Nevertheless, similar to the eoOPMD-associated *HNRNPA2B1* frameshift variants, frameshift variants of *HNRNPA1* showed reduced propensity for fibrillization, and one variant with an extension of amino acids (\*321Eext\*6) accumulated in cytoplasmic stress granules in response to stress, which can often be attributed to increased concentration of the protein in the cytoplasm<sup>7</sup>. Thus, the frameshift variants of *HNRNPA1* and *HNRNPA2B1* might share a common underlying pathomechanism.

Among the myopathies, emerging data suggest a specific role for LCD-containing RBPs such as TDP-43 and hnRNPA2/B1 in regenerating muscle, where they are thought to stabilize large muscle-specific transcripts (e.g., TTN, NEB), aid in their transport, and facilitate pre-mRNA splicing<sup>57</sup>. In particular, wild-type hnRNPA2/B1 retains its exclusive nuclear localization during muscle regeneration<sup>57</sup>. In association with their altered nucleocytoplasmic distribution, expression of *HNRNPA2B1* frameshift variants in our study resulted in cell toxicity in differentiating C2C12 myoblasts (Fig. 4) and degeneration of flight muscle in *Drosophila* (Fig. 7), suggesting an essential role for hnRNPA2/B1 in maintaining skeletal muscle integrity. Thus, with ongoing use and stress of the muscle, impaired hnRNPA2/B1 nuclear import may be inadequate to maintain muscle homeostasis over time, clinically manifesting as the progressive muscle weakness seen in our patients. The phenotypic relationship of our patients to typical OPMD is also of great interest, as OPMD is caused by polyaniline expansions in PABPN1, whose normal splicing in the nucleus is partially regulated by hnRNPA2/B1<sup>58</sup>. Thus, future studies should focus on the role of RBPs such as hnRNPA2/B1 in specific cell types, specific subcellular compartments, and in response to different physiologic states to identify their respective contributions to pathogenic processes and subsequent clinical manifestations.

Our data expand the clinical spectrum of *HNRNPA2B1* variants from MSP to include a distinct early-onset OPMD-like phenotype. Understanding how these seemingly divergent phenotypes emerge from common molecular and cellular events will likely uncover fundamental insights into RBP function and regulation that will be applicable to a broad array of neurodegenerative diseases.

## Methods

**Patient recruitment and sample collection.** Patients were recruited through local neurology and genetics clinics and clinical information was obtained based on the local standard clinical care. DNA and tissues (e.g., muscle, skin biopsy) and medical records were obtained based on standard procedures. The authors affirm that human research participants and/or legal guardian have seen and read the material to be published and have provided informed consent for publication of the images in Fig. 1. Ethical approval was obtained from the NIH, National Institute of Neurological Disorders and Stroke (NINDS), Institutional Review Board (Protocol 12-N-0095), National Center of Neurology and Psychiatry (Protocol A2019-123), University of Strasbourg (Protocol DC-2012-1693), Cambridge South, UK Research

Ethics Committee (approval 13/EE/0325), Health Research Authority, NRES Committee East of England—Hatfield (REC 13/EE/0398; REC 06/Q0406/33) and National Research Ethics Service (NRES) Committee North East—Newcastle & North Tyneside 1 (reference 08/H0906/28).

**Exome sequencing.** Quartet exome sequencing in family 1 was performed through the NIH Intramural Sequencing Center (NISC) using the Illumina TruSeq Exome Enrichment Kit and Illumina HiSeq 2500 sequencing instruments. Variants were analyzed using seqr (Center for Mendelian Genomics, Broad Institute). Trio exome sequencing in family 2 was performed at GeneDX with exon targets isolated by capture using the Agilent SureSelect Human All Exon V4 (50 Mb) kit or the Clinical Research Exome (Agilent Technologies). The sequencing methodology and variant interpretation protocol has been previously described<sup>59</sup>.

Patient 3 was sequenced as a singleton as part of the MYO-SEQ project<sup>60</sup>. Exome sequencing was performed by the Genomics Platform at the Broad Institute. Libraries were created with an Illumina exome capture (38 Mb target) and sequenced with a mean target coverage of >80x. Exome sequencing data were analyzed on seqr (<https://seqr.broadinstitute.org/>).

For family 4, exome sequencing of patients 4 and 5 was undertaken using Agilent SureSelect Human All Exon 50 Mb capture kit followed by sequencing on a 5500XL SOLiD sequencer. The affected mother and unaffected father were sequenced using Nextera Rapid Capture Expanded Exome for target selection followed by sequencing on an Illumina HiSeq2000. Genomic data were processed as previously described<sup>61</sup>. Confirmation of variants and segregation was performed by Sanger sequencing.

WES analysis in proband (P6) from family 5 was performed by deCODE genetics, Iceland. The alignment to the human reference genome (hg19) and the variant calling was done by deCODE genetics, Iceland. Data analysis was carried out using Clinical Sequence Miner platform, NextCODE Health.

Trio exome sequencing in family 6 was performed at the Centre National de Recherche en Génomique Humaine (Evry, France) using the Agilent SureSelect Human All Exon V4. Sequence analysis and variant interpretation was performed as described<sup>62</sup>.

For family 7, the patient underwent standard clinical next generation sequencing panels that included LGMD and congenital myopathy genes. Due to the remarkably similar phenotype to the other patients in this cohort, direct Sanger sequencing of the terminal exon of *HNRNPA2B1* was pursued and identified the de novo variant.

For family 8, the patient's DNA was analyzed using the Illumina TruSeq ONE Expanded kit on an Illumina NextSeq 550 sequencer as described<sup>63</sup>. Segregation analysis was performed by Sanger sequencing.

For family 9, genome sequencing in patient 10 and her parents was performed using the Illumina HiSeq X Ten platform and variant calling and interpretation were performed as previously described<sup>64</sup>. Confirmation of variants was performed by Sanger sequencing.

For family 10, exome sequencing in patient 11 was performed at Beijing Genomics Institute (BGI) using the DNBSEQ. Exome sequencing and variant calling were performed as previously described<sup>65</sup>. Confirmation of variants was performed by Sanger sequencing.

**Muscle histology, immunofluorescence, and confocal microscopy.** Clinical muscle biopsy slides were obtained and reviewed. These included hematoxylin and eosin, modified Gömöri trichrome, NADH, and other histochemical stains as well as electron microscopy images when available. For immunofluorescence and confocal microscopy, frozen muscle biopsy tissues were cryo-sectioned (10 µm), fixed (100% acetone, −20 °C for 10 min), and blocked and permeabilized in 5% normal goat serum (Sigma) with 0.5% Triton X in PBS for 1 h at room temperature. Primary antibody incubation was performed overnight at 4 °C as follows: hnRNPA2B1 (Santa Cruz, sc-32316; mouse, 1:200), TDP-43 (Proteintech, 10782-2-AP; rabbit, 1:200), ubiquitin (Stressgen, SPA-200; rabbit, 1:200), ubiquilin-2 (Abcam, Ab190283; mouse IgG1, 1:500), TIA1 (Abcam, Ab140595; rabbit, 1:100), P62/SQSTM1 (Santa Cruz, sc-28359; mouse IgG1, 1:250). Secondary antibody incubation was performed for 1 h at room temperature. Sections were then washed in PBS, the nuclei were stained with DAPI, and sections were mounted and coverslipped. Z-stack images were obtained using a Leica TCS SP5 II confocal microscope.

**Muscle MRI.** Muscle MRI was performed using conventional T1-weighted spin echo and short tau inversion recovery (STIR) of the lower extremities on different scanners at different centers.

**Plasmid constructs.** cDNA containing frameshift mutations of hnRNPA2 were synthesized by GenScript. For mammalian expression, FLAG-tagged WT, D290V, N323Tfs\*36, G328Afs\*31, G331Efs\*28, Δ323–341, and N323Lfs\*31 were cloned into pCAGGS vector at *SacI* and *SbfI* sites. N-terminal GFP-tagged WT, D290V, N323Tfs\*36, G328Afs\*31, and G331Efs\*28 were cloned into pEGFP-C1 vector at *BsrGI* and *XhoI* sites. N-terminal GFP-tagged Δ323–341, N323Lfs\*31, P318A, Y319A, ΔPY, ΔPY-NLS, and N323Tfs+cNLS were subjected to site-directed

mutagenesis using Q5 Site Directed Mutagenesis Kit (E0554S; NEB). C-terminal GFP-tagged WT, D290V, N323Tfs\*36, G328Afs\*31, and G331Efs\*28 were cloned into pEGFP-N1 vector at *XhoI* and *PstI* sites. For transgenic *Drosophila*, mutant hnRNA2 cDNAs were subcloned into the pUASTattB vector using *EcoRI* and *XhoI*. For bacterial expression, codon optimized cDNA containing frameshift mutations of hnRNA2 were synthesized and subcloned into the pGST-Duet vector using *BamHI* and *EcoRI* by GenScript. All clones were verified by restriction enzyme digestion and sequence analysis.

**Cell lines.** The following cell lines were purchased from the American Type Culture Collection: HEK293T (ATCC CRL-11268), HeLa (ATCC CCL-2), and C2C12 (ATCC CRL-1772). Cells were authenticated by short tandem repeat profiling. All cell lines were tested to be mycoplasma negative.

**Cell culture, transfection, and immunofluorescence.** HEK293T and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, and 1% L-glutamate. Cells were transfected using FuGene 6 (Promega) according to the manufacturer's instructions. For immunofluorescence, HeLa cells were seeded on eight-well glass slides (Millipore) and transfected with appropriate FLAG-tagged or GFP-tagged hnRNA2 constructs. 24 h post transfection, cells were stressed with 500  $\mu$ M sodium arsenite (Sigma-Aldrich) for indicated times. Cells were then fixed with 4% paraformaldehyde (Electron Microscopy Sciences), permeabilized with 0.5% Triton X-100, and blocked in 3% bovine serum albumin. Primary antibodies used were mouse monoclonal anti-FLAG (M2, F1804; Sigma), rabbit polyclonal anti-eIF4G (H-300, sc-11373; Santa Cruz Biotechnology), and mouse monoclonal anti-hnRNA2B1 (EF-67, sc-53531; Santa Cruz Biotechnology) antibodies. For visualization, the appropriate host-specific Alexa Fluor 488, 555 or 647 (Molecular Probes) secondary antibody was used. Slides were mounted using Prolong Gold Antifade Reagent with DAPI (Life Technologies). Images were captured using a Leica TCS SP8 STED 3X confocal microscope (Leica Biosystems) with a  $\times 63$  objective. Images were quantified using Cell Profiler (Broad Institute) for cytoplasmic accumulation. Briefly, images were subjected to segmentation and integrated fluorescent intensity was calculated for whole cell, nucleus, and cytoplasm. Cytoplasmic percent is simply cytoplasmic intensity divided by whole cell intensity. For fluorescent distribution across the cell, ImageJ (NIH) was used. A straight line was overlaid across the cell and then the fluorescent intensity was measured across the line using the built-in function.

**Western blot analysis.** Cell lysates were prepared by lysing cells in 1 $\times$  lysis buffer (150 mM NaCl, 25 mM Tris-HCl pH 7.5, 1 mM EDTA, 5% glycerol, and 1% NP-40) with Complete Protease Inhibitor Cocktail (Clontech Laboratories). Samples were resolved by electrophoresis on NuPAGE Novex 4–12% Bis-Tris gels (Invitrogen). Primary antibodies used were mouse monoclonal anti-FLAG (M2, F1804; Sigma), rabbit polyclonal anti-FLAG (F7425; Sigma), rabbit polyclonal anti-GFP (2555S; Cell Signaling), rabbit-polyclonal anti-hnRNA2B1 (HPA001666, Sigma), rabbit-polyclonal anti-hnRNA2B1 (NBP2-56497; NOVUS Biologicals), and mouse monoclonal anti-GAPDH antibodies (6C5, sc-32233; Santa Cruz Biotechnology). Blots were subsequently incubated with IRDye fluorescence-labeled secondary antibodies (LI-COR) and protein bands were visualized using the Odyssey Fc system (LI-COR) and Image Studio (LI-COR). Quantification was performed using ImageJ (NIH).

**Cell toxicity.** C2C12 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, and 1% L-glutamate. Cells were transfected using Lipofectamine 300 reagent (ThermoFisher Scientific) according to the manufacturer's instructions. C2C12 cells were counted using ADAM-CellT (NanoEntek Inc., Seoul, Korea), plated in six-well dishes (Corning), and transfected with appropriate GFP-tagged hnRNA2 constructs. 24 h post transfection, the media was changed to differentiation media (DMEM supplemented with 0.5% FBS, 1% penicillin/streptomycin, and 1% L-glutamate) and counted as day 0. After 1 or 2 days, the GFP signal of the cells were imaged with an EVOS microscope. Apoptotic cells were determined by staining with annexin-V-APC and propidium iodide (BD Biosciences) and measured by flow cytometry. Data was analyzed using FlowJo\_v10.6.1.

**siRNA knockdown.** C2C12 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin and 1% L-glutamate. Cells were transfected using RNAiMAX reagent (ThermoFisher Scientific; 13778075) according to the manufacturer's instructions. C2C12 cells were counted using ADAM-CellT (NanoEntek Inc., Seoul, Korea), plated in six-well dishes (Corning), and transfected with either ON-TARGETplus Non-Targeting Pool siRNA (Dharmacon; D-001810-10-05) or ON-TARGETplus Mouse HnRNA2b1 siRNA (Dharmacon; L-040194-01-0005). 72 h post transfection, cells were used for subsequent experiments. Knockdown efficiency was determined by Western blot using mouse monoclonal anti-hnRNA2/B1 (DP3B3, Santa Cruz Biotechnology; sc-32316) and goat polyclonal anti-actin (Santa Cruz Biotechnology; sc-1616). Target sequences for control siRNA were: UGUUUUACAUGUCGACUAA, UGUUUUACAUGUUGUGUGA, UGUUUUACAUGUUU

UCUGA, UGUUUUACAUGUUUCCUA. Target sequences for *HNRNA2B1* siRNA were: GGAUCUGAUGGAUACGGAA, GGAUGGCUAUAUUGGGUA, ACCGAUAGGCAGUCUGGAA, GGUGGAAUUAAGGAAGAA.

**Protein purification.** GST-tagged Kap $\beta$ 2 WT protein was purified as described previously<sup>66</sup> with modifications. Briefly, recombinant protein was expressed in BL21 (DE3) *Escherichia coli* cells by induction with 1 mM isopropyl- $\beta$ -D-thiogalactoside overnight at 16 °C. Cells were lysed by sonication in buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, 2 mM EDTA, 2 mM DTT, 15% (v/v) glycerol with protease inhibitors and centrifuged. GST-Kap $\beta$ 2 was then purified using Glutathione Sepharose 4B protein purification resin and eluted in buffer described previously with 30 mM glutathione, adjusted to pH 7.5. GST was cleaved using TEV protease overnight at 4 °C or kept on the protein and Kap $\beta$ 2 was further purified by ion-exchange and size exclusion chromatography in buffer containing 20 mM HEPES pH 7.5, 110 mM potassium acetate, 2 mM magnesium acetate, 2 mM DTT, and 10% (v/v) glycerol. Purified proteins were flash-frozen and stored at –80 °C. GST-tagged hnRNA2 PY-NLS WT and N323Tfs\*36 peptides were purified similarly except no ion-exchange step was included.

GST-tagged hnRNA2 proteins were purified as described<sup>6</sup>. Briefly, WT and mutant hnRNA2 with N-terminal GST tag were over-expressed into *E. coli* BL21(DE3)-RIL (Invitrogen). Bacteria were grown at 37 °C until reaching an OD<sub>600</sub> of ~0.6 and expression was induced by addition of 1 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside (IPTG) for 15–18 h at 15 °C. Protein was purified over a Glutathione-Sepharose column (GE) according to manufacturer instructions. GST-hnRNA2 was eluted from the Glutathione Sepharose with 40 mM HEPES-NaOH, pH 7.4, 150 mM potassium chloride, 5% glycerol, and 20 mM reduced glutathione. Eluted proteins were centrifuged at 16,100 $\times$ g for 10 min at 4 °C to remove any aggregated material before being flash-frozen in liquid N<sub>2</sub> and stored at –80 °C. Before each experiment, protein was centrifuged at 16,100 $\times$ g for 10 min at 4 °C to remove any aggregated material. After centrifugation, the protein concentration in the supernatant was determined by Bradford assay (Bio-Rad) and these proteins were used for aggregation reactions. TEV protease was purified as described<sup>67</sup>. Briefly, His-tagged TEV were over-expressed into *E. coli* BL21(DE3)-RIL (Invitrogen). Bacteria were grown at 37 °C until reaching an OD<sub>600</sub> of ~0.7 and expression was induced by addition of 1 mM IPTG for 15–18 h at 15 °C. Protein was lysed in TEV Lysis Buffer (25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 25 mM imidazole, 10 mM  $\beta$ -mercaptoethanol, with protease inhibitors) and purified over Ni-NTA resin (Qiagen). His-tagged TEV protein were eluted in buffer containing 25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 300 mM imidazole, 10 mM  $\beta$ -mercaptoethanol. Protein were dialyzed into buffer containing 25 mM HEPES-NaOH, pH 7.0, 5 mM  $\beta$ -mercaptoethanol, 5% (v/v) glycerol, flash frozen, and stored at –80 °C.

**Fluorescence polarization.** Synthesized TAMRA-tagged hnRNA2 PY-NLS WT, N323Tfs, G328Afs, or G331Efs peptide were incubated with increasing amounts of purified Kap $\beta$ 2 WT in buffer containing 20 mM HEPES pH 7.5, 110 mM potassium acetate, 2 mM magnesium acetate, 2 mM DTT, and 10% (v/v) glycerol in Corning 96-well solid black polystyrene plates. Fluorescence polarization was measured with a Cytation 5 multi-mode plate reader (Biotek) using Gen5 software with 561-nm polarization cube. Analysis was performed using MATLAB. Briefly, fluorescence polarization was converted to anisotropy and  $K_d$  values were calculated by fitting the resulting curve to the equation

$$F_{sb} = \frac{K_d + L + R - \sqrt{(K_d + L + R)^2 - 4 * L * R}}{2 * L};$$

where  $F_{sb}$  is the fraction bound,  $L$  is the concentration of peptide, and  $R$  is the concentration of Kap $\beta$ 2. Each peptide was run in triplicate.

**Generation of *Drosophila* lines and *Drosophila* stocks.** The hnRNA2 WT and D290V *Drosophila* stocks have been previously published<sup>6</sup>. Flies carrying pUASTattB-hnRNA2 N323Tfs, G328Afs, or G331Efs transgenes were generated by a standard injection and  $\phi$ C31 integrase-mediated transgenesis technique (BestGene Inc.). GMR-GAL4 was used to express transgenes in eyes; MHC-GAL4 was used to express transgenes in muscle. All *Drosophila* stocks were maintained in a 25 °C incubator with a 12-h day/night cycle.

**Adult *Drosophila* muscle preparation and immunohistochemistry.** Adult flies were embedded in a drop of OCT compound (Sakura Finetek) on a slide glass, frozen with liquid nitrogen, and bisected sagittally by a razor blade. After fixing with 4% paraformaldehyde in PBS, hemithoraces were stained by Texas Red-X phalloidin (Invitrogen) and DAPI according to manufacturer's instructions. Stained hemi-thoraces were mounted in 80% glycerol and the musculature was examined by DMIRE2 (Leica). For hnRNA2 staining, hemithoraces were permeabilized with PBS containing 0.2% Triton X-100 and stained with anti-hnRNA2B1 antibody (EF-67, sc-53531; Santa Cruz Biotechnology) and Alexa-488-conjugated secondary antibody (Invitrogen). Stained muscle fibers were dissected and mounted in Fluoromount-G (Southern Biotech) and imaged with a Marianas confocal microscope (Zeiss).



**Solubility and biochemical analyses of adult *Drosophila* muscles and**

**C-terminally GFP-tagged hnRNPA2 in C2C12 cells.** Sequential extractions were performed to examine the solubility profile of hnRNPA2. Adult fly thoraces or cells transfected with the appropriate GFP-tagged constructs were lysed in cold RIPA buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, and 1 mM EDTA) and sonicated. Cell lysates were cleared by centrifugation at 100,000×g for 30 min at 4 °C to generate RIPA-soluble samples. To prevent carry-overs, the resulting pellets were washed twice with PBS and RIPA-insoluble pellets were then extracted with urea buffer (7 M urea, 2 M thiourea, 4% CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulphonate), 30 mM Tris, pH 8.5), sonicated, and centrifuged at 100,000×g for 30 min at 22 °C. Protease inhibitors were added to all buffers before use. Protein concentration was determined by the bicinchoninic acid method (Pierce) and proteins were resolved by NuPAGE Novex 4–12% Bis–Tris gels (Invitrogen). For Western blot analysis, thoraces of adult flies were prepared and ground in NuPAGE LDS sample buffer (NP0007, Invitrogen). Samples were then boiled for 5 min and analyzed by standard Western blotting methods provided by Odyssey system (LI-COR) with 4–12% NuPAGE Bis–Tris gels (Invitrogen).

**In vitro fibrillization assays.** hnRNPA2 (5 μM) fibrillization (100 μl reaction) was initiated by addition of 1 μl of 2 mg/ml TEV protease in A2 assembly buffer (40 mM HEPES–NaOH, pH 7.4, 150 mM KCl, 5% glycerol, 1 mM DTT, and 20 mM glutathione). The hnRNPA2 fibrillization reactions were incubated at 25 °C for 0–24 h with agitation at 1200 rpm in an Eppendorf Thermomixer. For sedimentation analysis, at indicated time points, fibrillization reactions were centrifuged at 16,100×g for 10 min at 4 °C. Supernatant and pellet fractions were then resolved by SDS–PAGE and stained with Coomassie Brilliant Blue, and the relative amount in each fraction was determined by densitometry in ImageJ (NIH). All fibrillization assays were performed in triplicate. For electron microscopy, fibrillization reactions (10 μl) were adsorbed onto glow-discharged 300-mesh Formvar/carbon coated copper grids (Electron Microscopy Sciences) and stained with 2% (w/v) aqueous uranyl acetate. Excess liquid was removed and grids were allowed to air dry. Samples were viewed on a JEOL 1010 transmission electron microscope.

**Data collection and statistical analysis.** The NCPR was collected using CIDER v1.7 (<http://pappulab.wustl.edu/CIDER/>). Exome sequencing data were analyzed on seqr (<https://seqr.broadinstitute.org/>). FACS data was collected using BD FACS-Canto II Flow Cytometer (Version 9.0). All flow cytometry data were analyzed by FlowJo (Version 10.6.1). Images were quantified using Cell Profiler (Broad Institute, Version 4.1.3) and Image J (NIH, Version 2.1.0). Fluorescence polarization analysis was performed using MATLAB (Version 8.4). Statistical analyses were performed using Prism 9 (GraphPad, Version 9.3.1) software. Statistical tests used for individual experiments are described in corresponding legends.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

Patient-related data, including genetic sequencing data not included in the manuscript or its supplements, were generated as part of clinical care and may be subject to patient confidentiality. All requests for raw and analyzed data and materials related to patients presented in this article will be reviewed by the respective institution to verify if the request is subject to any intellectual property or confidentiality obligations. Data requests for anonymized data are typically shared with qualified investigators after a material transfer agreement; such requests should be directed to corresponding author Carsten G. Bönnemann. All other data generated or analyzed during this study are included in this published article (and its Supplementary Information/Source Data file).

Received: 28 April 2021; Accepted: 25 March 2022;

Published online: 28 April 2022

**References**

1. Tauber, D., Tauber, G. & Parker, R. Mechanisms and regulation of RNA condensation in RNP granule formation. *Trends Biochem. Sci.* **45**, 764–778 (2020).
2. Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *J. Biol. Chem.* **294**, 7115–7127 (2019).
3. Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 215–235 (2021).
4. Kabashi, E. et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.* **40**, 572–574 (2008).
5. Sreedharan, J. et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319**, 1668–1672 (2008).
6. Kim, H. J. et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
7. Beijer, D. et al. Characterization of HNRNPA1 mutations defines diversity in pathogenic mechanisms and clinical presentation. *JCI Insight* **6**, e148363 (2021).
8. Naruse, H. et al. Molecular epidemiological study of familial amyotrophic lateral sclerosis in Japanese population by whole-exome sequencing and identification of novel HNRNPA1 mutation. *Neurobiol. Aging* **61**, 255 e9–255 e16 (2018).
9. Liu, Q. et al. Whole-exome sequencing identifies a missense mutation in hnRNPA1 in a family with flail arm ALS. *Neurology* **87**, 1763–1769 (2016).
10. Qi, X. et al. Familial early-onset paget's disease of bone associated with a novel hnRNPA2B1 mutation. *Calcif. Tissue Int.* **101**, 159–169 (2017).
11. Berardo, A. et al. HNRNPDL-related muscular dystrophy: expanding the clinical, morphological and MRI phenotypes. *J. Neurol.* **266**, 2524–2534 (2019).
12. Sun, Y. et al. Limb girdle muscular dystrophy D3 HNRNPDL related in a Chinese family with distal muscle weakness caused by a mutation in the prion-like domain. *J. Neurol.* **266**, 498–506 (2019).
13. Vieira, N. M. et al. A defect in the RNA-processing protein HNRPDL causes limb-girdle muscular dystrophy 1G (LGMD1G). *Hum. Mol. Genet.* **23**, 4103–4110 (2014).
14. Straub, V., Murphy, A. & Udd, B., group, L.w.s. 229th ENMC international workshop: limb girdle muscular dystrophies—nomenclature and reformed classification Naarden, the Netherlands, 17–19 March 2017. *Neuromuscul. Disord.* **28**, 702–710 (2018).
15. Kwiatkowski, T. J. Jr. et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205–1208 (2009).
16. Vance, C. et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323**, 1208–1211 (2009).
17. Klar, J. et al. Welander distal myopathy caused by an ancient founder mutation in TIA1 associated with perturbed splicing. *Hum. Mutat.* **34**, 572–577 (2013).
18. Mackenzie, I. R. et al. TIA1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron* **95**, 808–816.e9 (2017).
19. Benatar, M. et al. Motor neuron involvement in multisystem proteinopathy: implications for ALS. *Neurology* **80**, 1874–1880 (2013).
20. Mathieu, C., Pappu, R. V. & Taylor, J. P. Beyond aggregation: pathological phase transitions in neurodegenerative disease. *Science* **370**, 56–60 (2020).
21. Nedelsky, N. B. & Taylor, J. P. Pathological phase transitions in ALS-FTD impair dynamic RNA-protein granules. *RNA* **28**, 97–113 (2022).
22. Ryan, V. H. et al. Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Mol. Cell* **69**, 465–479.e7 (2018).
23. Pesiridis, G. S., Lee, V. M. & Trojanowski, J. Q. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **18**, R156–R162 (2009).
24. Shang, Y. & Huang, E. J. Mechanisms of FUS mutations in familial amyotrophic lateral sclerosis. *Brain Res.* **1647**, 65–78 (2016).
25. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
26. Tada, M. et al. Matrin 3 Is a component of neuronal cytoplasmic inclusions of motor neurons in sporadic amyotrophic lateral sclerosis. *Am. J. Pathol.* **188**, 507–514 (2018).
27. Batlle, C. et al. hnRNPD phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation. *Cell Rep.* **30**, 1117–1128.e5 (2020).
28. Guo, W. et al. An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat. Struct. Mol. Biol.* **18**, 822–830 (2011).
29. Trollet, C. et al. Oculopharyngeal muscular dystrophy. In *GeneReviews*(R) (eds. Adam, M. P. et al.) (University of Washington, Seattle, WA, 1993).
30. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
31. Siomi, H. & Dreyfuss, G. A nuclear localization domain in the hnRNP A1 protein. *J. Cell Biol.* **129**, 551–560 (1995).
32. Michael, W. M., Choi, M. & Dreyfuss, G. A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* **83**, 415–422 (1995).
33. Pollard, V. W. et al. A novel receptor-mediated nuclear protein import pathway. *Cell* **86**, 985–994 (1996).
34. Lee, B. J. et al. Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* **126**, 543–558 (2006).



35. Allemand, E. et al. Regulation of heterogeneous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. *Proc. Natl Acad. Sci. USA* **102**, 3605–3610 (2005).
36. Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **6**, 365–387 (2011).
37. Zhang, Z. C. & Chook, Y. M. Structural and energetic basis of ALS-causing mutations in the atypical proline-tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). *Proc. Natl Acad. Sci. USA* **109**, 12017–12021 (2012).
38. Kalderon, D., Richardson, W. D., Markham, A. F. & Smith, A. E. Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* **311**, 33–38 (1984).
39. Duffy, J. B. GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. *Genesis* **34**, 1–15 (2002).
40. Freibaum, B. D. et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129–133 (2015).
41. Li, S. et al. Genetic interaction of hnRNP A2B1 and DNAJB6 in a *Drosophila* model of multisystem proteinopathy. *Hum. Mol. Genet.* **25**, 936–950 (2016).
42. Prayson, R. A. & Cohen, M. L. Ubiquitin immunostaining and inclusion body myositis: study of 30 patients with inclusion body myositis. *Hum. Pathol.* **28**, 887–892 (1997).
43. Martinsson, T. et al. Autosomal dominant myopathy: missense mutation (Glu-706→Lys) in the myosin heavy chain IIa gene. *Proc. Natl Acad. Sci. USA* **97**, 14614–14619 (2000).
44. Izumi, R. et al. Isolated inclusion body myopathy caused by a multisystem proteinopathy-linked hnRNP A1 mutation. *Neurol. Genet.* **1**, e23 (2015).
45. Hackman, P. et al. Weller distal myopathy is caused by a mutation in the RNA-binding protein TIA1. *Ann. Neurol.* **73**, 500–509 (2013).
46. Brais, B. et al. Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat. Genet.* **18**, 164–167 (1998).
47. Seelen, M. et al. No mutations in hnRNP A1 and hnRNP A2B1 in Dutch patients with amyotrophic lateral sclerosis, frontotemporal dementia, and inclusion body myopathy. *Neurobiol. Aging* **35**, 1956e9–1956e11 (2014).
48. Le Ber, I. et al. hnRNP A2B1 and hnRNP A1 mutations are rare in patients with “multisystem proteinopathy” and frontotemporal lobar degeneration phenotypes. *Neurobiol. Aging* **35**, 934 e5–6 (2014).
49. Calini, D. et al. Analysis of hnRNP A1, A2/B1, and A3 genes in patients with amyotrophic lateral sclerosis. *Neurobiol. Aging* **34**, 2695 e11–2 (2013).
50. Gasset-Rosa, F. et al. Cytoplasmic TDP-43 de-mixing independent of stress granules drives inhibition of nuclear import, loss of nuclear TDP-43, and cell death. *Neuron* **102**, 339–357.e7 (2019).
51. Yoshizawa, T. et al. Nuclear import receptor inhibits phase separation of FUS through binding to multiple sites. *Cell* **173**, 693–705.e22 (2018).
52. Guo, L. et al. Nuclear-import receptors reverse aberrant phase transitions of RNA-binding proteins with prion-like domains. *Cell* **173**, 677–692.e20 (2018).
53. Lee, K. H. et al. C9orf72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. *Cell* **167**, 774–788.e17 (2016).
54. Boeynaems, S. et al. Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Mol. Cell* **65**, 1044–1055.e5 (2017).
55. Gwon, Y. et al. Ubiquitination of G3BP1 mediates stress granule disassembly in a context-specific manner. *Science* **372**, eabf6548 (2021).
56. Van Deerlin, V. M. et al. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *Lancet Neurol.* **7**, 409–416 (2008).
57. Vogler, T. O. et al. TDP-43 and RNA form amyloid-like myo-granules in regenerating muscle. *Nature* **563**, 508–513 (2018).
58. Bergeron, D., Pal, G., Beaulieu, Y. B., Chabot, B. & Bachand, F. Regulated intron retention and nuclear pre-mRNA decay contribute to PABPN1 autoregulation. *Mol. Cell Biol.* **35**, 2503–2517 (2015).
59. Tanaka, A. J. et al. Mutations in SPATA5 are associated with microcephaly, intellectual disability, seizures, and hearing loss. *Am. J. Hum. Genet.* **97**, 457–464 (2015).
60. Topf, A. et al. Sequential targeted exome sequencing of 1001 patients affected by unexplained limb-girdle weakness. *Genet. Med.* **22**, 1478–1488 (2020).
61. Ashton, J. J. et al. Identification of variants in genes associated with single-gene inflammatory bowel disease by whole-exome sequencing. *Inflamm. Bowel Dis.* **22**, 2317–2327 (2016).
62. Schartner, V. et al. Dihydropyridine receptor (DHPR, CACNA1S) congenital myopathy. *Acta Neuropathol.* **133**, 517–533 (2017).
63. Macchione, F. et al. Multiple acyl-CoA dehydrogenase deficiency in elderly carriers. *J. Neurol.* **267**, 1414–1419 (2020).
64. French, C. E. et al. Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Med.* **45**, 627–636 (2019).
65. Iida, A. et al. A novel PAK3 pathogenic variant identified in two siblings from a Japanese family with X-linked intellectual disability: case report and review of the literature. *Cold Spring Harb Mol Case Stud.* **5**, a003988 (2019).
66. Zhang, Z. C., Satterly, N., Fontoura, B. M. & Chook, Y. M. Evolutionary development of redundant nuclear localization signals in the mRNA export factor NXF1. *Mol. Biol. Cell* **22**, 4657–4668 (2011).
67. Cupo, R. R. & Shorter, J. Expression and purification of recombinant Skd3 (human ClpB) protein and tobacco Etch virus (TEV) protease from *Escherichia coli*. *Bio Protoc.* **10**, e3858 (2020).

## Acknowledgements

We thank the patients and their families for their participation. We thank Christopher Mendoza and Gilberto (“Mike”) Averion for their help in clinic. We also thank the NIH Intramural Sequencing Center staff and Daniel MacArthur and Fengmei Zhao (Analytic and Translational Genetics Unit at Massachusetts General Hospital in collaboration with the Broad Institute) for help with exome analysis. We thank Darren Chambers and Lucy Feng for technical assistance in processing muscle pathology specimens. We thank Natalia Nedelsky for editorial assistance. Work in C.G.B.’s laboratory is supported by intramural funds from the NIH National Institute of Neurological Disorders and Stroke. Exome sequencing was in part funded through the Clinical Center Genomics Opportunity, which is sponsored by the National Human Genome Research Institute, the NIH Deputy Director for Intramural Research, and the NIH Clinical Center. J.P.T. is supported by the Howard Hughes Medical Institute, R35NS097974, and the ALS Association (18-1IA-419). This study was supported in part by an Intramural Research Grant (2-5 and 29-4 to I.N.; 2-5 and 30-9 to A.I.) for Neurological and Psychiatric Disorders of NCNP and by AMED under grants 20ek0109490h0001 and JP19ek0109285h0003 (to I.N.) and Joint Usage and Joint Research Programs, the Institute of Advanced Medical Sciences, Tokushima University (2020, 2A19 to A.I.). Work in F.M.’s group was supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London and the MRC Centre for Neuromuscular Diseases Biobank, and by the HSS England Diagnostic and Advisory Service for Congenital Myopathies and Congenital Muscular Dystrophies in London, UK for their financial support to the DNC Muscle Pathology Service. F.M. was also funded by the European Community’s Seventh Framework Program (FP7/2007–2013) under grant agreement 2012-305121 “Integrated European-omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)”. E.P. and L.B. are members of the European Reference Network for Neuromuscular Diseases—Project ID 870177 and acknowledge support from the Telethon Network of Genetic Biobanks (GTB12001D) and EuroBioBank network. The work in J.L.’s group is supported by the France Génomique National infrastructure and by the Fondation Maladies Rares within the frame of the “Myo-capture” sequencing project, and by Association Française contre les Myopathies (22734). C.E.F. and F.L.R. are funded by Cambridge NIHR Biomedical Research Centre and the Rosetree Foundation. L.G. was supported by an Ellison Medical Foundation/American Federation for Aging Research fellowship, Alzheimer’s Association Research fellowship, and a Target ALS Springboard Fellowship. C.M.F. was supported by NIH grants T32GM008275 and F31NS111870. A.F.F. was supported by NIH grants T32AG00255 and F31NS087676. J.S. was supported by Target ALS, Packard Foundation for ALS Research, The ALS Association, The G. Harold and Leila Y. Mathers Charitable Foundation, and NIH grant R01GM099836. MYOSEQ was funded by Sanofi Genzyme, Ultragenyx, LGMD21 Research Fund, Samantha J. Brazzo Foundation, LGMD2D Foundation and Kurt+Peter Foundation, Muscular Dystrophy UK, and Coalition to Cure Calpain 3. Analysis was provided by the Broad Institute Center for Mendelian Genomics and was funded by the National Human Genome Research Institute, the National Eye Institute, and the National Heart, Lung, and Blood Institute grant UM1 HG008900, and in part by National Human Genome Research Institute grant R01HG009141. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Author contributions

H.J.K., P.M., and S.D. designed and performed laboratory experiments, evaluated patients, analyzed data, and drafted the manuscript. K.O., L.G. and C.M.F. performed laboratory experiments, analyzed data, and drafted the manuscript. X.L., N.F., S.R.H., A.R.F., M.O., A.S., A.T., P.M., G.A., R.P., D.G.O., N.D., R.B., M.G. A.T., I.T.Z., L.B., T.E.L., A.K., M.S., A.K., S.M., P.Ma., Y.P., E.F., A.M., S.E., R.U.G., L.B., C.B., E.P., L.S., C.E.F., F.L.R., T.H., S.Q.-R., J.B., I.N., T.S., T.E., V.S., N.B.R., J.L., F.M., I.N., and M.A.T. provided and interpreted clinical and genetic information, crucial samples and material, and/or analyzed data. A.F.F., Y.H., and M.C. performed laboratory experiments, and/or analyzed data. J.S., C.G.B., and J.P.T. supervised the overall study, interpreted data, drafted, and revised the manuscript. All authors reviewed and edited the manuscript.

## Competing interests

J.P.T. is a consultant for 5AM and Third Rock Ventures. J.S. is a consultant for Dewpoint Therapeutics, ADRx, Neumora, Vivid Science, and Korro Bio. M.A.T. is the founder and CEO of Exerkine and CSO for Cora Therapeutics. The remaining authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-30015-1>.

**Correspondence** and requests for materials should be addressed to Carsten G. Bönnemann or J. Paul Taylor.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

<sup>1</sup>Department of Cell and Molecular Biology, St. Jude Children's Research Hospital, Memphis, TN, United States. <sup>2</sup>National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, United States. <sup>3</sup>Department of Biochemistry & Biophysics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, United States. <sup>4</sup>Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA, United States. <sup>5</sup>Département Médecine Translationnelle et Neurogénétique, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Institut National de la Santé et de la Recherche Médicale U1258, Centre National de la Recherche Scientifique UMR7104, Université de Strasbourg, Illkirch, France. <sup>6</sup>Wessex Clinical Genetics Services, Princess Anne Hospital, Academic Unit of Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, England. <sup>7</sup>Wessex Neurological Centre, University Hospital Southampton, Southampton, UK. <sup>8</sup>Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry (NCNP), 4-1-1 Ogawahigashi, Kodaira, Tokyo 187-8502, Japan. <sup>9</sup>Medical Genome Center, NCNP, Kodaira, Tokyo, Japan. <sup>10</sup>Department of Neurology, Niigata City General Hospital, Niigata, Japan. <sup>11</sup>The Dubowitz Neuromuscular Centre, 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. <sup>12</sup>Department of Paediatric Neurology, Cambridge University Hospital NHS Trust, Addenbrookes Hospital, Cambridge CB2 0QQ, UK. <sup>13</sup>Division of Neuropathology, University College London Hospitals NHS Foundation Trust National Hospital for Neurology and Neurosurgery London, UK and Division of Neuropathology, UCL Institute of Neurology, Dubowitz Neuromuscular Centre, London, UK. <sup>14</sup>Department of Histopathology Box 235, Level 5 John Bonnett Clinical Laboratories Addenbrooke's Hospital, Cambridge, UK. <sup>15</sup>Institute of Medical Genetics and Applied Genomics, University of Tuebingen, Tuebingen, Germany. <sup>16</sup>John Walton Muscular Dystrophy Research Centre, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. <sup>17</sup>Division of Neuromuscular & Neurometabolic Disorders, Department of Pediatrics, McMaster University, Hamilton Health Sciences Centre, Hamilton, ON, Canada. <sup>18</sup>Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, United States. <sup>19</sup>Division of Neuropaediatrics, Development and Rehabilitation, Department of Pediatrics, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. <sup>20</sup>Pediatric Neurology, University Children's Hospital Basel, University of Basel, Basel, Switzerland. <sup>21</sup>Department of Neurometabolism, University Hospital of Nantes, Nantes, France. <sup>22</sup>CHU Nantes, Service de génétique médicale, Centre de Référence des Maladies Neuromusculaires AOC, 44000 Nantes, France. <sup>23</sup>Université de Nantes, CNRS, INSERM, l'institut du thorax, 44000 Nantes, France. <sup>24</sup>Service d'anatomopathologie, CHU Brest and EA 4685 LIEN, Université de Bretagne Occidentale, Brest, France. <sup>25</sup>CHU de Nantes, Centre de Référence des Maladies Neuromusculaires, Filnemus, Euro-NMD, Hôtel-Dieu, Nantes, France. <sup>26</sup>Etablissement de Santé pour Enfants et Adolescents de la région Nantaise, Nantes, France. <sup>27</sup>Human Genetics and Genomic Medicine, Faculty of Medicine, University of Southampton, Southampton, UK. <sup>28</sup>Department of Neurosciences, DNS, University of Padova, Padova, Italy. <sup>29</sup>Clinical Genetics Unit, Department of Women and Children's Health, University of Padova, IRP Città della Speranza, Padova, Italy. <sup>30</sup>Clinical Genetics Unit, Department of Women and Children's Health, CIR-Myo Myology Center, University of Padova, IRP Città della Speranza, Padova, Italy. <sup>31</sup>Department of Paediatrics, University of Cambridge, Cambridge, UK. <sup>32</sup>Institute of Neurology, Psychiatry and Narcology of NAMS of Ukraine, Kharkiv, Ukraine. <sup>33</sup>Cambridge Institute of Medical Research, University of Cambridge, Cambridge, UK. <sup>34</sup>Neuromuscular Unit, Pediatric Neurology and ICU Department, Raymond Poincaré Hospital (UVSQ), AP-HP Université Paris-Saclay, Garches, France. <sup>35</sup>Sorbonne Université, INSERM, Centre of Research in Myology, UMRS974 Paris, France. <sup>36</sup>APHP, Centre de Référence des Maladies Neuromusculaires Nord/Est/Ile de France, Institut de Myologie, Sorbonne Université, Hôpital Pitié-Salpêtrière, Paris, France. <sup>37</sup>Unité de Morphologie Neuromusculaire, Institut de Myologie, Sorbonne Université, Hôpital Pitié-Salpêtrière, Paris, France. <sup>38</sup>Howard Hughes Medical Institute, Chevy Chase, MD, United States. <sup>39</sup>These authors contributed equally: Hong Joo Kim, Payam Mohassel.

<sup>✉</sup>email: [carsten.bonnemann@nih.gov](mailto:carsten.bonnemann@nih.gov); [jpaul.taylor@stjude.org](mailto:jpaul.taylor@stjude.org)

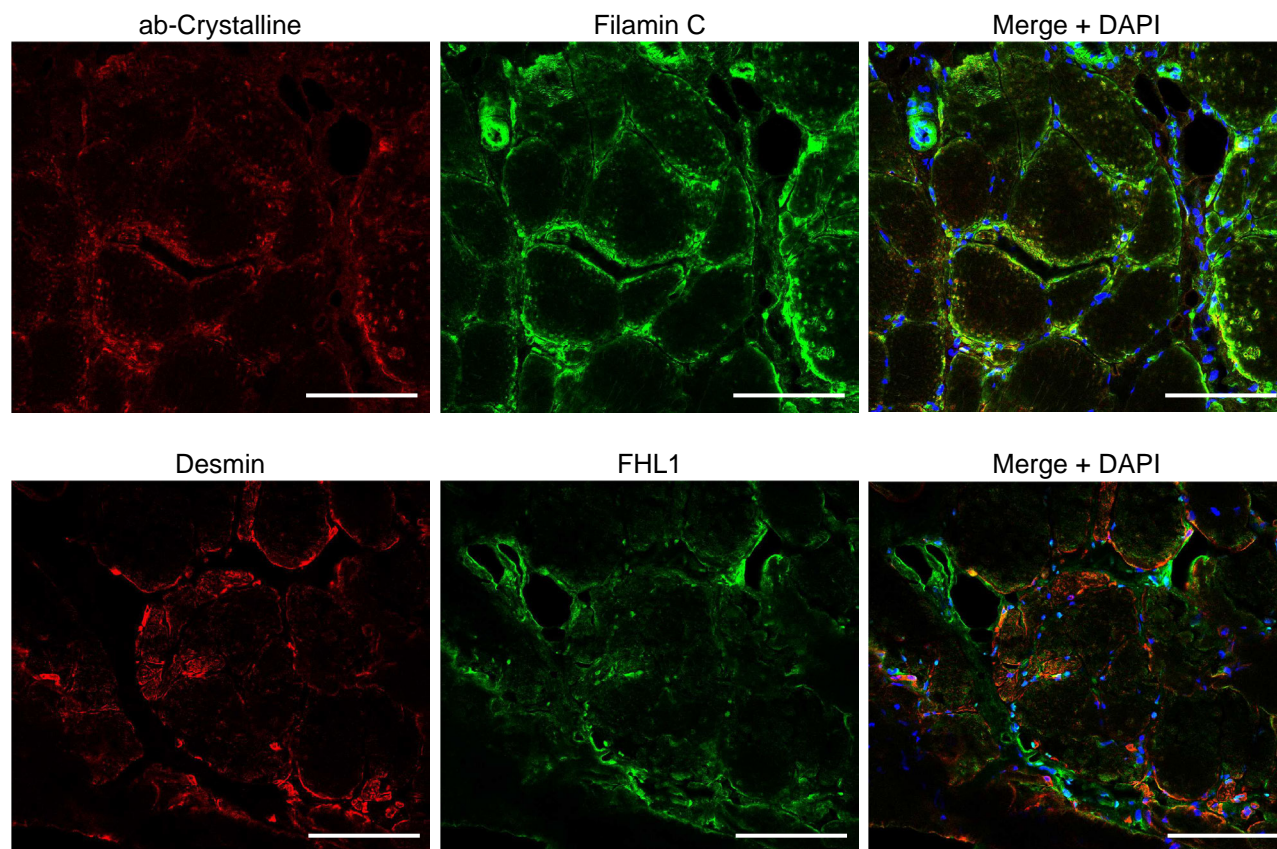
## Supplementary Information for

### Heterozygous frameshift variants in *HNRNPA2B1* cause early-onset oculopharyngeal muscular dystrophy

Hong Joo Kim<sup>#</sup>, Payam Mohassel<sup>#</sup>, Sandra Donkervoort, Lin Guo, Kevin O'Donovan, Maura Coughlin, Xaviere Lornage, Nicola Foulds, Simon R. Hammans, A. Reghan Foley, Charlotte M. Fare, Alice F. Ford, Masashi Ogasawara, Aki Sato, Aritoshi Iida, Pinki Munot, Gautam Ambegaonkar, Rahul Phadke, Dominic G O'Donovan, Rebecca Buchert, Mona Grimm, Ana Töpf, Irina T. Zaharieva, Lauren Brady, Ying Hu, Thomas E. Lloyd, Andrea Klein, Maja Steinlin, Alice Kuster, Sandra Mercier, Pascale Marcorelles, Yann Péréon, Emmanuelle Fleurence, Adnan Manzur, Sarah Ennis, Rosanna Upstill-Goddard, Luca Bello, Cinzia Bertolin, Elena Pegoraro, Leonardo Salviati, Courtney E. French, Andriy Shatillo, F Lucy Raymond, Tobias B. Haack, Susana Quijano-Roy, Johann Böhm, Isabelle Nelson, Tanya Stojkovic, Teresinha Evangelista, Volker Straub, Norma B. Romero, Jocelyn Laporte, Francesco Muntoni, Ichizo Nishino, Mark A. Tarnopolsky, James Shorter, Carsten G. Bönnemann\* and J. Paul Taylor\*

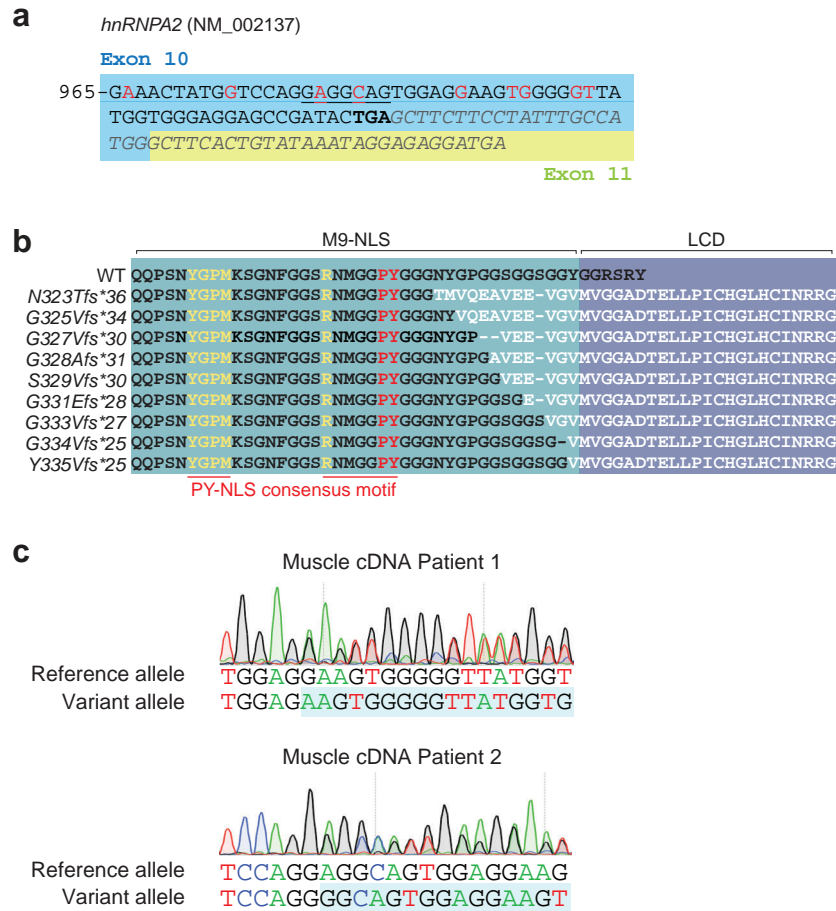
<sup>#</sup> These authors contributed equally: Hong Joo Kim, Payam Mohassel

\*email: [carsten.bonnemann@nih.gov](mailto:carsten.bonnemann@nih.gov); [jpaul.taylor@stjude.org](mailto:jpaul.taylor@stjude.org)

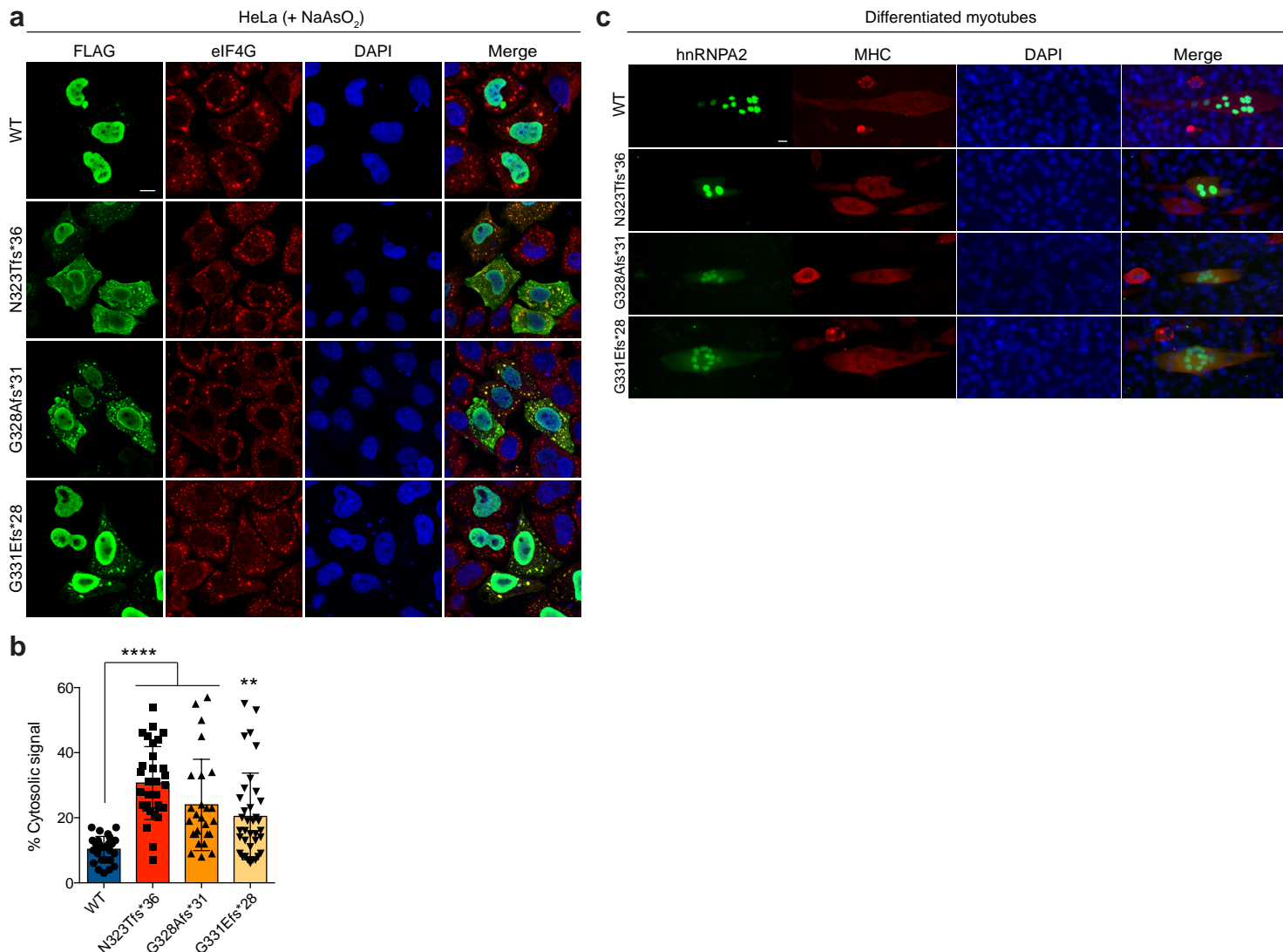


**Supplementary Fig. 1. Cytoplasmic myofiber inclusions do not contain myofibrillar proteins.** Immunostaining of patient 1 muscle biopsy with  $\alpha\beta$ -crystallin, filamin C, desmin and FHL1 showing absence of cytoplasmic or perinuclear inclusions or protein aggregates that contain these Z-disk associated myofibrillar proteins. Scale bar, 100  $\mu$ m. The micrographs shown are representative images of a single diagnostic muscle biopsy in each indicated patient. No independent replicates were performed.

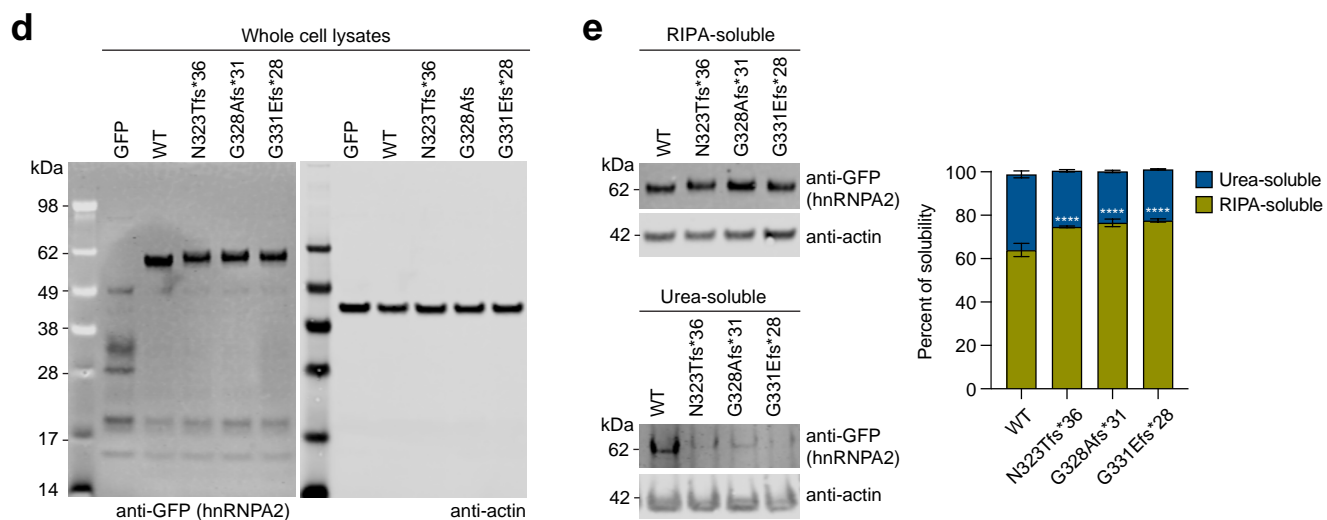
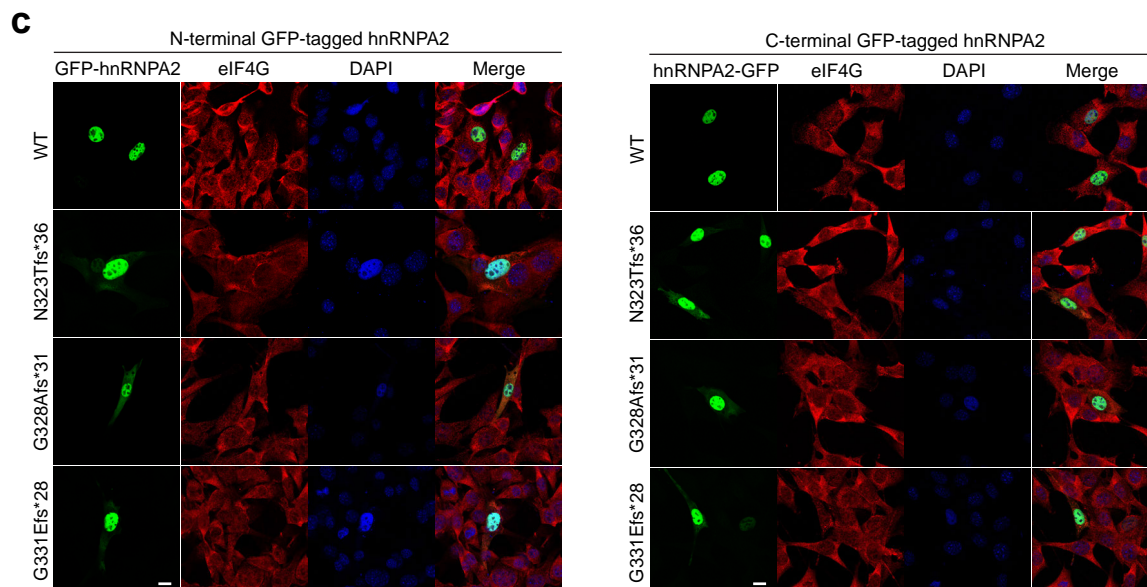
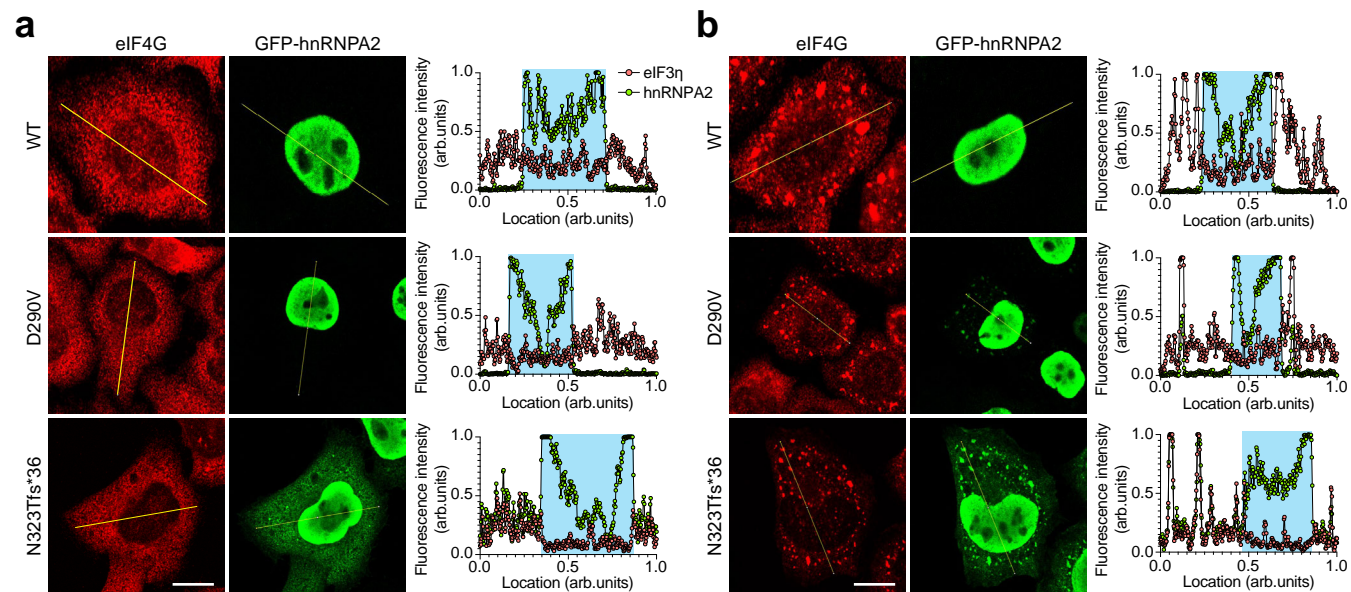




**Supplementary Fig. 2. Frameshift *HNRNPA2B1* variant transcripts are stable in muscle tissue and cause rearrangement of amino acid sequence in the C terminus.** **a**, cDNA sequence of *HNRNPA2* (NM\_002137) in exon 10 and exon 11. Red nucleotides indicate nucleotides deleted in patients; underlined nucleotides indicate nucleotides deleted in patient 8. The original stop codon (TGA) is indicated in bold. **b**, Amino acid sequences in WT and each frameshift mutant. The consensus PY-NLS motif in M9-NLS is underlined in red on bottom row. **c**, RNA was extracted from muscle tissue of patients 1 (c.992delG) and 2 (c.981delA). After reverse transcriptase reaction, the cDNA was amplified by PCR and sequenced. The intensity of the chromatogram peaks corresponding to the reference and frameshift alleles appear nearly equal, suggesting little to no degradation of the frameshift transcript in muscle tissue.

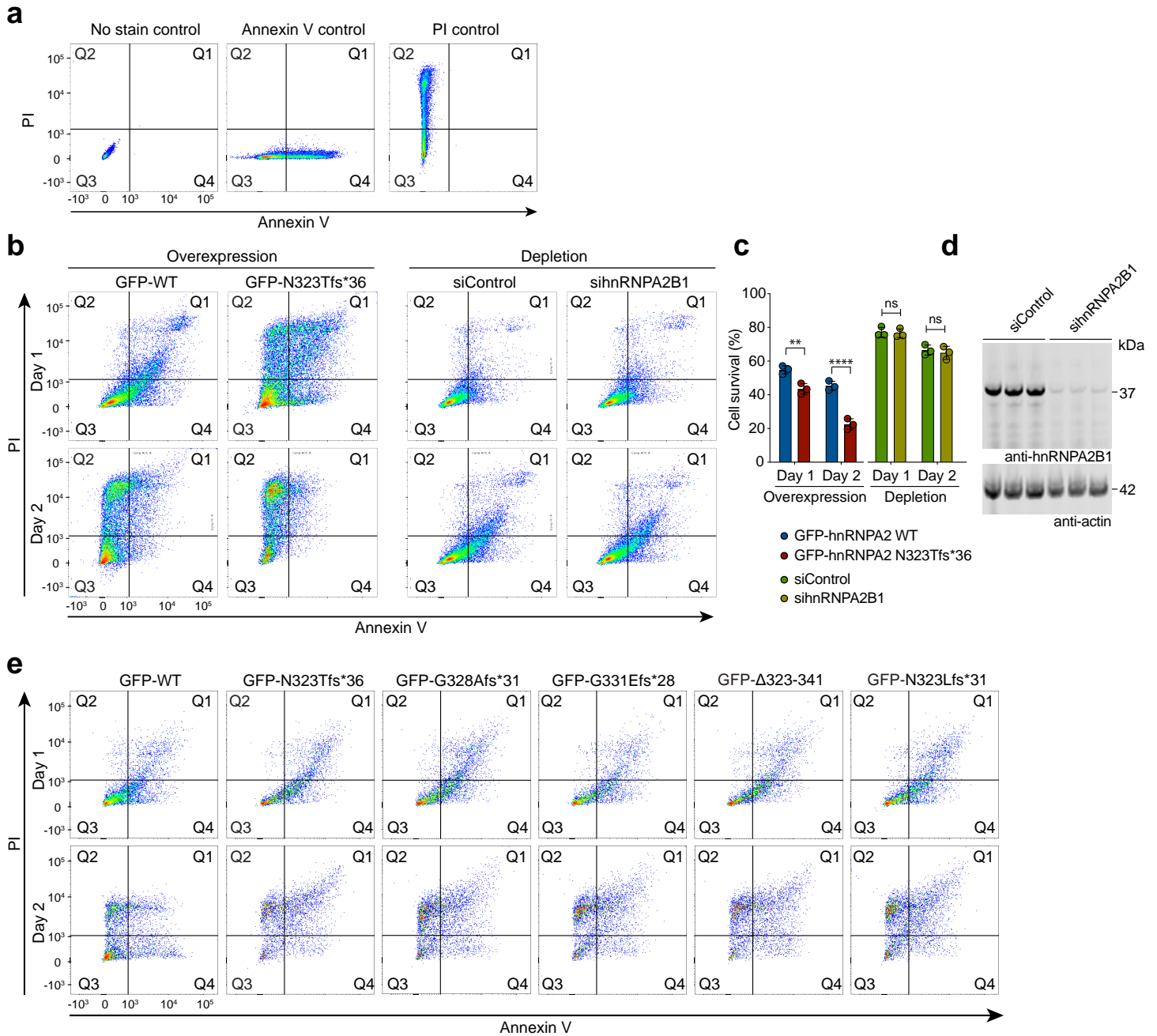


**Supplementary Fig. 3. Frameshift variants enhance hnRNP A2B1 cytoplasmic localization and recruitment to RNP granules.** **a**, Intracellular localization of FLAG-tagged hnRNP A2 WT, N323Tfs\*36, G328Afs\*31, and G331Efs\*28 mutants upon 30 min 0.5 mM NaAsO<sub>2</sub> stress in HeLa cells. Representative images from three independent experiments are shown. Scale bar, 10  $\mu$ m. **b**, Quantification of hnRNP A2 cytosolic signal intensity in cells as shown in (a). An interleaved scatter plot with individual data points is shown; error bars represent mean  $\pm$  s.d. For WT, N323Tfs\*36, G328Afs\*31, and G331Efs\*28 mutants,  $n = 27, 31, 26,$  and  $36$  cells, respectively. \*\*\*\* $P < 0.0001$  (WT vs. N323Tfs\*36 and G328Afs\*31), \*\* $P = 0.0022$  (WT vs. G331Efs\*28) by two-way ANOVA with Dunnett's multiple comparisons test. **c**, Intracellular localization of hnRNP A2 WT, N323Tfs\*36, G328Afs\*31, and G331Efs\*28 mutants in differentiated C2C12 myotubes. Representative images from three independent experiments are shown. Scale bar, 10  $\mu$ m.



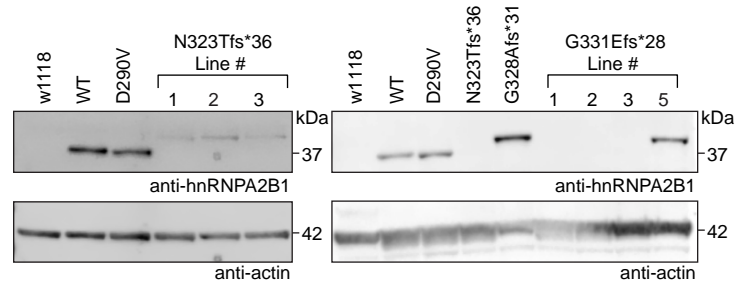


**Supplementary Fig. 4. Frameshift variants cause diffuse cytoplasmic distribution, showing divergence from MSP-associated D290V variant. a-b,** Fluorescent intensity across unstressed (**a**) or sodium arsenite-stressed (**b**) cells expressing GFP-tagged hnRNPA2 WT, D290V or N323Tfs\*36. Representative cells from three independent experiments were chosen for analysis. Graphs represent fluorescent intensity across the line in images; blue shaded area denotes nucleus. Scale bar, 10  $\mu$ m. **c,** Intracellular localization of N-terminal (left) or C-terminal (right) tagged GFP-hnRNPA2 WT, N323Tfs\*36, G328Afs\*31, and G331Efs\*28 in undifferentiated C2C12 cells. eIF4G was used as a cytoplasmic marker. Representative images from three independent experiments are shown. Scale bar, 10  $\mu$ m. **d,** Representative images from three independent immunoblot analyses of C2C12 cells expressing C-terminal tagged hnRNPA2 constructs. Whole cell lysates were immunoblotted with actin (loading control) and GFP (hnRNPA2) antibodies. **e,** C2C12 cells expressing C-terminal tagged hnRNPA2 constructs were subjected to sequential extractions to examine solubility profiles. Quantification (right) of RIPA- and urea-soluble fractions. Data represents mean  $\pm$  s.d. n = 3 independent experiments. \*\*\*\* $P < 0.0001$  by one-way ANOVA with Dunnett's multiple comparisons test.

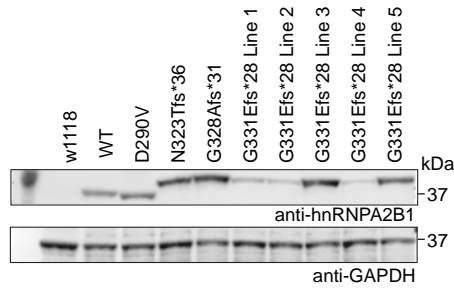


**Supplementary Fig. 5. Alteration or absence of the wild type hnRNPA2 C-terminal sequence causes apoptotic cell death in differentiating cells by a gain of function mechanism.** **a**, Gate strategy for analysis. Gates were drawn on controls of cells stained for only annexin V or PI to separate from the negative populations. **b**, Representative scatter plots of GFP-positive C2C12 cells (overexpression) and whole cell population (depletion) stained with propidium iodide (PI) and annexin V, one (top) or two (bottom) days after change to differentiation media. **c**, Quantification of cells that were negative for both PI and annexin V. Values represent means  $\pm$  s.e.m. ( $n = 3$  independent experiments).  $**P = 0.0014$  and  $****P < 0.0001$  by two-way ANOVA with Tukey's multiple comparisons test. ns; Not significant. **d**, Immunoblot showing knockdown of *HNRNPA2B1* by sihnRNPA2B1. Three independently prepared samples were loaded on the gel. **e**, Representative scatter plots of GFP-positive C2C12 cells stained with propidium iodide (PI) and annexin V, one (top) or two (bottom) days after change to differentiation media.

**a hnRNP A2 expression driven by GMR-Gal4**

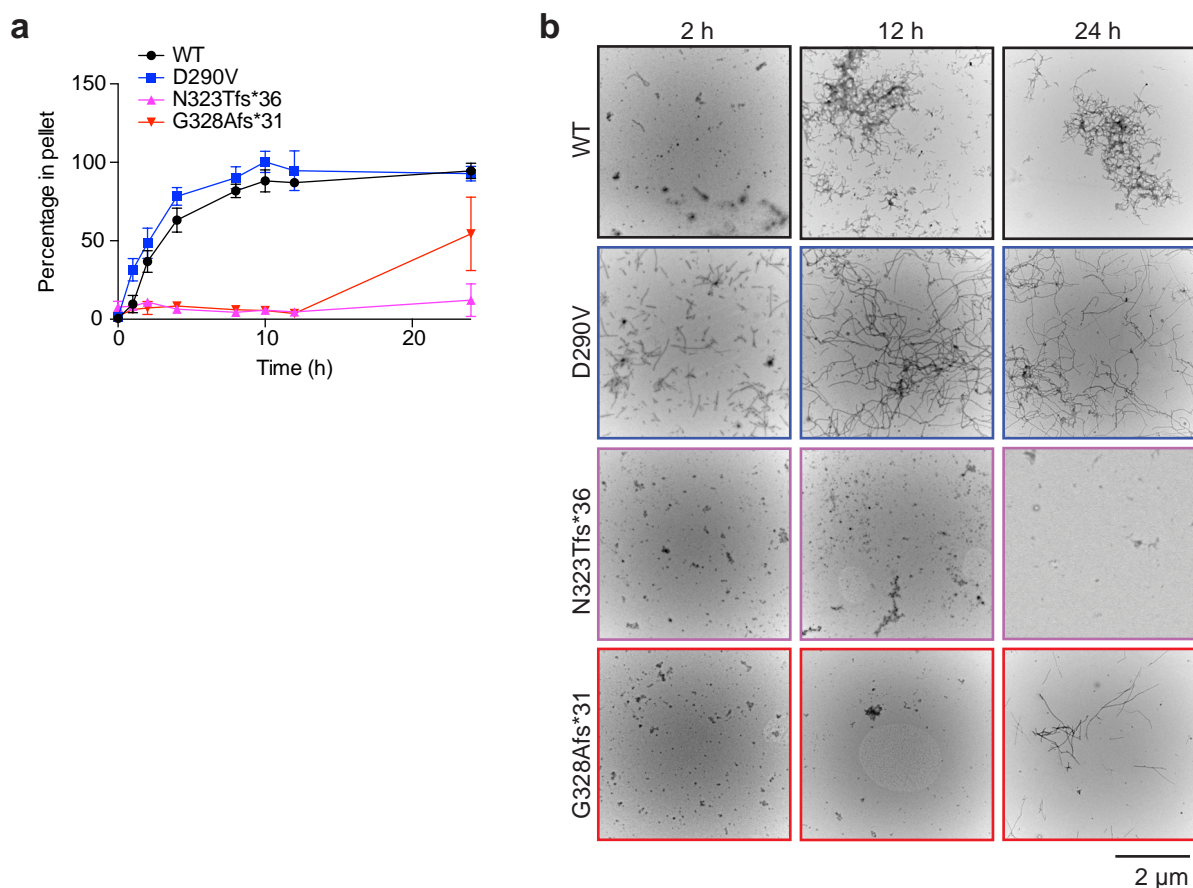


**b hnRNP A2 expression driven by MHC-Gal4**



**Supplementary Fig. 6. Expression of frameshift variants in a *Drosophila* model. a,** Representative immunoblots from three independent experiments showing hnRNP A2 expression driven by GMR-GAL4 in transgenic flies. Lysates were prepared from heads of adult flies and analyzed by Western blot using an antibody against hnRNP A2B1. Actin was used as a loading control. **b,** Representative immunoblots from three independent experiments showing hnRNP A2 expression driven by MHC-GAL4 in transgenic flies. Lysates were prepared from thoraces of adult flies and analyzed by Western blot using an antibody against hnRNP A2B1. GAPDH was used as a loading control.





**Supplementary Fig. 7. Frameshift variants reduce the intrinsic ability of hnRNPA2 to fibrillize.** **a**, Full-length purified hnRNPA2 WT, D290V, N323Tfs\*36, or G328Afs\*31 (5  $\mu$ M) were incubated at 25°C with agitation for 0-24 h. The amount of aggregated hnRNPA2 was determined by sedimentation analysis at indicated times. Values represent means  $\pm$  s.e.m. (n = 3 independent experiments). Notably, when lower protein concentrations were used (3  $\mu$ M), the frameshift variants did not assemble fibrils, whereas D290V protein assembled at a faster rate than WT, as previously reported<sup>1</sup>. **b**, Representative electron micrographs from three independent experiments showing hnRNPA2 fibrillization reactions after 2, 12 and 24 h at 25°C. Fibrils were scarce for WT after 2 h, abundant for D290V after 2 h, and absent for N323Tfs\*36 and G328fs\*31 after 2 h. Fibrils remained absent after 12 h for hnRNPA2 N323Tfs\*36 and G328fs\*31, but were abundant for D290V and WT. After 24 h, all hnRNPA2 variants formed fibrils, but fibrils were less abundant for N323Tfs\*36 and G328fs\*31. Scale bar, 2  $\mu$ m.

**Supplementary Table 1:** *HNRNPA2B1* variants identified in the cohort. The ACMG variant categories shown are based on data presented in this manuscript. All variants were initially classified as having “uncertain significance.” Those variants with confirmed *de novo* status were subsequently re-classified as “likely pathogenic.” In those cases where functional data further suggested pathogenicity, the variants were upgraded to a classification of “pathogenic”.

Genomic location <sup>a</sup>	Nucleotide Change <sup>b/c</sup>	Predicted Amino Acid Change	gnomAD AF	ACMG Variant Classification	Family/Patient
7:26232169	c.992delG / c.1028delG	p.(G331Efs*28)/ p.(G343Efs*28)	0	Pathogenic (PS2, PS3, PM2, PM4, ,PP3)	F1/P1
7: 26232181	c.981delA / c.1017delA	p.(G328Afs*31)/ p.(G340Afs*31)	0	Pathogenic (PS2, PS3, PM2, PM4, PP3)	F2/P2
7: 26232178	c.984delC/ c.1020delC	p.(S329Vfs*30)/ p.(S341Vfs*30)	0	Likely Pathogenic (PS2, PM2, PM4, PP3)	F3/P3
7: 26232194	c.966delA/ c.1004delA	p.(N323Tfs*36)/ p. (N335Tfs*36)	0	Pathogenic (PS3, PM2, PM4, PP1, PP3)	F4/P4 and P5
7: 26232161	c.1001delG/ c.1037delG	p.(G334Vfs*25)/ p.(G346Vfs*25)	0	Uncertain Significance (PM2, PM4, PP3)	F5/P6
7: 26232165	c.996_997dupTG/ c.1032_1033dupTG	p.(G333Vfs*27)/ p.(G345Vfs*27)	0	Likely Pathogenic (PS2, PM2, PM4, PP3)	F6/P7 and F9/P10
7: 26232176	c.980_986del/ c.1016_102del	p.(G327Vfs*30)/ p.(G339Vfs*30)	0	Likely Pathogenic (PS2, PM2, PM4, PP3)	F7/P8
7: 26232188	c.974delG/ c.1010delG	p.(G325Vfs*34)/ p.(G337Vfs*34)	0	Likely Pathogenic (PS2, PM2, PM4, PP3)	F8/P9
7: 2623161	c.1001_1002dupGT/ c.1037_1038dupGT	p. (Y335Vfs*25)/ p.(Y347Vfs*25)	0	Uncertain Significance (PM2, PM4, PP3)	F10/P11

a.Genome reference GRCh37; *HNRNPA2B1* is expressed as two alternatively spliced isoforms: Variants are mapped on both hnRNPA2 (b. Transcript ID: NM\_002137) and hnRNPB1 (c. Transcript ID: NM\_031243). AF = allele frequency; ACMG=American College of Medical Genetics.

## Supplementary Note

### Clinical history and details

**Family 1.** Patient 1 is a 12-year-old boy of Italian and Colombian descent who was first noted to have dysphagia at approximately 2 years of age. By 4 years of age, ophthalmoplegia and non-fatigable ptosis became apparent. His early motor development was otherwise unremarkable. He subsequently experienced progressive difficulty with running, getting up from the floor, and gaining and maintaining weight. He has a history of recurrent respiratory infections and a decreased forced vital capacity (FVC) of 38% predicted. Examination at 12 years of age showed normal cognitive function, complete ptosis, near-complete ophthalmoplegia, facial weakness, a high arched palate, tongue weakness and globally atrophic muscles. He also had dysphonia and dysphagia. There was a severe pectus excavatum chest wall deformity. Mild large joint contractures in the ankles, knees, hips and jaw were present. Muscle strength examination showed proximal greater than distal weakness. Osteotendinous reflexes were reduced. He had a hyperlordotic posture and waddling gait, ambulating with hyperextended knees. He was unable to run. Serum creatine kinase (CK) was 724 units/L (normal 0-174 units/L). Nerve conduction studies (NCS) showed normal sensory responses and CMAP amplitudes. Electromyography (EMG) showed abnormal spontaneous activity in the form of fibrillation potentials and positive sharp waves and early recruitment of small amplitude and short duration motor unit action potentials. Repetitive stimulation and single fiber EMG (SFEMG) were normal. Muscle biopsy obtained at 8 years of age from the right rectus femoris muscle showed increased fiber size variability, increased endomysial fibrosis and cytoplasmic rimmed vacuoles. There was no inflammation or mitochondrial abnormalities. He was the only affected family member and whole exome sequencing identified a *de novo* heterozygous variant in *HNRNPA2B1* c.992delG,



p.(G331Efs\*28)/c.1028delG; p.(G343Efs\*28). Echocardiogram was normal. Cardiac MRI showed a severe pectus excavatum with a Haller index of 13 (normal < 2.5). Brain MRI at 7 years of age was normal. Serum screen for Paget's disease of the bone was normal.

**Family 2.** Patient 2 is a 17-year-old young man of Northern European descent who was first noted to have difficulty with feeds in the neonatal period. His early motor development was otherwise unremarkable. Ptosis and ophthalmoplegia were noted at approximately 5 years of age. He subsequently developed mild lower extremity predominant weakness involving the proximal muscles more than distal ones. Swallowing difficulties were also progressive and persistent. There was no history of respiratory insufficiency. Examination at 17 years of age showed bilateral ptosis, abduction and adduction deficit of the eye movements bilaterally and mild facial weakness. Contractures were not present. Muscle strength testing showed mild proximal greater than distal weakness. He had a normal gait. Mild scoliosis was also noted. Family history was non-contributory. A *de novo* heterozygous variant in *HNRNPA2B1* c.981delA, p.(G328Afs\*31)/c.1017delA (p.G340Afs\*31) was identified on whole exome sequencing. Serum CK was 602 units/L. NCS/EMG was not performed. Muscle biopsy obtained at 13 years of age from the left vastus lateralis muscle showed fiber size variability and rimmed vacuoles. Electron microscopy identified cytoplasmic, membrane-bound autophagic vacuoles and tubulofilamentous inclusions predominantly in the cytoplasm and peri-nuclear areas. A few angular atrophic fibers were also present. There was no inflammation or mitochondrial abnormalities. Brain MRI was normal.

**Family 3.** Patient 3 is a 17-year-old young woman of Ukrainian descent who was born premature at 33 weeks of gestation with bilateral congenital hip dysplasia. She had delays in reaching motor milestones. Ptosis and ophthalmoplegia were noted by 2.5 years of age, and global muscle weakness was apparent by 6 years of age. Her symptoms have been slowly progressive over time. Examination revealed ptosis, ophthalmoplegia, facial weakness and dysphonia. Mild proximal lower and upper extremity weakness was noted. Mild scoliosis was also present. Reflexes were reduced and gait exam showed a hyperlordotic posture. She was unable to run. Serum CK ranged between 638 and 994 units/L. Family history was non-contributory. A *de novo* heterozygous variant in *HNRNPA2B1* c.984delC, p.(S329Vfs\*30)/c.1020delC;p.(S341Vfs\*30) was identified on whole exome sequencing and was absent in the unaffected parents. Pulmonary function testing, brain MRI, NCS/EMG and muscle biopsy were not performed.

**Family 4.** This pedigree consists of a 73-year-old affected woman (recently deceased) with two affected daughters (Patient 4: 40 years and Patient 5: 43 years respectively), and an unaffected son 48 years of age. The index case is the only patient in her sibship and her parents died in their 80s with no neuromuscular problems. She first presented to medical attention at 41 years of age with a history of increasing limb weakness and dysphagia. She was noted to have a marked ptosis and almost complete ophthalmoplegia that reportedly had been present for many years. The limb weakness was most marked proximally and necessitated wheelchair use by her late 50s. The younger daughter, Patient 4, presented at 18 years of age with ophthalmoplegia and ptosis. Her symptoms progressed and she also noted a poor cough and significant dysphagia, and proximal limb weakness in her 30s. Serum CK was 808 units/L. The older daughter, Patient 5, presented at 17 years of age with bilateral ptosis, ophthalmoplegia, and dysphagia. By her 40s

she developed proximal upper and lower extremity weakness with progression of ophthalmoplegia and significant dysphagia. Serum CK was 1105 units/L. She also had respiratory insufficiency (FVC 45% predicted). Quadriceps muscle biopsy at 40 years of age showed significant variation in muscle fiber size with an increased number of internal nuclei and rimmed vacuoles. There was no evidence of inflammation or mitochondrial abnormalities. Electron microscopy was reported as normal. All affected family members carried a heterozygous variant in *HNRNPA2B1* c.966delA, p.(N323Tfs\*36)/c.1002delA;p.(N335Tfs\*36) with a dominant pattern of inheritance. OPMD gene analysis showed no evidence of an expansion of the GCG repeat sequence within *PABN1*. There were no abnormalities of bone biochemistry and no evidence of dementia or cognitive decline.

**Family 5.** Patient 6 is a 12-year-old girl of English ancestry who developed respiratory difficulties immediately after birth with subsequent tracheostomy, which remained in place at the time of evaluation. She also was noted to have a hemifacial macrosomia complex with left anophthalmia. She had mild delays in reaching gross motor milestones but was independently ambulatory with mild proximal greater than distal weakness. Her gait was waddling in quality with a hyperlordotic posture. She also had ptosis and complete ophthalmoplegia and dysphagia, requiring a percutaneous gastrostomy tube for nutrition. She had a jaw hamartoma that was resected. She also had respiratory insufficiency (FVC 21% predicted). Mild large joint contractures were present. Serum CK was 770 units/L. EMG/NCS showed an irritable myopathy with myotonic and pseudomyotonic discharges. Repetitive nerve stimulation and SFEMG was normal. Family history was non-contributory. A presumed *de novo* heterozygous variant in



*HNRNPA2B1* c.1001delG, p.(G334Vfs\*25)/c.1037delG;p.(G346Vfs\*25) was identified on whole exome sequencing.

**Family 6.** Patient 7 is a 9-year-old French girl whose gestation was notable for reduced fetal movements. She was born with intrauterine growth retardation. Ptosis was noted by her parents at birth, which became more pronounced with fatigue and as she aged. Respiratory impairment began at 6 months of age with bronchiolitis and recurrent pneumonias. She walked at 14-15 months of age and did not have a speech delay. Axial weakness and ophthalmoplegia were noted at 4.5 years of age. At age 5, she also developed dysphonia and dysphagia, requiring a percutaneous gastrostomy tube for nutrition and respiratory insufficiency necessitated the use of nocturnal non-invasive ventilatory support. At age 8, she her respiratory insufficiency worsened and she required continuous non-invasive ventilation and her muscle weakness progressed, resulting in loss of independent ambulation and use of a motorized wheelchair. Serum CKs ranged between 287 and 536 units/L. NCS/EMG was normal except for increased latency of tendon reflexes (T reflex). Repetitive nerve stimulation was normal. Quadriceps muscle biopsy obtained at 4 years and 9 months of age showed myofiber necrosis and regeneration, and subsarcolemmal vacuoles. Family history was non-contributory. A *de novo* heterozygous variant in *HNRNPA2B1* c.996\_997dup, p.(G333Vfs\*27)/c.1032\_1033dup; p.(G345Vfs\*27) was identified on trio-based whole exome sequencing.

**Family 7.** Patient 8 is a 20-year-old French man with delays in achieving early motor milestones as an infant. During childhood, he had a propensity to fall and had difficulty participating in sports. He had ptosis, ophthalmoplegia and dysphonia. He also had axial and proximal greater

than distal weakness in all limbs that was progressive, resulting in loss of independent ambulation. He also has respiratory insufficiency and uses non-invasive ventilatory support at nighttime. Large joint contractures and scoliosis were also noted. Serum CK was 2484 units/L. NCS/EMG was consistent with a myopathic process with fibrillation potentials at rest. Repetitive nerve stimulation was normal. Muscle biopsy at 14 years of age showed variability in muscle fiber size, myofiber atrophy, and cytoplasmic rimmed vacuoles and internalized nuclei. There was also a moderate increase in the endomysial connective tissue. Electron microscopy studies showed marked autophagic changes and vacuoles containing membranous, myelin-like remnants. Many myofibrils contained areas with tubulofilamentous inclusions, which on occasion were also present in myonuclei. Brain MRI was normal. Family history was non-contributory. Next generation sequencing panels including LGMD genes and congenital myopathy genes were non-diagnostic. Due to the similarity of the phenotype with other cases in this cohort, Sanger sequencing of the *HNRNPA2B1* gene was pursued. A *de novo* heterozygous variant in *HNRNPA2B1*, c.980\_986del, p.(G327Vfs\*30)/c.1016\_1022del, p.(Gly339Valfs\*30), was identified.

**Family 8.** Patient 9 is a 37-year-old man born of non-consanguineous parents of Italian descent. There was no delay in the acquisition of psychomotor milestones, although he was reported to always have been slower than his peers in running. Starting at 5 years of age, he developed progressive muscle weakness especially in the lower limbs, skeletal deformities of talipes equinovarus and scoliosis, bilateral ptosis and ophthalmoplegia. He lost the ability to run at 13 years of age and lost the ability to independently ambulate at 26 years of age. He underwent a blepharoplasty for severe bilateral ptosis at 28 years of age. He also developed a severe

dysphagia and currently needs a semi-liquid diet and thickened liquids. Echocardiography was normal and pulmonary function testing showed a restrictive defect (FVC 34% predicted), resulting in the need for non-invasive nocturnal ventilation since 31 years of age. At 34 years of age, he underwent surgery for a retraction of the temporomandibular joint. The most recent neurological examination (35 years of age) showed severe weakness in all extremities with anti-gravity movements only possible in the distal segments of upper limbs, inability to sit without support, absent deep tendon reflexes, normal sensation, surgically corrected ptosis, complete loss of vertical and near-complete loss of horizontal eye movements, moderate facial weakness, hypomobility of the velum palatinum and a hypotrophic tongue. Serum CK was elevated between 280 and 7067 units/L. A muscle biopsy at age 9 years showed nonspecific myopathic changes, while a second biopsy at age 13 was notable for a severe myopathic picture, with core-like areas of reduced oxidative reaction in both fiber types and rimmed vacuoles. Genetic testing for OPMD (GCN repetitions in *PABPN1*) was normal. Next generation sequencing (Illumina Trusight One expanded panel of human disease-causing genes) identified a heterozygous c.974delG, p.(G325Vfs\*34)/c.1010delG, p.(Gly337ValfsTer34) variant in *HNRNPA2B1* which was confirmed to be *de novo* by Sanger sequencing of parental samples.

**Family 9.** Patient 10 is a 7-year-old girl of Indian origin who presented with a respiratory infection at 18 months of age with acute onset bilateral ptosis lasting a week. Her early motor milestones and motor function were normal until 4.5 years of age. Axial weakness, ptosis and ophthalmoplegia were first noted at 4.5 years of age, coinciding again with a respiratory illness. She also developed dysphonia and dysphagia, requiring a percutaneous gastrostomy tube for nutrition at 6 years of age. She has a weak cough. Her weakness in the extremities developed and



progressed, resulting in loss of independent ambulation by 7 years of age. She also has long finger flexor and ankle contractures. She demonstrates increased weakness during intercurrent chest infections. Serum CK levels were elevated at 424 units/L. NCS/EMG was consistent with a myopathic process. Stimulated, single fiber EMG was normal and brain MRI was unremarkable. Plasma lactate was 3.9; urine organic acids showed strongly raised pyruvate with mildly raised lactate and respiratory chain studies detected low complex I and IV activity. Mitochondrial DNA sequencing did not reveal any pathogenic variants. Quadriceps muscle biopsy obtained at 5 years and 6 months of age showed rimmed vacuoles, sparse protein aggregates and few central/internal nuclei with scattered basophilic coarsely granular fibers, some of which contained large eosinophilic inclusions. Acid phosphatase activity was high in many of the rimmed vacuoles. Family history was non-contributory. A *de novo* heterozygous variant in *HNRNPA2B1* c.996\_997dup, p.(G333Vfs\*27)/c.1032\_1033dup; p.(G345Vfs\*27) was identified on genome sequencing.

**Family 10.** Patient 11 is an 18-year-old woman of Japanese origin with normal motor milestones, who was noted to be a slow runner in childhood. At 10 years of age, short stature was noted; however, growth hormone levels were within the normal range. She developed progressive weakness: at 15 years of age, she was noted to have easy fatiguability and difficulty running and climbing stairs. On examination, bilateral ptosis, ophthalmoplegia, lower proximal muscle weakness, scoliosis, and waddling gate were noted at 17 years of age. She has a decreased forced vital capacity (FVC) of 58% predicted. Serum CK levels were elevated at 501 Units/L. NCS was normal. Brain MRI was unremarkable. Muscle CT revealed severely affected thigh muscle with relative sparing of the rectus femoris, gracilis, and adductor longus muscle.

Mitochondrial DNA sequencing did not reveal any pathogenic variants. Vastus medialis muscle biopsy obtained at 17 years and 11 months of age showed scattered fibers with rimmed vacuoles, small angular fibers, and moderate to marked fiber size variation. Exome sequencing identified a heterozygous variant in *HNRNPA2B1* c.1001\_1002dup, p.(Y335Vfs\*25)/c.1037\_1038dup; p.(Y347Vfs\*25).

In summary, in ten unrelated families, we identified a cluster of heterozygous frameshift variants in *HNRNPA2B1* that clinically manifest with a remarkably similar phenotype of ophthalmoplegia, ptosis, dysphagia, gradually progressive proximal and distal limb weakness, and variable degrees of respiratory insufficiency.

### **Supplementary References**

1. Kim, H.J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467-73 (2013).