GGC repeat expansions within new open reading frames are translated into toxic polyglycine proteins in oculopharyngodistal myopathy

Manon Boivin, Jiaxi Yu, Nobuyuki Eura, Léa Schmitt, David Pietri, Erwan Grandgirard, Patrice Goetz-Reiner, Damien Plassard, Chadia Nahy, Anne Maglott, et al.

nature genetics

Article

https://doi.org/10.1038/s41588-026-02507-z

GGC repeat expansions within new open reading frames are translated into toxic polyglycine proteins in oculopharyngodistal myopathy

Received: 28 February 2025

Accepted: 9 January 2026

Published online: 17 February 2026

Check for updates

Manon Boivin[1,7] ✉, Jiaxi Yu[2,6], Nobuyuki Eura[3], Léa Schmitt[1], David Pietri[1], Erwan Grandgirard[1], Patrice Goetz-Reiner[1], Damien Plassard[1], Chadia Nahy[1], Anne Maglott[1], Bastien Morlet[1], Chao Gao[2], Elise Lefebvre[1], Muriel Philipps[1], Pascal Eberling[1], Angélique Pichot[1], Paola Rossolillo[1], Christelle Thibault[1], Mustapha Oulad-Abdelghani[1], Ichizo Nishino[3], Kang Yang[4], Ning Wang[4], Zhaoxia Wang[2], Jianwen Deng[2,5,7] ✉ & Nicolas Charlet-Berguerand[1,7] ✉

A total of 3–6% human genome is composed of microsatellite sequences, which are short DNA elements composed of two to six nucleotide motifs repeated in tandem. Expansion of a subset of these microsatellites is the leading cause of >60 diseases. However, most of these mutations are located in sequences annotated as noncoding, which raises questions about their pathogenicity. Here we found that GGC repeat expansions causing oculopharyngodistal myopathy with or without oculopharyngeal myopathy leukoencephalopathy are located within previously unrecognized open reading frames (ORFs), resulting in their translation into new polyglycine-containing proteins. Antibodies developed against these proteins stain the p62-positive inclusions typical of these diseases. Moreover, expression of these polyglycine proteins causes locomotor and skeletal muscle alterations associated with neurodegeneration in cell, fly and mouse models. Finally, we identified a compound, the cationic porphyrin TMPyP4, targeting the expression of these polyglycine proteins, raising hope to develop a therapy for these disorders. Overall, this work highlights the complexity and richness of the human genome and the importance of mutations in yet-unrecognized small ORFs.

A total of ~98% human genome is comprised of sequences annotated as noncoding, with half of them composed of repetitive DNA elements, including microsatellites, which are two- to six-nucleotide DNA motifs repeated in tandem. These microsatellites, estimated between ~1.5 and 2 million in humans, occupy 3–6% of our genome and are a source of genetic variation as they are highly heterogeneous in size and sequences[1]. Expansion of a subset of these microsatellites over a threshold size is also the leading cause of various human pathologies, including cancer and inherited diseases[2,3]. In that aspect, >60 neurodevelopmental, neuromuscular, and neurodegenerative disorders are known to be caused by expansions of trinucleotide, tetranucleotide, pentanucleotide or hexanucleotide repeats. Remarkably, this number is rapidly increasing as advances in long-read and whole human-genome sequencing have revealed ~20 new pathogenic microsatellite expansions causing human genetic diseases in recent years[4–12]. When embedded within a coding sequence, in-frame repeat expansions

A full list of affiliations appears at the end of the paper. ✉e-mail: boivinm@igbmc.fr; jianwendeng@pkufh.com; ncharlet@igbmc.fr

are translated into a mutant protein containing a stretch of repeated amino acids. The archetype of this mechanism is the polyglutamine (polyQ) group of diseases, where expansions of CAG repeats, embedded within the open reading frames (ORFs) of diverse genes, are translated into toxic polyQ-containing proteins, ultimately resulting in neuronal cell dysfunction and death. However, a majority of microsatellite expansions, notably most of the recently discovered ones, are located in genomic sequences annotated by default as noncoding (5′-untranslated region (5′UTR) and 3′UTR, introns, antisense RNAs, long noncoding RNAs, etc.), thus questioning their pathogenicity[13].

Oculopharyngodistal myopathy (OPDM; OMIM 164310) is a rare adult-onset and slowly progressive neuromuscular disease first described in 1977 (ref. 14), while oculopharyngeal myopathy with leukoencephalopathy (OPML; OMIM 618637) is an autosomal dominant disorder with oculopharyngeal myopathy, diffuse limb weakness and leukoencephalopathy described more recently[4]. Key clinical features of OPDM and OPML comprise ptosis, external ophthalmoplegia, dysphagia and dysarthria associated with facial and distal limb muscle weakness. Their histopathology is characterized by the presence of large cytoplasmic rimmed vacuoles and rare, but typical, eosinophilic intranuclear inclusions, which are p62-positive and ubiquitin-positive but of unknown origin[4–10,15]. The genetic causes of OPDM and OPML were uncovered recently as similar expansions of ~50 to 200–300 repeats of the trinucleotide GGC sequence located within diverse genomic regions, transcribed but annotated as noncoding and embedded in at least six different genes (*LOC642361*, *LRP12*, *GIPC1*, *NOTCH2NLC*, *RILPL1* and *ABCD3*)[4–11]. Consequently, these pathologies are now classified into at least six subtypes according to the gene hosting the pathogenic GGC repeat expansion (*LOC642361*, OPML; *LRP12*, OPDM1; *GIPC1*, OPDM2; *NOTCH2NLC*, OPDM3; *RILPL1*, OPDM4; *ABCD3*, OPDM5). Of interest, recent clinical studies indicate that OPDM and OPML have a much wider clinical spectrum than previously thought, with evidence of variable neurological manifestations and reports of movement disorders, tremor, ataxia, visual disturbance, peripheral neuropathy, etc. Finally, OPDM3 shares the same genetic cause with neuronal intranuclear inclusion disease (NIID)[4,5,10]. NIID is a neurological disease characterized by variable muscle weakness associated with heterogeneous dysfunctions of the central and peripheral nervous system. These genetic similarities and clinical overlaps suggest that OPDM, OPML and NIID belong to a new continuum of neurological diseases, which possibly share a common pathophysiological mechanism[16,17]. However, a loss-of-function mechanism is unlikely as expression of the genes hosting these GGC repeat expansions is unaltered in tissue samples from individuals with OPDM, NIID or OPML, with observation of increased or unchanged mRNA levels and, for coding genes, normal protein expression[6–10]. These observations exclude a classical promoter-silencing mechanism but raise questions about how GGC repeat expansions, located within genomic regions annotated as noncoding, can lead to the formation of protein inclusions and cause muscle and neuronal dysfunction.

Here we found that the GGC repeats located in the long 'noncoding' *LOC642361* RNA, as well as in the 'noncoding' sequences of the *GIPC1*, *NOTCH2NLC* and *RILPL1* genes, are embedded within previously unrecognized ORFs, resulting in expression of new proteins where each GGC repeat encodes for a glycine amino acid. Consequently, these GGC repeat expansions are translated into new polyglycine-containing proteins. Antibodies developed against these proteins confirmed their expression in patients, notably their localization in p62-positive inclusions in muscle sections of individuals with OPDM and OPML. Moreover, expression of these polyglycine proteins in cell and animal models is sufficient to induce formation of the characteristic OPDM/OPML p62-positive inclusions, as well as muscle fiber atrophy associated with neurodegeneration and neuroinflammation, thus recapitulating key clinical features of OPDM, OPML and NIID. Of interest, side-by-side comparison of these diverse polyglycine proteins reveals unexpected variations in their biological properties and toxicity, highlighting a key contribution of the specific amino acid sequences flanking their common polyglycine core. Finally, we tested various pharmacological compounds and identified the cationic porphyrin TMPyP4 as a proof-of-concept therapeutic for these neurological disorders.

Overall, this study highlights the richness and complexity of the human genome, notably the existence of numerous uncharted small ORFs in sequences originally annotated as noncoding, resulting in translation of their embedded microsatellite mutations into new and toxic proteins.

Results

GIPC1, antisense *RILPL1* and *LOC642361* GGC repeats are translated into polyglycine

As noncanonical translation of repeat expansions is an established model of pathogenicity in microsatellite diseases[18,19], we investigated the potential translation of the GGC repeats causing OPDM and OPML. Three representative noncoding sequences, namely the 5′UTR of *GIPC1*, the antisense transcript of *RILPL1* and the *LOC642361* long noncoding RNA (lncRNA), which are the cause of OPDM2, OPDM4 and OPML, respectively, were cloned with ~50 GGC repeats and fused to green fluorescent protein (GFP) in the three possible frames potentially encoded by these repeats (glycine, alanine and arginine; Fig. 1a,b, Supplementary Note 1 and Supplementary Fig. 1a–f). Notably, transfection into HEK293 cells followed by fluorescence-activated cell sorting (FACS) analysis, direct observation of the GFP fluorescence and western blotting consistently indicate that the *GIPC1* GGC repeats are predominantly translated in the glycine frame, while GFP expression in the alanine or arginine frames is negligible (Fig. 1c–e). Similar results were obtained with the antisense *RILPL1* and *LOC642361* noncoding RNAs (Extended Data Fig. 1a–f). In contrast, analysis of the sense *RILPL1* RNA with a CCG expansion shows no detectable translation in any frames (Supplementary Fig. 1g–i). Controls indicate that the lack of GGC repeat translation into polyalanine or polyarginine is not caused by differences in RNA expression, a potential toxicity leading to cell loss or another bias impairing observation of GGC repeats translation in the alanine or arginine frames (Supplementary Note 1 and Supplementary Fig. 1j–l). Finally, the treatment of cell extracts with lysostaphin, a glycyl–glycine endopeptidase, cleaves these proteins into smaller products, thus confirming the presence of a polyglycine stretch within them (Extended Data Fig. 1g–i). Overall, these results indicate that the *GIPC1*, antisense *RILPL1* and *LOC642361* GGC repeat expansions, while located in sequences annotated as noncoding, are translated into new polyglycine-containing proteins that have yet to be characterized.

GIPC1, antisense *RILPL1* and *LOC642361* GGC repeats are embedded in small ORFs

To uncover how these microsatellites are translated, constructs with 50 GGC repeats embedded in the upstream *GIPC1*, antisense *RILPL1* or *LOC642361* sequences and fused to GFP in the glycine frame were transfected into HEK293 cells and their corresponding encoded GFP-tagged polyglycine proteins were immunoprecipitated and analyzed by mass spectrometry to determine their N-terminal sequences (Fig. 2a). In all three sequences, translation starts with a typical acetylated methionine (M^{ac}), which corresponds to initiation at standard start codons located upstream of the GGC repeats (Fig. 2a and Supplementary Fig. 2a–d). Translation initiations of the *RILPL1* antisense RNA and of the *LOC642361* lncRNA occur at classical ATG start codons, while translation of the *GIPC1* 5′UTR occurs in the absence of any ATG start codon but instead initiates at a CTG near-cognate start codon located ahead of the repeats (Supplementary Fig. 2a and Supplementary Note 2). Near-cognate start codons (CTG, GTG, ACG, TTG) are codons differing from the cognate AUG start codon by one nucleotide, but that can nonetheless initiate translation through mispairing with the initiator methionine-tRNA[20,21]. Western blotting, fluorescence observation and FACS analyses show that the deletion of these ATG or CTG start codons abolishes polyglycine expression, highlighting the requirement of these start codons to drive GGC repeats translation (Fig. 2b and Supplementary Fig. 2e–p). Next,

https://doi.org/10.1038/s41588-026-02507-z

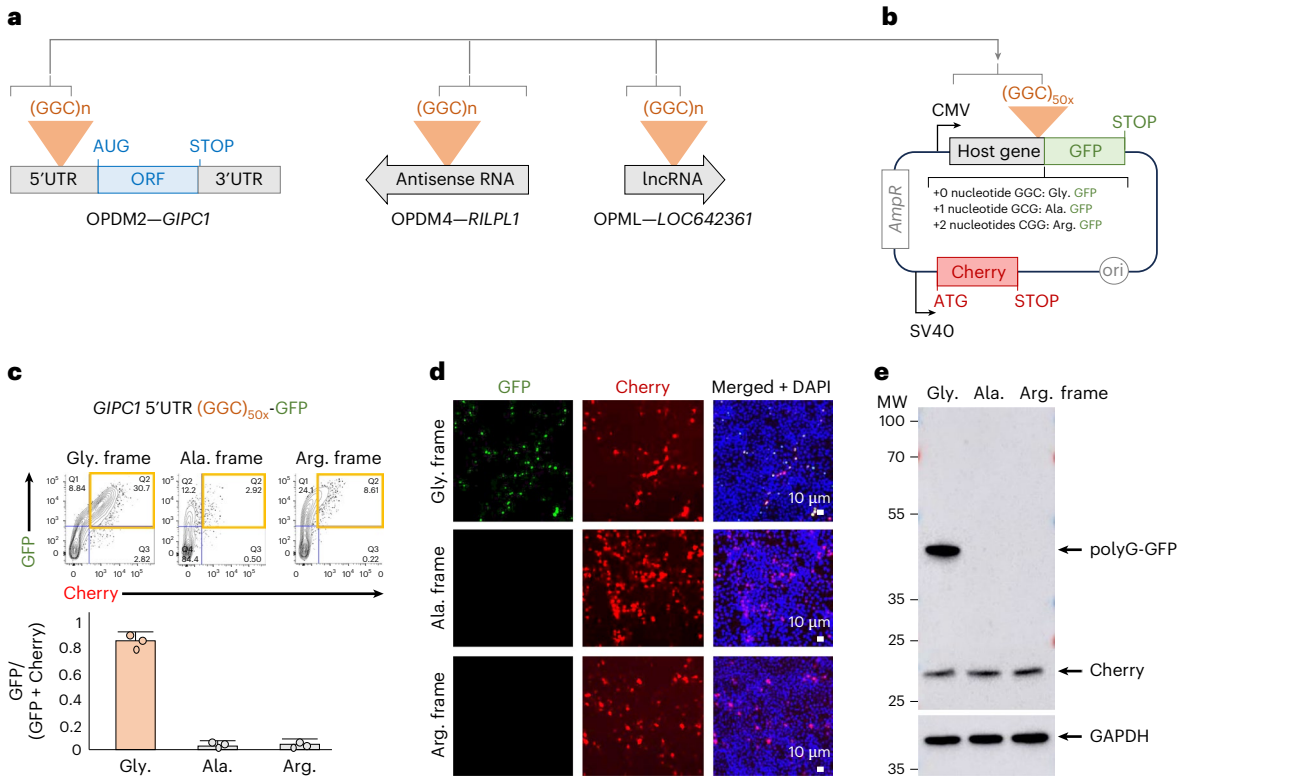


Fig. 1 | *GIPC1*, *RILPL1* and *LOC642361* GGC repeats are translated into polyglycine. a, Scheme of the GGC repeat expansions located within the *GIPC1*, antisense *RILPL1* and *LOC642361* sequences. **b**, Scheme of the construct expressing ~50 pure GGC repeats cloned with their upstream host sequences and fused to the GFP in their three potential encoded frames. This plasmid also contains an independent Cherry expression cassette. **c**, GFP and Cherry FACS analysis (top) with its quantification (bottom) of HEK293 cells transfected for 24 h with a plasmid expressing 50 GGC repeats embedded within the *GIPC1* 5′UTR sequence fused to the GFP in the glycine, alanine or arginine frames, while the Cherry is expressed independently. Bar heights represent the mean and error bars represent the mean ± s.e.m. Sample size, $n = 3$ independent biological replicates. **d**, As in **c** but with microscopy fluorescence analysis. Scale bars = 10 µm. **e**, As in **c** but with immunoblotting analyses (Extended Data Fig. 1 and Supplementary Fig. 1). Full-length blots are provided as Source data. Gly., glycine; Ala., alanine; Arg., arginine.

we noted that the *GIPC1* 5′UTR and the *LOC642361* lncRNA sequences with a control size of repeats (~10 GGC) are similarly translated, but into small and unstable peptides, which are hardly detected without inhibition of the cell degradation pathways (Extended Data Fig. 2a–c). In contrast, these microproteins become stable when carrying an expansion of their polyglycine stretch (Extended Data Fig. 2d,e). Overall, these data reveal that the 5′UTR of the *GIPC1* gene, the antisense transcript of *RILPL1* and the *LOC642361* long noncoding RNA contain previously unrecognized ORFs, which are translated independently of the length of their GGC microsatellites. Their translation initiates at either cognate AUG or near-cognate start codons located ahead of the GGC repeats and in the glycine frame, resulting in expression of new proteins where each GGC repeat encodes for a glycine amino acid. In the absence of a GGC expansion and thus with a normal size of glycine stretch, their encoded peptides are of small size (<100 amino acids) and thus unstable and hardly observable. In contrast, when encompassing an expansion over ~50 GGC repeats, these ORFs are translated into new polyglycine-containing proteins, which were named uGIPpolyG, asRILpolyG and LOC6polyG for upstream of GIPC1, antisense of RILPL1 and LOC642361-encoded polyglycine proteins, respectively (Fig. 2c–e, Supplementary Note 2 and Supplementary Fig. 2q–y).

Polyglycine proteins colocalize with p62 inclusions in OPDM/OPML muscle sections

To confirm that these GGC repeat expansions are translated into new polyglycine-containing proteins in individuals with OPDM/OPML, we developed antibodies against their distinct N-terminal or C-terminal sequences (Supplementary Note 3 and Supplementary Fig. 3a–l). Immunofluorescence staining performed on skeletal muscle sections of individuals with OPDM2, OPDM4 and OPML revealed the presence of their respective polyglycine proteins (uGIPpolyG, asRILpolyG and LOC6polyG) within the p62-positive cytoplasmic rimmed vacuoles and intranuclear inclusions typical of these diseases (Fig. 3a–c and Supplementary Fig. 3m–o). Moreover, as OPDM3 and NIID have an identical genetic cause, namely an expansion of GGC repeats in the 5′UTR of the *NOTCH2NLC* gene, and as this expansion was recently found to belong to a small ORF translated in a polyglycine protein, uN2CpolyG[22–24], we developed antibodies against this protein and uncovered its presence within the typical p62-positive inclusions in muscle sections of individuals with OPDM3 (Fig. 3d and Supplementary Fig. 3p). No or only faint staining was observed in non-OPDM individuals (Fig. 3a–d and Supplementary Fig. 3m–p), as without a GGC repeat expansion and thus without a polyglycine stretch, these microproteins do not aggregate, and their small sizes impair their stability and their detection. Moreover, as each of these antibodies is directed against a specific ORF sequence encoding a distinct polyglycine protein, these antibodies are specific to each OPDM subtype and indeed do not stain p62-positive inclusions in other OPDM/OPML subtypes (Extended Data Fig. 3a–d). These controls confirm the specificity of our antibodies and support the existence of distinct ORFs, each encoding a unique polyglycine protein, which are consequently specific to each OPDM subtype. Finally, as various microsatellite expansions have been reported to be RAN translated in their three potential frames[18,19], and as a short expansion of GCN repeats in *PABPN1* is translated into a protein with an extended polyalanine stretch that causes oculopharyngeal muscular dystrophy[25], we also investigated a potential

https://doi.org/10.1038/s41588-026-02507-z

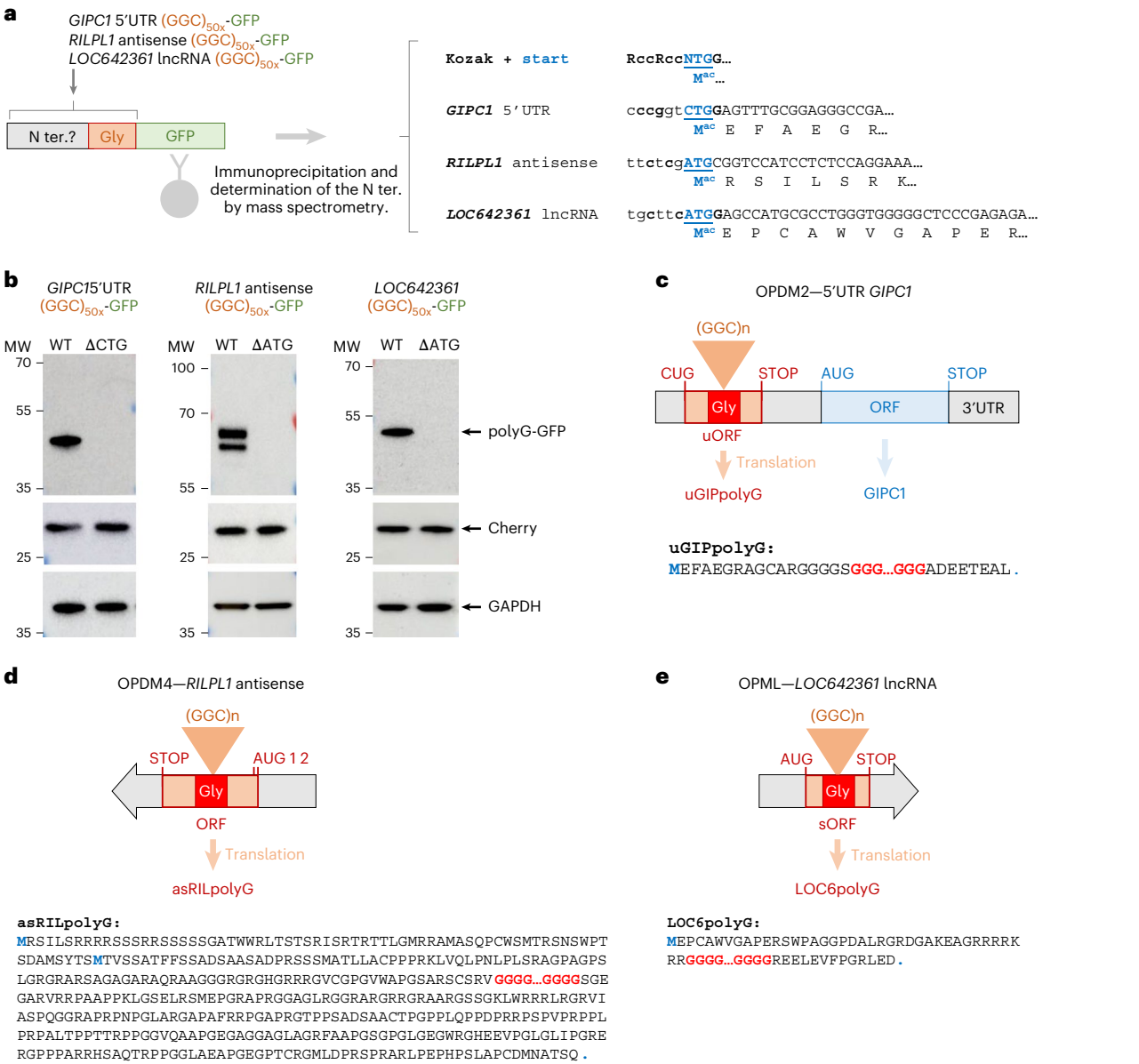


Fig. 2 | *GIPC1*, *RILPL1* and *LOC642361* GGC repeats are embedded in small ORFs. a, Scheme of HEK293 cells transfected for 24 h with a plasmid expressing 50 GGC repeats cloned with their upstream *GIPC1*, antisense *RILPL1* and *LOC642361* sequences and fused to GFP in the glycine frame, followed by GFP-immunoprecipitation and mass spectrometry analysis to determine the N-terminal sequences of their encoded polyglycine proteins. **b**, Immunoblot against the GFP, Cherry or GAPDH of proteins extracted from HEK293 cells transfected as in **a** but with WT or mutant (ΔCTG/ATG start codons) constructs expressing 50 GGC repeats cloned with their upstream *GIPC1*, antisense *RILPL1* and *LOC642361* sequences and fused to GFP in the glycine frame. **c**–**e**, Schemes and amino acid sequences of the new ORFs and their encoded polyglycine proteins identified in the *GIPC1* 5′UTR (**c**), *RILPL1* antisense transcript (**d**) and *LOC642361* lncRNA (**e**) (Extended Data Fig. 2 and Supplementary Fig. 2). Full-length blots are provided as Source data. WT, wild type; sORF, short ORF; ter., terminal.

translation of the *GIPC1* GGC repeats in the alanine frame. However, two independent antibodies developed against a putative GIPC1 polyalanine protein do not stain intranuclear inclusions or rimmed vacuoles in muscle sections of individuals with OPDM2, arguing against translation of GGC repeats in the alanine frame (Supplementary Fig. 3q–s). Overall, these data confirm that the GGC repeat expansions causing OPDM and OPML are embedded in previously unrecognized ORFs and consequently translated into new polyglycine-containing proteins.

Expression of polyglycine proteins forms inclusions and is pathogenic in muscle cells

To study further these polyglycine proteins, we cloned their cDNAs with an expansion of 100 GGN glycine-encoded codons, a strategy preventing repeat instability (Fig. 4a, Supplementary Note 4 and Supplementary Fig. 4a). As a control, expression of 100 pure GGC RNA repeats, deprived of any translation start codon, is not toxic, dismissing a potential RNA toxicity mechanism (Supplementary Note 4 and Supplementary Fig. 4b–e). Of interest, expression of these diverse polyglycine proteins in human LHCN-M2 differentiated muscle cells followed by immunofluorescence revealed that they form cytoplasmic and intranuclear inclusions, which are p62-positive and thus reminiscent of the OPDM, OPML and NIID histopathological features (Fig. 4b,c and Supplementary Fig. 4f). Live imaging suggests that cytoplasmic polyglycine inclusions may be too large to penetrate nuclei but may directly aggregate and grow within the nucleus from soluble polyglycine species and/or microaggregates (Supplementary Videos 1 and 2). Correlative light and electron microscopy (EM) shows that these polyglycine inclusions appear as round-shaped electron-dense deposits composed of

https://doi.org/10.1038/s41588-026-02507-z

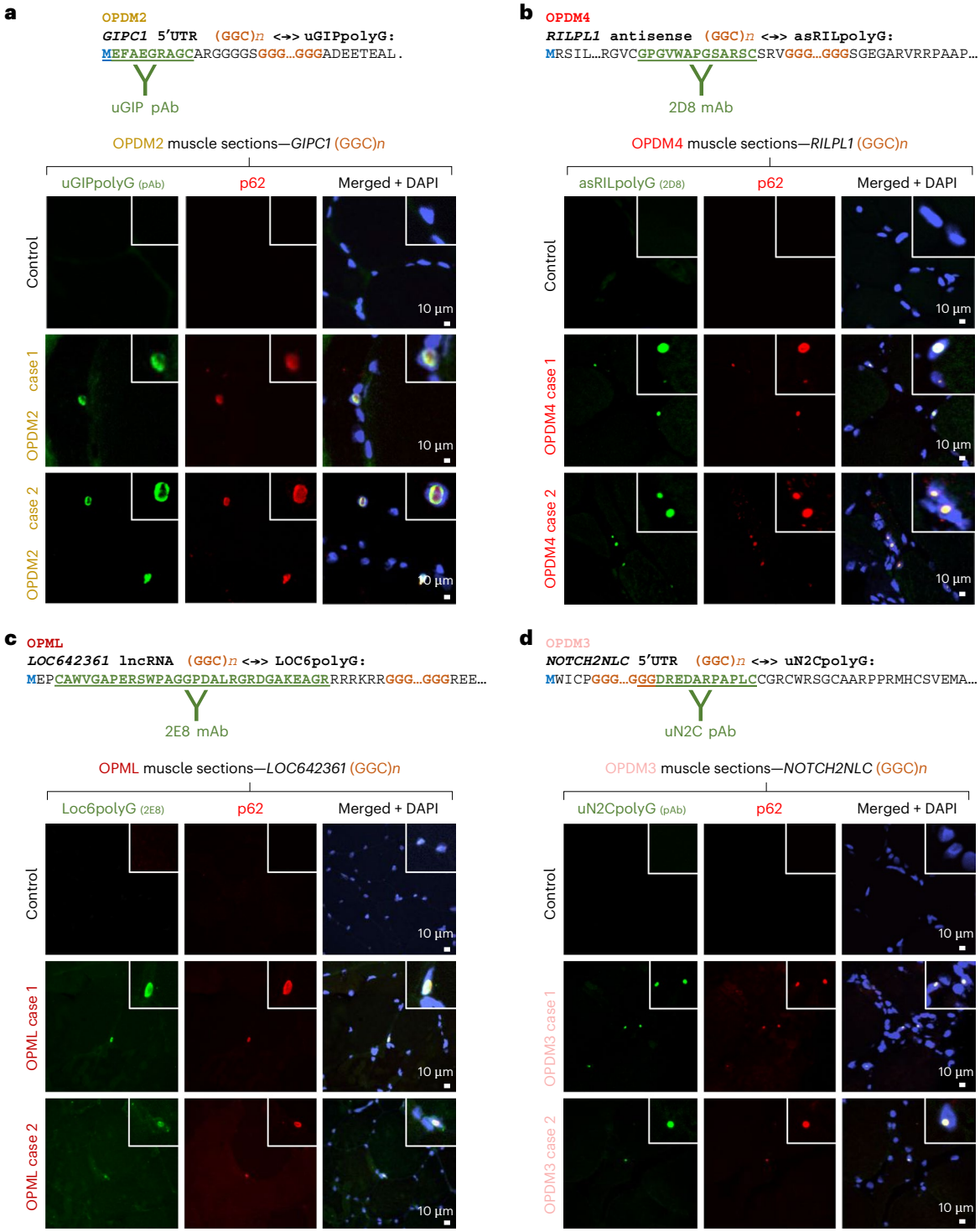


Fig. 3 | Polyglycine proteins are present in typical OPDM/OPML p62-positive inclusions. a, Top, partial amino acid sequence of the uGIPpolyG protein encoded by the expanded GGC repeats embedded in the *GIPC1* 5′UTR sequence causing OPDM2. The peptide sequences against which the uGIP antibody is directed are indicated in bold and underlined. Bottom, immunofluorescence staining against p62 and the uGIPpolyG protein on skeletal muscle sections of individuals with OPDM2 or age-matched control individuals. **b**, As in **a** but with the asRILpolyG protein stained in individuals with OPDM4. **c**, As in **a** but with the LOC6polyG protein stained in individuals with OPML. **d**, As in **a** but with the uN2CpolyG protein stained in individuals with OPDM3. Scale bar = 10 µm (**a**–**d**; Extended Data Fig. 3 and Supplementary Fig. 3).

filamentous structures without membrane boundaries (Fig. 4d), which is consistent with observations in individuals with OPDM and NIID. Of interest, we noted that these different polyglycine proteins present some differences in their localization, with the OPDM4 asRILpolyG protein being systematically more nuclear than the other (Fig. 4e). Similarly, protein expression assessed by immunoblotting revealed further unexpected variations and different protein half-life, with the uN2CpolyG (OPDM3/NIID) protein consistently less observed (Fig. 4f, Supplementary Note 4 and Supplementary Fig. 4g). As polyglycine proteins accumulate in cellular inclusions that may correspond to insoluble

https://doi.org/10.1038/s41588-026-02507-z

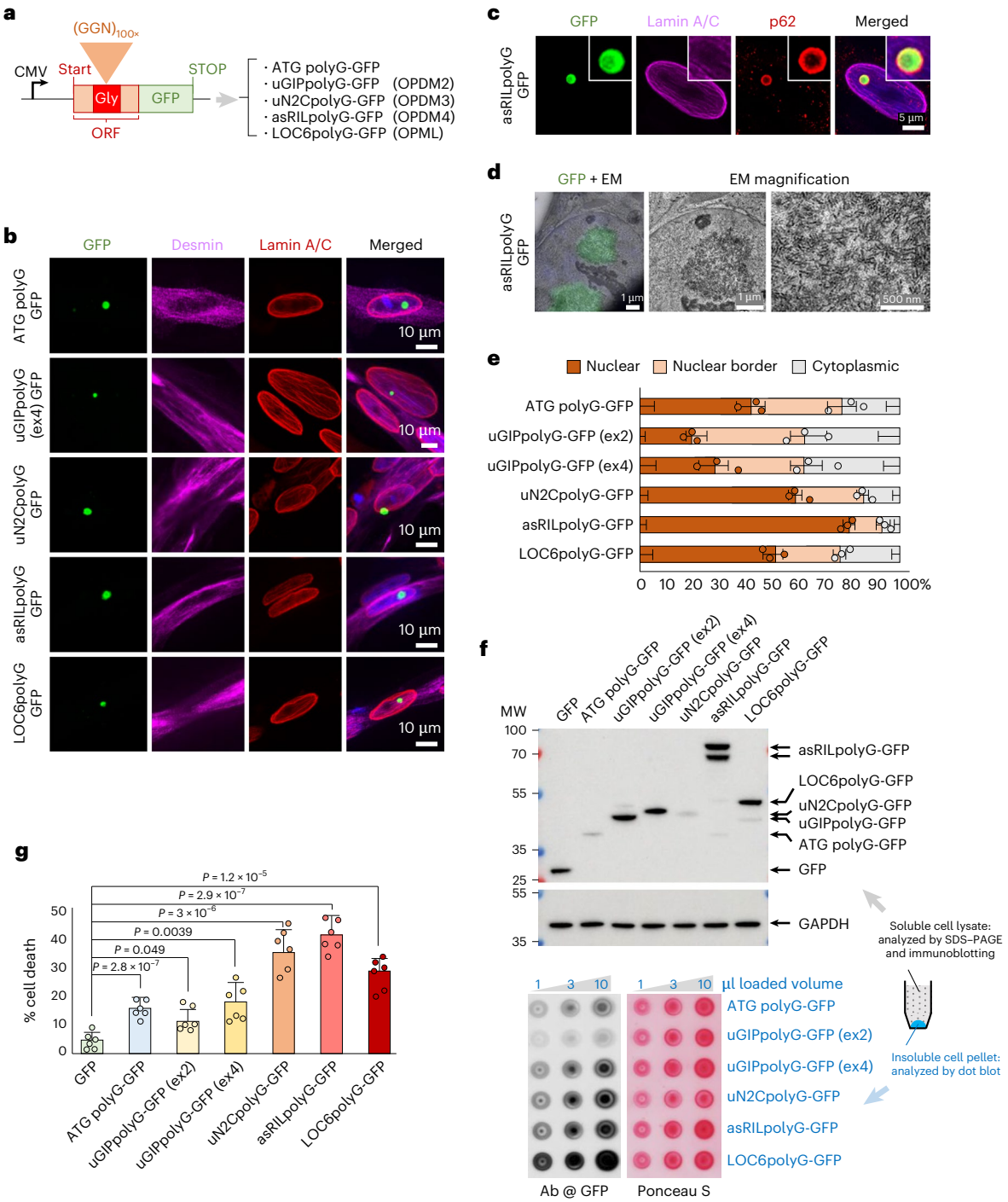


Fig. 4 | Expression of polyglycine proteins forms inclusions and is pathogenic in muscle cells. a, Scheme of the constructs encoding GFP-tagged polyGly, uGIPpolyG, uN2CpolyG, asRILpolyG or LOC6polyG protein cloned with an optimized expansion of 100 GGN repeats. **b**, GFP fluorescence and immunofluorescence against the desmin and lamin A/C proteins of LHCN-M2 cells differentiated into myotubes for 4 days and expressing either GFP-tagged ATG polyG, uGIPpolyG, uN2CpolyG, asRILpolyG or LOC6polyG. Scale bars = 10 µm. **c**,**d**, GFP fluorescence with immunofluorescence against p62 and lamin A/C (**c**) or CLEM (**d**) of LHCN-M2 cells expressing asRILpolyG-GFP and differentiated into myotubes for 4 days. Scale bars, as indicated. **e**, Quantification of nuclear versus cytoplasmic localization of the polyglycine proteins studied in **b**. Bar heights represent the mean. Error bars represent the mean ± s.e.m. Sample size, $n = 3$ independent biological replicates. **f**, SDS–PAGE gel and immunoblot against the GFP or the GAPDH of soluble proteins (top), and dot blot against the GFP or Ponceau staining of the insoluble proteins (bottom) extracted from 48-h differentiated LHCN-M2 muscle cells expressing either the GFP or GFP-tagged ATG polyG, uGIPpolyG, uN2CpolyG, asRILpolyG or LOC6polyG. **g**, Cell viability of LHCN-M2 muscle cells differentiated for 3 days and expressing GFP or GFP-tagged ATG polyG, uGIPpolyG, uN2CpolyG, asRILpolyG or LOC6polyG. Bar heights represent the mean and error bars represent the mean ± s.e.m. Sample size, $n = 6$ independent biological replicates. Unpaired two-tailed t test compared to the GFP control condition (Extended Data Fig. 4 and Supplementary Fig. 4). Full-length blots are provided as Source data. CLEM, correlative light and EM.

protein aggregates, which classically escape to immunoblot detection performed on the soluble cell fraction, we also performed dot blot analysis of the cell lysate pellet sonicated in 2× Laemmli buffer (Fig. 4f). This assay exposed further disparities between these polyglycine proteins, with the uN2CpolyG, asRILpolyG and LOC6polyG proteins notably more present in the insoluble protein fraction. These results were confirmed by quantification of the localization of these diverse polyglycine proteins in LHCN-M2 muscle cells, notably their presence in inclusions versus a diffuse localization (Extended Data Fig. 4a). Next, immunoprecipitation of these diverse polyglycine proteins followed by mass spectrometry unveiled specific interactants, notably some that are exclusive to peculiar polyGly proteins (Extended Data Fig. 4b and Supplementary Data 1). In that aspect, the uN2CpolyG protein interacts with the KU70/KU80 dimer involved in DNA repair, while the LOC6polyG interacts with ribosomal proteins. These interactions are independent of their glycine stretches, suggesting that the newly identified ORFs may encode functional microproteins, whose physiological importance remains to be thoroughly studied. Finally, expression of these diverse polyglycine proteins is toxic and causes LHCN-M2 muscle cell death, but with some differences with a higher toxicity of the uN2CpolyG, asRILpolyG and LOC6polyG proteins, notably compared to the uGIPpolyG protein or an artificial ATG polyGly protein expressing 100 glycines with no OPDM flanking sequences (Fig. 4g). Live cell tracking indicates that cell death was observed both in cells with polyGly inclusions and in cells showing a diffuse localization of these polyglycine proteins (Extended Data Fig. 4c). Moreover, formation of polyGly inclusions can be abrupt, with a diffuse localization observed during dozens of hours and sudden aggregation in minutes. Notably, some polyGly aggregates were even observed after cell death. Finally, no overt signs of apoptosis were noted, suggesting that these proteins are toxic by another pathway (Supplementary Fig. 4h,i). In conclusion, these diverse polyglycine proteins share the common properties of forming p62-positive cellular inclusions and inducing cell death, recapitulating key features of OPDM and OPML. However, these polyglycine proteins also show different biological properties (localization, half-life, aggregation, toxicity, etc.), suggesting a modulation of their central and common polyglycine core by their specific flanking amino acid sequences.

Polyglycine proteins form inclusions and are pathogenic for muscles in mice

To determine the physiological impact of these polyglycine proteins, we expressed them through a recombinant adeno-associated viral (rAAV) strategy in mouse skeletal muscles (Fig. 5a). Histological analyses of the tibialis anterior muscles up to 10 months after rAAV injection show that polyglycine proteins are toxic and promote muscle fiber atrophy with the presence of internalized or centralized nuclei, but with some striking differences among these proteins. Indeed, expression of the OPDM4 asRILpolyG, OPML LOC6polyG and OPDM3 uN2CpolyG proteins causes histological changes in 3–5 months after rAAV injection, while the OPDM2 uGIPpolyG protein shows a lesser toxicity with some muscle changes detected only 9 months after rAAV injection (Fig. 5b,c, Supplementary Note 5 and Supplementary Fig. 5a). Similarly, expression of ATG polyGly, a protein deprived of any OPDM natural bordering sequences, shows a limited and delayed pathogenicity. Analyses of p62 staining revealed numerous p62-positive cytoplasmic and intranuclear inclusions, as observed in OPDM patients (Fig. 5d and Supplementary Fig. 5b). These inclusions are eosinophilic, which is especially apparent in the uN2CpolyG expressing mice (Fig. 5b). Of interest, all OPDM polyglycine proteins form inclusions, but with some notable differences, with observation of frequent OPML LOC6polyG and OPDM3 uN2CpolyG aggregates, while ATG polyG and OPDM2 uGIPpolyG inclusions are less represented (Fig. 5e). Moreover, the localization of these polyglycine proteins varies, with the OPDM4 asRILpolyG protein more observed in nuclei compared to the other polyGly proteins (Fig. 5e). Single nuclei RNA sequencing revealed an increase in macrophages and B-cells, as well as in regenerative muscle fibers, in OPDM versus control mice (Extended Data Fig. 5a,b). These results indicate signs of inflammation and muscle regeneration consistent with myopathic changes in OPDM mice. However, these alterations were mild, with limited transcriptomic changes and only minor changes in muscle fiber types (Supplementary Fig. 5c,d and Supplementary Data 2). Correspondingly, animal performances were only slightly altered in rotarod and open field locomotor tests (Supplementary Fig. 5e,f). These data indicate that expression of polyglycine proteins in mice causes progressive muscle fiber atrophy and histological changes reminiscent of OPDM, but with specific and limited myopathic alterations, at least in the time frame and AAV-driven mouse models analyzed here. Finally, expression of the asRILpolyG and LOC6polyG proteins are remarkably deleterious as these mice die suddenly around 5–6 months or 8–9 months after rAAV injection, respectively (Fig. 5f). These mice present dilated cardiomyopathy with the presence of numerous p62-positive inclusions in cardiomyocytes (Extended Data Fig. 5c). Abundance of these aggregates mirrors their toxicity with rare ATG polyG and uGIPpolyG inclusions, an intermediate situation for uN2CpolyG, while the asRILpolyG and LOC6polyG proteins form numerous large aggregates associated with notable myopathic changes (Extended Data Fig. 5c). These data are reminiscent of the cardiac dysfunctions reported in OPDM and NIID[26,27] and led to investigate the toxicity of these polyglycine proteins in other tissues, notably the central nervous system (CNS), especially in regards of the neurological manifestations reported in individuals with OPDM2, OPDM3/NIID and OPML[4,5,10,28–30].

Polyglycine proteins form inclusions and are pathogenic for the CNS in mice

Expression of polyglycine in the mouse CNS is toxic, resulting in progressive motor performances and coordination changes, associated with a reduced lifespan (Fig. 6a,b and Extended Data Fig. 6a–e). However, we noted some differences among these diverse polyglycine proteins, with mice expressing the ATG polyG or the OPDM2 uGIPpolyG protein showing a milder pathogenicity and longer lifespan compared to mice expressing the OPML LOC6polyG or the OPDM3 uN2CpolyG proteins. Next, p62 staining revealed that these polyglycine proteins form cytoplasmic and intranuclear inclusions, recapitulating key histopathological features of OPDM, OPML and NIID (Fig. 6c, Supplementary Note 6 and Supplementary Fig. 6a). Abundance of these polyglycine inclusions mirrors their toxicity, and their accumulation is age dependent (Fig. 6c and Extended Data Fig. 6f). Finally, polyglycine expression leads to neuroinflammation and neuronal cell death, notably loss of Purkinje cells (Fig. 6d and Supplementary Fig. 6b,c). Overall, these data confirm that expression of polyglycine-containing proteins is toxic and recapitulate key features of OPDM, OPML and NIID, notably myopathic changes and neurodegeneration associated with the presence of typical p62-positive inclusions. Moreover, side-by-side analysis of these diverse polyglycine proteins revealed some notable differences in their expression, localization and toxicity, highlighting the importance of their specific flanking amino acid sequences to modulate the toxic properties of their central polyglycine core.

Porphyrin TMPyP4 alleviates aggregation and toxicity of polyglycine proteins

To alleviate the toxicity of these polyglycine proteins, we tested various compounds and identified one, the cationic porphyrin TMPyP4, that efficiently reduces their abundance and toxicity in cell cultures (Fig. 7a,b, Supplementary Note 7 and Supplementary Fig. 7a–e). RNA sequencing and mass spectrometry revealed that TMPyP4 induces only limited changes and no global transcriptomic or proteomic alterations (Supplementary Fig. 7f–i and Supplementary Data 3 and 4). Pathway analysis revealed that TMPyP4 acts principally on translation (Supplementary Data 5), which is consistent with its known inhibitory function

https://doi.org/10.1038/s41588-026-02507-z

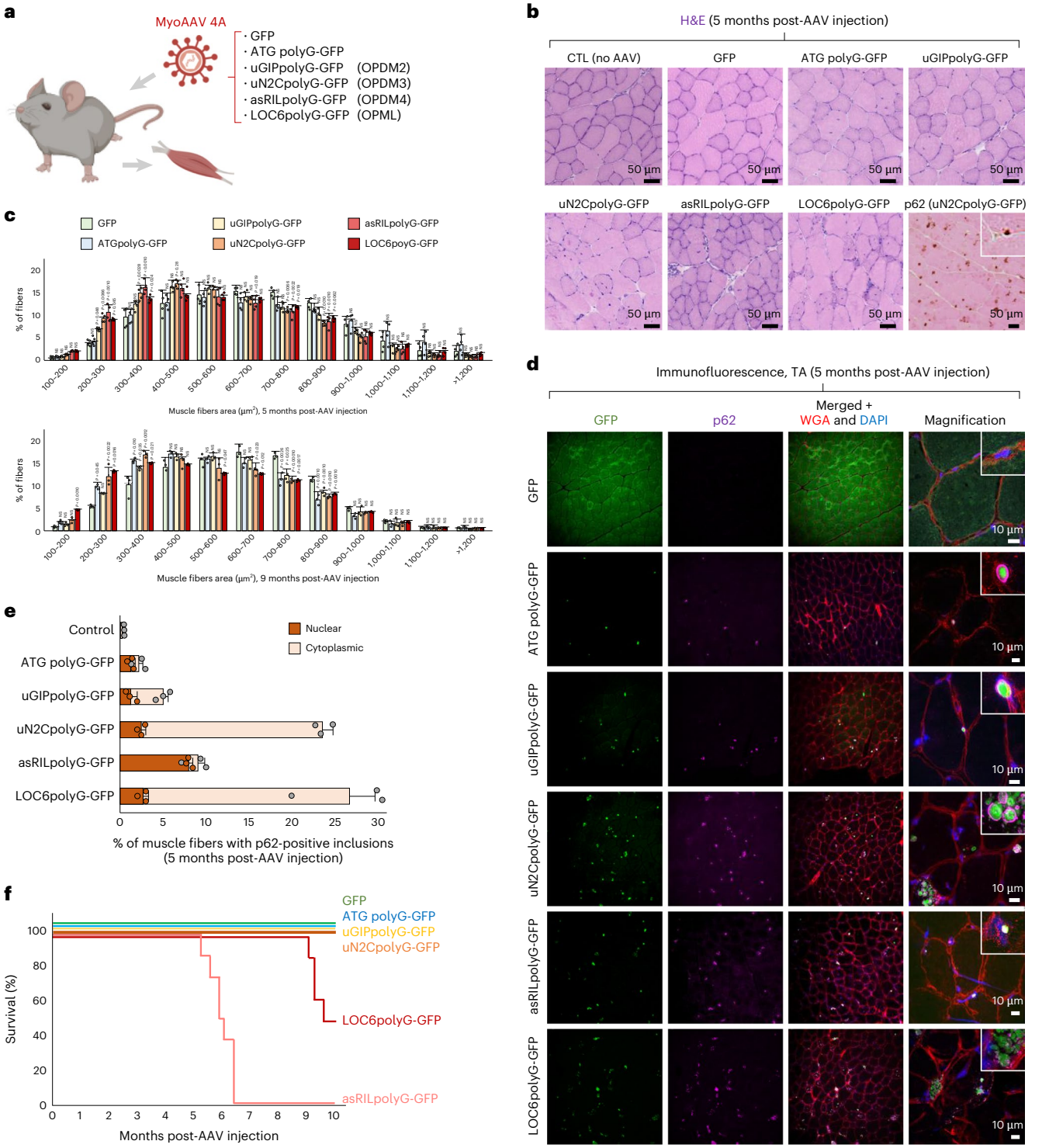


Fig. 5 | Expression of polyG proteins forms inclusions and is pathogenic in mouse muscles. a, Scheme of the AAV strategy to study OPDM/OPML polyglycine toxicity in mouse skeletal muscles. Panel **a** was created with BioRender.com. **b**, H&E staining of TA frozen sections of 5-month-old AAV-injected male mice expressing GFP, ATG polyG-GFP or GFP-tagged OPDM2 uGIPpolyG, OPDM3 uN2CpolyG, OPDM4 asRILpolyG or OPML LOC6polyG. The last image shows a representative p62 immunohistochemistry, which reveals numerous protein inclusions. Scale bars = 50 µm. **c**, Quantification of mouse TA muscle fiber area 5 months (top) or 9 months (bottom) postinjection of AAV expressing the OPDM/OPML polyglycine proteins and controls shown in **a**. Bar heights represent the mean. Error bars represent the mean ± s.e.m. Sample size, $n = 4$ mice per condition and with at least 1,000 muscle fibers counted per animal. One-way ANOVA with Tukey's post hoc test. **d**, GFP fluorescence and immunofluorescence against p62 with counterstaining of membranes by fluorescence-conjugated WGA and nuclear DNA by DAPI on frozen TA muscle sections 5-months postinjection of AAV expressing the OPDM/OPML polyglycine proteins and controls described in **a**. Scale bars = 10 µm. **e**, Quantification of GFP-positive inclusions in TA frozen sections of controls and OPDM/OPML polyglycine-expressing mice. Bar heights represent the mean and error bars represent the mean ± s.e.m. Sample size, $n = 3$ mice per condition, with at least 1,000 muscle fibers counted per animal. **f**, Kaplan–Meier survival curve of controls and OPDM/OPML polyglycine-expressing mice. Sample size, $n = 8$ mice per condition (Extended Data Fig. 5 and Supplementary Fig. 5). ANOVA, analysis of variance; H&E, hematoxylin and eosin; WGA, wheat germ agglutinin; TA, tibialis anterior.

https://doi.org/10.1038/s41588-026-02507-z

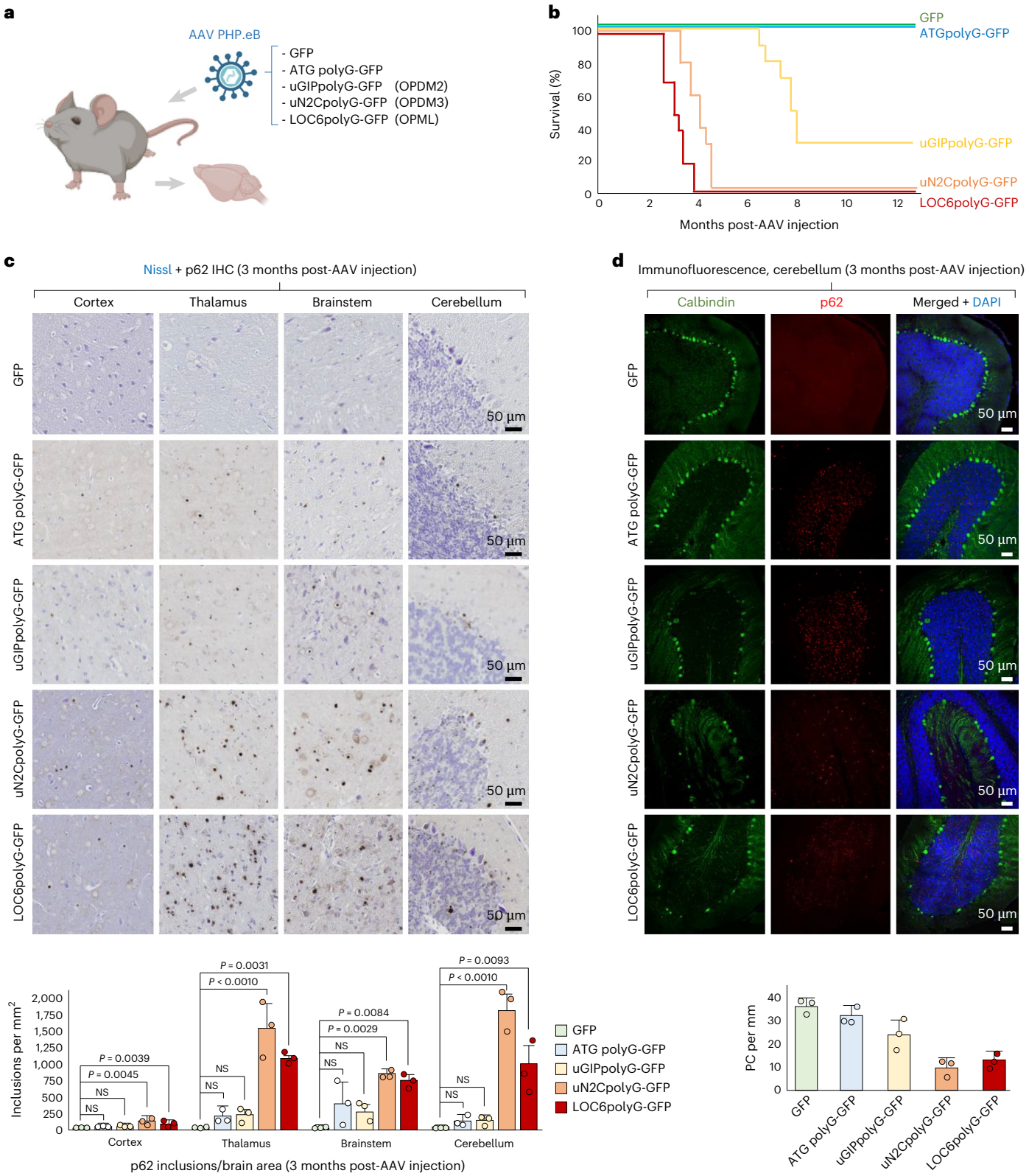


Fig. 6 | Expression of polyglycine proteins forms inclusions and is pathogenic in mouse CNS. a, Scheme of the AAV strategy to study OPDM/OPML polyglycine toxicity in mouse CNS. Panel **a** was created with BioRender.com. **b**, Kaplan–Meier survival curve of controls and OPDM/OPML polyglycine-expressing mice. Sample size, $n = 10$ mice per condition. **c**, Top, immunohistochemistry against p62 with cresyl violet (Nissl) counterstaining of various mouse brain areas 3-month postinjection of AAV expressing GFP, ATG polyG-GFP or GFP-tagged OPDM2 uGIPpolyG, OPDM3 uN2CpolyG or OPML LOC6polyG. Scale bars = 50 µm. Bottom, quantification of p62-positive inclusions. Bar heights represent the mean. Error bars represent the mean ± s.e.m. Sample size, $n = 3$ mice per condition, with at least 200 nuclei counted per brain region and per animal. **d**, Top, immunofluorescence against p62 and calbindin on the cerebellum 3-months postinjection of AAV expressing controls or OPDM/OPML GFP-tagged polyG proteins. Scale bars = 50 µm. Bottom, quantification of PC number. Bar heights represent the mean. Error bars represent the mean ± s.e.m. Sample size, $n = 3$ mice per condition, with 4 mm² cerebellum area counted per animal (Extended Data Fig. 6 and Supplementary Fig. 6). PC, Purkinje cell.

https://doi.org/10.1038/s41588-026-02507-z

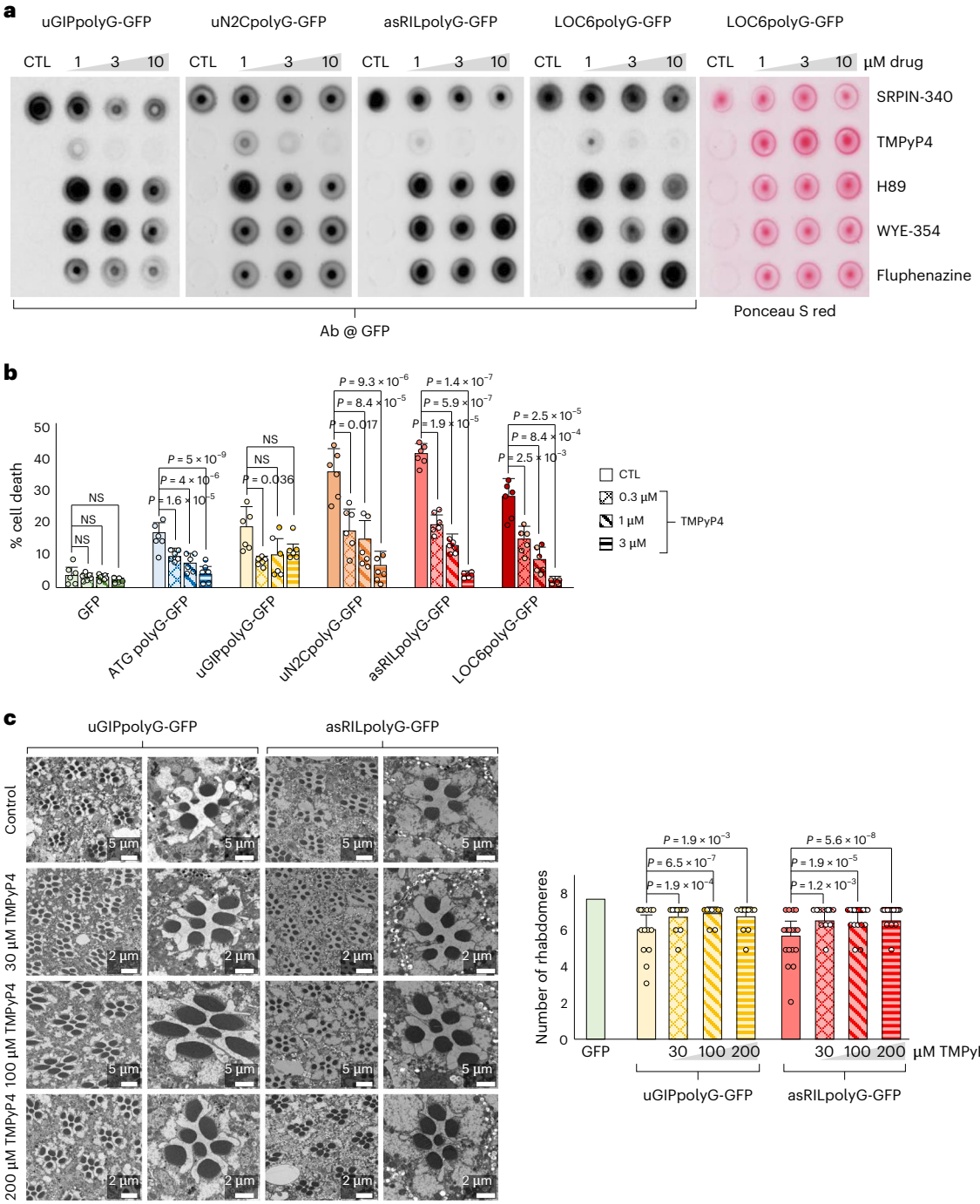


Fig. 7 | The porphyrin TMPYP4 alleviates aggregation and toxicity of polyglycine proteins. a, Dot blot against GFP or Ponceau staining of the insoluble proteins extracted from 48-h differentiated LHCN-M2 muscle cells expressing GFP-tagged ATG polyG, OPDM2 uGIPpolyG, OPDM3 uN2CpolyG, OPDM4 asRILpolyG or OPML LOC6polyG and treated overnight with the indicated drug concentration. **b**, Cell viability of LHCN-M2 muscle cells differentiated for 3 days and expressing GFP, ATG polyG-GFP or GFP-tagged OPDM2 uGIPpolyG, OPDM3 uN2CpolyG, OPDM4 asRILpolyG or OPML LOC6polyG and treated overnight with no or 0.3, 1 or 3 µM of TMPyP4. Bar heights represent the mean. Error bars represent the mean ± s.e.m. Sample size, $n = 6$ independent biological replicates. Unpaired two-tailed t test compared to each nontreated control condition. **c**, TMPyP4 ameliorates ommatidial degeneration in *Drosophila* models of polyG-expanded proteins. Left, fly eyes representative EM images of 20-day-old OPDM2 uGIPpolyG or OPDM4 asRILpolyG-expressing *Drosophila* fed with no, 30, 100 or 200 µM of TMPyP4. Scale bars = 5 µm in columns 1 and 3, and 2 µm in columns 2 and 4. Right, quantification analysis revealed that TMPyP4 significantly preserved ommatidial integrity in OPDM polyglycine-expressing flies. Bar heights represent the mean. Error bars represent the mean ± s.e.m. All quantitative data are given as the number of rhabdomeres per ommatidium. Sample size, $n = 40–56$ ommatidia from three flies per condition. One-way ANOVA with Bonferroni post hoc test (Extended Data Fig. 7 and Supplementary Fig. 7).

on the translation of GC-rich microsatellites[31,32]. Next, to investigate TMPyP4 effects in animals, we developed *Drosophila* expressing polyglycine proteins. Ubiquitous expression of uGIPpolyG (OPDM2) leads to a progressively reduced mobility and shortened lifespan, while expression of asRILpolyG (OPDM4) was particularly toxic with no, or very few, animals surviving to the adult stage (Supplementary Note 7 and Extended Data Fig. 7a,b). Expression of these polyglycine proteins in *Drosophila* eyes led to ommatidial degeneration and loss of

https://doi.org/10.1038/s41588-026-02507-z

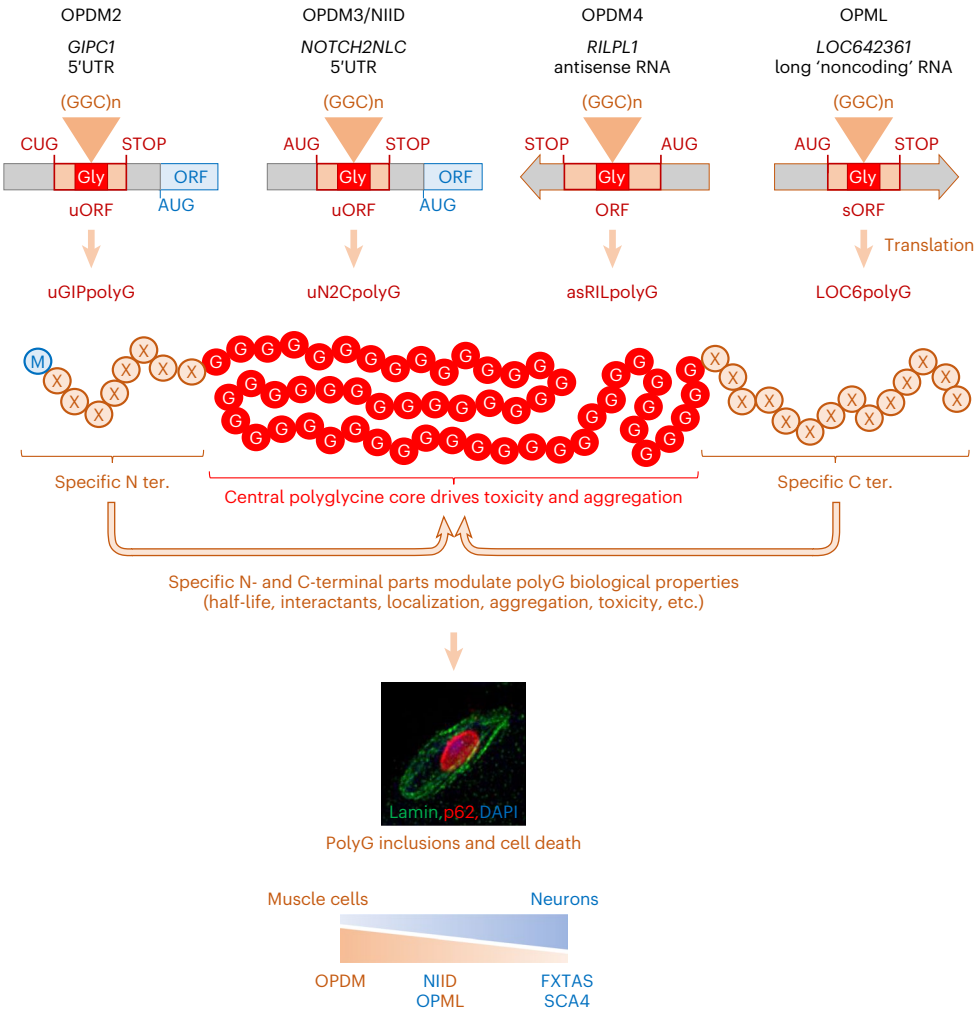


Fig. 8 | Model of polyG toxicity in OPDM/OPML neurological diseases. Expanded GGC repeats within sequences originally annotated as noncoding are embedded in previously undescribed ORFs, resulting in expression of new polyglycine-containing proteins that form protein inclusions and are toxic for neuronal and muscle cells. Of interest, toxicity and biological properties of their central polyglycine core are modulated by their bordering sequences, which are specific to each hosting ORF. Moreover, these data recall the neurodegenerative NIID, FXTAS and SCA4 disorders, suggesting existence of a wider neurological spectrum of diseases caused by polyGly proteins. FXTAS, fragile X-associated tremor/ataxia syndrome; SCA4, spinocerebellar ataxia 4.

rhabdomeres, but with a higher toxicity of the OPDM4 asRILpolyG protein compared to the OPDM2 uGIPpolyG protein (Extended Data Fig. 7c and Supplementary Fig. 7j). These results are consistent with observations in cells and mice, highlighting in a third model the importance of the specific amino acid sequences flanking the common polyglycine core of these proteins to modulate their pathogenicity. Notably, TMPyP4 corrects polyglycine toxicity, restoring normal eye structure and rhabdomeres in uGIPpolyG and asRILpolyG-expressing *Drosophila* (Fig. 7c). Overall, these data highlight that expression of polyglycine proteins reproduces the locomotor and neurodegenerative clinical features observed in the OPDM, OPML and NIID disorders, and that modulating the expression and toxicity of these polyglycine proteins could be of therapeutic interest for these neurological diseases (Fig. 8).

Discussion

OPDM, NIID and OPML are neurological diseases caused by similar GGC repeat expansions, but embedded in sequences annotated as noncoding in diverse genes (*LOC642361*, *LRP12*, *GIPC1*, *NOTCH2NLC*, *RILPL1* and *ABCD3*). Here we found that some of these GGC repeats are embedded in previously uncharted ORFs and are translated into new polyglycine-containing proteins that form p62-positive protein inclusions and are toxic in cell and animal models. In addition, this work clarifies some prerequisites for these GGC microsatellite expansions to be translated into stable and detectable polypeptides, notably the necessity for these repeats to be (1) located in an RNA transcript, which may include ill-described sequences annotated by default as noncoding, but with the requirement that this RNA is exported within the cytoplasm where translation occurs; (2) embedded within an ORF, with the crucial point to be in frame with an upstream cognate ATG or near-cognate (ACG, CTG, GTG or TTG) start codon with its associated Kozak motif and (3) of sufficient size for the encoded polypeptide to be stable and thus reliably detected in tissues (Supplementary Note 8).

Overall, these data are reminiscent of the fragile X-associated tremor/ataxia syndrome and spinocerebellar ataxia 4, where GGC repeat expansions, respectively, located in a small upstream ORF (uORF) of the *FMR1* gene or within the main ORF of the ZFHX3 protein, are translated into polyglycine-containing proteins, which are toxic and accumulate in p62-positive inclusions[12,33–36]. Altogether, these observations support the existence of a new group of human disorders, the polyG (or polyGly) diseases, where similar expansions of GGC repeats are embedded in diverse, previously poorly characterized, ORFs and consequently translated into various polyglycine-containing proteins, which form protein inclusions and are toxic for muscle and neuronal cells. Moreover, this work reinforces the proposition that

OPDM, OPML, NIID, fragile X-associated tremor/ataxia syndrome and spinocerebellar ataxia 4 belong to a new continuum of neuromuscular and neurodegenerative diseases with similar genetic causes, overlapping clinical and histopathological presentations, and a shared mechanism of pathogenicity (Fig. 8). This recalls the pioneering discovery of the polyQ group of diseases where similar expansions of CAG repeats, embedded within the main ORFs of diverse genes, are translated in various polyQ-containing proteins that form protein aggregates and are pathogenic in various neurodegenerative disorders.

However, this work also raises several questions, notably whether other microsatellite expansions, embedded in yet unrecognized ORFs and thus translated into potentially toxic proteins, remain to be identified (Supplementary Note 8). Patent candidates would be the GGC repeat expansions found in the *LRP12* and *ABCD3* genes, causing OPDM1 and OPDM5, respectively[4,11]. Whether these microsatellite mutations located in sequences annotated as noncoding are nonetheless translated into new and toxic proteins is an exciting question for future studies. It also remains to be determined how these polyglycine proteins are toxic. They form cellular inclusions, which is consistent with the known self-aggregation properties of glycine homopolypeptides that form amyloid-like fibrils[37,38]. However, whether these polyglycine proteins are toxic under their soluble or aggregated form is unknown. Similarly, it is unclear whether their localization is important for their toxicity, and, in that aspect, how these polyglycines accumulate in the nucleus in the absence of an evident nuclear localization signal remains to be clarified (Supplementary Note 8). Another point of interest is the side-by-side comparisons of these diverse polyglycine proteins, which reveal that their abilities to form inclusions and promote cell death originate from their central and common polyglycine core, while their localization, expression, half-life, aggregation and interactions with other proteins are modulated by their specific N-terminal and C-terminal flanking sequences, originating from their hosting ORFs. It remains to be investigated thoroughly how these flanking sequences modulate the toxicity of these polyglycine proteins.

In conclusion, this work highlights a unified pathogenic mechanism for the skeletal muscle and CNS dysfunctions observed in individuals with OPDM, NIID and OPML, where GGC repeat expansions are embedded in previously unrecognized ORFs and consequently translated into new and toxic polyglycine-containing proteins. Consistent with a shared mechanism of toxicity, this work also provides proof of concept that a common therapeutic approach may be worth pursuing for these neurological disorders.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-026-02507-z.

References

1. Wright, S. E. & Todd, P. K. Native functions of short tandem repeats. *eLife* **12**, e84043 (2023).
2. Depienne, C. & Mandel, J. L. 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges?. *Am. J. Hum. Genet.* **108**, 764–785 (2021).
3. Malik, I., Kelley, C. P., Wang, E. T. & Todd, P. K. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **22**, 589–607 (2021).
4. Ishiura, H. et al. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat. Genet.* **51**, 1222–1232 (2019).
5. Sone, J. et al. Long-read sequencing identifies GGC repeat expansions in *NOTCH2NLC* associated with neuronal intranuclear inclusion disease. *Nat. Genet.* **51**, 1215–1221 (2019).
6. Deng, J. et al. Long-read sequencing identified repeat expansions in the 5′UTR of the *NOTCH2NLC* gene from Chinese patients with neuronal intranuclear inclusion disease. *J. Med. Genet.* **56**, 758–764 (2019).
7. Deng, J. et al. Expansion of GGC repeat in *GIPC1* is associated with oculopharyngodistal myopathy. *Am. J. Hum. Genet.* **106**, 793–804 (2020).
8. Yu, J. et al. The CGG repeat expansion in *RILPL1* is associated with oculopharyngodistal myopathy type 4. *Am. J. Hum. Genet.* **109**, 533–541 (2022).
9. Shi, Y. et al. CGG repeat expansion in *LOC642361*/*NUTM2B-AS1* typically presents as oculopharyngodistal myopathy. *J. Genet. Genomics* **51**, 184–196 (2024).
10. Yu, J. et al. The GGC repeat expansion in *NOTCH2NLC* is associated with oculopharyngodistal myopathy type 3. *Brain* **144**, 1819–1832 (2021).
11. Cortese, A. et al. A CCG expansion in *ABCD3* causes oculopharyngodistal myopathy in individuals of European ancestry. *Nat. Commun.* **15**, 6327 (2024).
12. Figueroa, K. P. et al. A GGC-repeat expansion in *ZFHX3* encoding polyglycine causes spinocerebellar ataxia type 4 and impairs autophagy. *Nat. Genet.* **56**, 1080–1089 (2024).
13. Vegezzi, E. et al. Neurological disorders caused by novel non-coding repeat expansions: clinical features and differential diagnosis. *Lancet Neurol.* **23**, 725–739 (2024).
14. Satoyoshi, E. & Kinoshita, M. Oculopharyngodistal myopathy. *Arch. Neurol.* **34**, 89–92 (1977).
15. Ogasawara, M. et al. Intranuclear inclusions in skin biopsies are not limited to neuronal intranuclear inclusion disease but can also be seen in oculopharyngodistal myopathy. *Neuropathol. Appl. Neurobiol.* **48**, e12787 (2022).
16. Ishiura, H., Tsuji, S. & Toda, T. Recent advances in CGG repeat diseases and a proposal of fragile X-associated tremor/ataxia syndrome, neuronal intranuclear inclusion disease, and oculophryngodistal myopathy (FNOP) spectrum disorder. *J. Hum. Genet.* **68**, 169–174 (2023).
17. Boivin, M. & Charlet-Berguerand, N. Trinucleotide CGG repeat diseases: an expanding field of polyglycine proteins? *Front. Genet.* **13**, 843014 (2022).
18. Gao, F. B., Richter, J. D. & Cleveland, D. W. Rethinking unconventional translation in neurodegeneration. *Cell* **171**, 994–1000 (2017).
19. Kearse, M. G. & Wilusz, J. E. Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.* **31**, 1717–1731 (2017).
20. Kozak, M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol. Cell. Biol.* **9**, 5073–5080 (1989).
21. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
22. Boivin, M. et al. Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: the polyG diseases. *Neuron* **109**, 1825–1835 (2021).
23. Zhong, S. et al. Upstream open reading frame with *NOTCH2NLC* GGC expansion generates polyglycine aggregates and disrupts nucleocytoplasmic transport: implications for polyglycine diseases. *Acta Neuropathol.* **142**, 1003–1023 (2021).
24. Liu, Q. et al. Expression of expanded GGC repeats within *NOTCH2NLC* causes behavioral deficits and neurodegeneration in a mouse model of neuronal intranuclear inclusion disease. *Sci. Adv.* **8**, eadd6391 (2022).
25. Brais, B. et al. Short GCG expansions in the *PABP2* gene cause oculopharyngeal muscular dystrophy. *Nat. Genet.* **18**, 164–167 (1998).

26. Gu, X. et al. *NOTCH2NLC*-related oculopharyngodistal myopathy type 3 with cardiomyopathy and nephropathy. *Muscle Nerve* **67**, E18–21E (2023).
27. Pan, Y. et al. Expression of expanded GGC repeats within *NOTCH2NLC* causes cardiac dysfunction in mouse models. *Cell Biosci.* **13**, 157 (2023).
28. Fan, Y. et al. *GIPC1* CGG repeat expansion is associated with movement disorders. *Ann. Neurol.* **91**, 704–715 (2022).
29. Pan, Y. et al. Assessment of GGC repeat expansion in *GIPC1* in patients with Parkinson's disease. *Mov. Disord.* **37**, 1557–1559 (2022).
30. Murayama, A. et al. Sequential development of parkinsonism in two patients with oculopharyngodistal type myopathy in *GIPC1*-related repeat expansion disorder. *Neuromuscul. Disord.* **44**, 104465 (2024).
31. Green, K. M. et al. High-throughput screening yields several small-molecule inhibitors of repeat-associated non-AUG translation. *J. Biol. Chem.* **294**, 18624–18638 (2019).
32. Mori, K. et al. The porphyrin TMPyP4 inhibits elongation during the noncanonical translation of the FTLD/ALS-associated GGGGCC repeat in the *C9orf72* gene. *J. Biol. Chem.* **297**, 101120 (2021).
33. Todd, P. K. et al. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**, 440–455 (2013).
34. Sellier, C. et al. Translation of expanded CGG repeats into FMRpolyG is pathogenic and may contribute to fragile X tremor ataxia syndrome. *Neuron* **93**, 331–347 (2017).
35. Wallenius, J. et al. Exonic trinucleotide repeat expansions in *ZFHX3* cause spinocerebellar ataxia type 4: a poly-glycine disease. *Am. J. Hum. Genet.* **111**, 82–95 (2024).
36. Chen, Z. et al. Adaptive long-read sequencing reveals GGC repeat expansion in *ZFHX3* associated with spinocerebellar ataxia type 4. *Mov. Disord.* **39**, 486–497 (2024).
37. Lorusso, M., Pepe, A., Ibris, N. & Bochicchio, B. Molecular and supramolecular studies on polyglycine and poly-L-proline. *Soft Matter* **7**, 6327 (2011).
38. Plumley, J. A., Tsai, M. I. & Dannenberg, J. J. Aggregation of capped hexaglycine strands into hydrogen-bonding motifs representative of pleated and rippled β-sheets, collagen, and polyglycine I and II crystal structures. A density functional theory study. *J. Phys. Chem. B* **115**, 1562–1570 (2011).

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Université de Strasbourg, CNRS UMR 7104, Inserm UMR-S, Illkirch, France. [2]Department of Neurology, Peking University First Hospital, Beijing, China. [3]Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry (NCNP), Tokyo, Japan. [4]Department of Neurology and Institute of Neurology of First Affiliated Hospital, Institute of Neuroscience, and Fujian Key Laboratory of Molecular Neurology, Fujian Medical University, Fuzhou, China. [5]Rare Diseases Medical Center, Peking University First Hospital, Beijing, China. [6]These authors contributed equally: Manon Boivin, Jiaxi Yu. [7]These authors jointly supervised this work: Manon Boivin, Jianwen Deng, Nicolas Charlet-Berguerand. ✉e-mail: boivinm@igbmc.fr; jianwendeng@pkufh.com; ncharlet@igbmc.fr

Methods

Human samples

Human muscle samples were sampled with the informed consent of individuals and families and approved by the Institutional Review Board of the Peking University First Hospital, First Affiliated Hospital of Fujian Medical University and National Center of Neurology and Psychiatry. This study was approved by the ethics committees of Peking University First Hospital, First Affiliated Hospital of Fujian Medical University and National Center of Neurology and Psychiatry, and all procedures were conducted in accordance with relevant guidelines and regulations. Muscle biopsy samples from patients with OPDM, OPML and NIID and age-matched control participants were examined. All clinical materials were obtained for diagnostic purposes after informed consent was provided. Before this study, all samples had been analyzed using routine histology techniques and EM. Fresh-frozen samples were stored at −80 °C until use.

Mice

All animal work were performed with approval from the IGBMC/ICS Animal Care Committee and the French agency for research on animals, Direction générale de la recherche et de l'innovation (DGRI), authorization APAFIS33864-2021111217327782. C57BL/6 wild-type male mice were retro-orbitally AAV-injected at 2 months and then housed for 6–8 months in a temperature-controlled room (19–22 °C) with a 12-h light/12-h dark cycle and free access to food and water. Mice were killed by carbon dioxide (CO_2) inhalation to dissect the different skeletal muscles, heart and brain, which were subsequently frozen for molecular biology, freezing using prechilled isopentane or paraformaldehyde (PFA)-fixed and embedded in paraffin for histology.

Cell cultures

U2OS and HEK293 cells were grown in DMEM containing 1 g l^{-1} glucose with 10% FCS and gentamicin at 37 °C in 5% CO_2. LHCN-M2 cells were grown in DMEM containing 4.5 g l^{-1} glucose with 20% FCS, without PyrNa/M199, supplemented with 25 µg ml^{-1} fetuin, 5 mg ml^{-1} hEGF, 0.5 mg ml^{-1} human bFGF, 5 µg ml^{-1} human insulin, 0.2 µg ml^{-1} dexamethasone and gentamicin at 37 °C in 5% CO_2. Differentiation of the LHCN-M2 cells was induced by serum removal. U2OS T-Rex cells (Thermo Fisher Scientific) stably expressing Nup50-Cherry were lipofectamine-transfected with Pci1-linearized pcDNA3 expressing Nup50 fused to the mCherry and selected for neomycin resistance for 2 weeks.

Constructs

Human *GIPC1* exon 1, antisense *RILPL1* and *LOC642361* lncRNA sequences upstream of their GGC repeats were cloned into pcDNA3.1 fused to GFP lacking its start codon (ΔATG), with each insert cloned in all three reading frames. Mutations of the ATG or CTG start codons, or within ORFs, were achieved by inverse PCR or by oligonucleotide ligations. GIPC1 uORF, antisense RILPL1 and LOC642361 small ORFs with either 12 or 100 optimized GGN repeats were synthesized by GenScript and fused to GFP into a pAAV2-CAG vector. To ensure repeat expansions stability, all GGC repeat-containing plasmids were transformed into STBL3 bacterial strain (Invitrogen), and all constructs were confirmed by sequencing.

Cell transfection and treatments

For transient transfection, cells were plated and transfected the following day in medium with 0.1% FCS or without serum for LHCN-M2 cells for 5 h using Lipofectamine 2000 (Thermo Fisher Scientific). After 5 h to 4 days post-transient transfection, cells were analyzed by live imaging, immunofluorescence, real-time qPCR (RT-qPCR), cell viability, dot blotting or western blotting. For treatments, LHCN-M2 cells were incubated overnight with indicated concentration of SRPIN340, H-89, fluphenazine TMPyP4, 5,10,15,20-tetra(4-pyridyl)-21H,23H-porphine, 5,10,15,20-tetraphenyl-21H,23H-porphine (Sigma-Aldrich), 5,10,15,20-tetrakis(4-aminophenyl)-21H,23H-porphine, 5,10,15,20-tetrakis (4-ethynylphenyl)-21H,23H-porphine, 5,10,15,20-tetra(pyridin-2-yl) porphyrin, (porphyrin-5,10,15,20-tetrayltetrakis(benzene-4,1-diyl)) tetraboronic acid, 4,4′,4″,4‴-(21H,23H-porphine-5,10,15,20-tetrayl) tetrakis-phenol (BLDpharm). For cycloheximide treatment, HEK293 cells were treated 1 day post-transfection with 50 µg ml^{-1} of cycloheximide for 1, 3, 8 or 24 h.

Cell viability assay

LHCN-M2 cells were transiently transfected for 72 h with the different polyglycine-expressing constructs and treated overnight with the indicated drug concentration. After the addition of 0.5 µM of TO-PRO-3 (Thermo Fisher Scientific), live cells were imaged using the CX7 Cellular Imaging System (25 fields per well at ×10 magnification), followed by a cell-to-cell analysis using Cellomics HCS Studio software (CellHealth Bioapplication). Transfected cells were detected using GFP staining and dead cells were identified using TO-PRO-3 intensity within cell mask.

FACS analysis

HEK293 cells transfected for 24 h with the different frame constructs were trypsinized, centrifuged for 5 min at 100*g* and resuspended in 500 µl of PBS. Cells were analyzed by the BD LSRFortessa X-20 and results were construed by FlowJo.

Western blotting

Proteins were denatured for 3 min at 95 °C, separated on 4–12% Bis–Tris gel (NuPAGE), transferred on nitrocellulose membranes (Amersham Protan), blocked with 5% nonfat dry milk in Tris-buffered saline with 0.1% Tween-20 (TBS-T), incubated with anti-GFP (Abcam, ab290; 1:10,000), anti-GFP (Abcam, ab1218; 1:10,000), mCherry (Abcam, ab167453; 1:5,000), GAPDH (Abcam, ab8245; 1:10,000), KU70 (SantaCruz, sc-56129; 1:5,000), KU80 (Abcam, ab119935; 1:10,000), RPL10A (Thermo Fisher Scientific, MA5-27171; 1:3,000), RPL36 (Abcam, ab241584; 1:10,000), HA (Abcam, ab130275; 1:5,000), uGIP pAb or uN2C pAb (rabbit polyclonal homemade; 1:1,000), uN2C 4D12, asRIL 2D8 or 4B9 or LOC6 2E8 (mouse monoclonal homemade; 1:1,000) in TBS-T with 5% nonfat dry milk overnight at 4 °C. The membranes were washed thrice and incubated with antirabbit or mouse peroxidase antibody Cell Signaling Technology, 7074S or 7076S; 1:10,000) 1 h in TBS, followed by washing and ECL Prime chemiluminescence revelation kit (Millipore).

Dot blotting

LHCN-M2 cells transfected with ATG polyG-GFP, uGIPex2polyG-GFP, uGIPex4polyG-GFP, uN2CpolyG-GFP, asRILpolyG-GFP or LOC6polyG-GFP constructs during 48 h were scraped and centrifuged for 10 min at 700*g* at 4 °C. The cell pellet was frozen, resuspended in 200 µl of RIPA, frozen and centrifuged for 10 min at 20,000*g* at 4 °C. The pellet was resuspended in 200 µl of 2× Laemmli buffer, sonicated for 5 s at 20% amplitude and warmed for 3 min at 95 °C. Proteins were directly loaded on nitrocellulose membranes (Amersham Protan), washed twice with Towbin buffer, blocked with 5% nonfat dry milk in TBS-T and incubated with anti-GFP (Abcam, ab290; 1:10,000) in TBS-T with 5% nonfat dry milk overnight at 4 °C. The membranes were washed thrice and incubated with antirabbit peroxidase antibody (Cell Signaling Technology, 7074S; 1:10,000) 1 h in TBS, followed by washing and ECL Prime chemiluminescence revelation kit (Millipore).

Lysostaphin treatment

HEK293 cells transfected with uGIPpolyG-GFP, asRILpolyG-GFP or LOC6polyG-GFP constructs were scraped and centrifuged for 10 min at 700*g* at 4 °C. The cell pellet was resuspended in 300 µl of RIPA and centrifuged for 10 min at 20,000*g* at 4 °C. Thirty microliters of supernatant extract were incubated with 10 ng $µl^{-1}$ of lysostaphin (Prospec,

ENZ-269) for 1–20 min at 37 °C. Laemmli buffer was added to the mix and proteins were analyzed by western blot.

AAV production and retro-orbital injection

Recombinant AAV were generated by triple-transfection of HEK293T/17 cell line with the pAAV expression plasmids (expressing−GFP, ATG polyG-GFP, uGIPex2polyG-GFP, uGIPex4polyG-GFP, uN2CpolyG-GFP, asRILpolyG-GFP or LOC6polyG-GFP), the auxiliary plasmid pHelper (Agilent) encoding the adenovirus helper functions and the capsid plasmid pUCmini-iCAP-PHP.eB (Addgene, 103005) or pMyoAAV-4A. The pMyoAAV-4A was previously generated by the IGBMC facility using available literature[39]. The rAAV were collected from cell lysate and treated with Benzonase (Merck) at 100 U ml^{-1}. Recombinant vectors were purified by Iodixanol gradient ultracentrifugation (OptiPrep, Axis Shield), followed by dialysis and concentration (Amicon Ultra-15 Centrifugal Filter Device 100K) against sterile PBS (Dulbecco's PBS containing 0.5 mM $MgCl_2$). Particles were quantified by RT-PCR and vector titers were expressed as viral genomes per ml (vg ml^{-1}). The 2-month-old C57BL/6 male mice were injected retro-orbitally with 100 µl of sterile NaCl with 1.5×10^{13} vg kg^{-1} of AAV.

Mouse phenotyping

Rotarod test (Bioseb) was performed with three testing trials during which the rotation speed accelerated from 4 to 40 rpm in 5 min. Trials were separated by 10–15-min interval. The average latency was used as index of motor coordination performance.

Notched bar test. Mice were tested under 100-lux lighting on a 2-cm-wide and 50-cm-long natural wooden piece notched bar comprising 12 platforms of 2 cm, spaced by 13 gaps of 2 cm and bearing a 6-cm^2 terminal platform. Animals had to cross the notched bar twice for training and thrice for the test. Every instance of a back paw passing through the gap was counted as an error, and the global error percentage was calculated.

Open field test. Mice were tested in automated open fields, each of which was virtually divided into central and peripheral regions. The open fields were placed in a room homogeneously illuminated at 120 lux. Each mouse was placed in the periphery of the open field and allowed to explore freely the apparatus for 30 min, with the experimenter out of the animal's sight. The distance traveled, the number of rears and time spent in the central and peripheral regions were recorded over the test session. The number of entries and the percent time spent in the center area are used as an index of emotionality/anxiety.

Immunofluorescence on PFA-fixed cells

Glass coverslips containing plated cells were fixed for 15 min in PBS with 4% PFA, washed with PBS and incubated in PBS with 0.5% Triton X-100 for 5 min. The coverslips were incubated during 1 h with primary antibody against p62 (Abcam, ab56416; 1:1,000), p62 (Abcam, ab109012; 1:1,000), desmin (Abcam, ab32362; 1:500), lamin A/C (Abcam, ab238303; 1:1,000), uGIP pAb or uN2C pAb (rabbit polyclonal homemade; 1:100) and uN2C 4D12, asRIL 2D8 or 4B9 or LOC6 2E8 (mouse monoclonal homemade; 1:100). After washing with PBS, the coverslips were incubated with donkey antimouse or donkey antirabbit secondary antibodies conjugated with Alexa 488, CY3 or CY5 (Jackson Immunoresearch; 1:500) for 1 h, washed twice with PBS and incubated for 3 min in PBS/DAPI (1:10,000 dilution). Coverslips were rinsed twice before mounting in Pro-Long media (Molecular Probes).

Immunofluorescence or immunochemistry on PFA-fixed tissue sections

For immunochemistry followed by cresyl violet counterstaining, buffers were DEPC-treated and autoclaved. Brain sections were deparaffinized for 10 min in Sub-X (Leica) and dehydrated as follows: ethanol 100% (10 min), ethanol 90% (5 min), ethanol 70% (5 min) and rinsed in water. Antigen retrieval was done in pressure cooker in 10 mM Tris pH 9, 1 mM EDTA or 10 mM sodium citrate pH 6. For immunochemistry, endogenous peroxidase activity was blocked for 15 min with 3% H_2O_2. Slides were blocked for 1 h with PBS, 0.5% Triton X-100 and 5% horse serum for immunofluorescence of PBS with 0.1% Tween-20 and 5% BSA for immunochemistry followed by overnight incubation at 4 °C with primary antibody against Calbindin (Cell Signaling Technology, 13176S; 1:800), GFAP (Abcam, ab68428; 1:10,000), p62 (Abcam, ab56416; 1:1,000), p62 (Cell Signaling Technology, 23214S; 1:500) or tyrosine hydroxylase (Abcam, ab112; 1:2,000). For immunofluorescence, slides were washed with PBS containing 0.1% Triton X-100, incubated with donkey antimouse or donkey antirabbit secondary antibodies conjugated with Alexa 488 or CY3 (Jackson Immunoresearch; 1:500) for 1 h, washed twice with PBS containing 0.1% Triton X-100 and incubated for 3 min in PBS/DAPI (1:1,000 dilution). Slides were rinsed twice in PBS before mounting in Pro-Long media (Molecular Probes). For immunochemistry, slides were washed with PBS containing 0.1% Tween-20, incubated with horse antimouse or antirabbit coupled to peroxidase (Vector, MP-7402 or MP-7401) for 30 min, washed with PBS containing 0.1% Tween-20 and then revealed by DAB EqV substrate (Vector, SK-4103) under binocular magnifier. The reaction was stopped by immersing the slide in PBS. Then, the slides were washed for 15 min in water, and stained in 1% cresyl violet solution at 55 °C for 10 min. Slides were washed in water, quickly dehydrated in 100% ethanol, immersed in Sub-X and mounted in CV Ultra mounting medium (Leica).

Immunofluorescence or immunochemistry on isopentane-frozen sections

For immunochemistry, endogenous peroxidase activity was blocked for 15 min with 3% H_2O_2. Muscle sections were blocked for 1 h with PBS and 3% BSA and incubated overnight at 4 °C with primary antibody directed against p62 (Abcam, ab109012; 1:1,000), p62 (Cell Signaling Technology, 23214S; 1:500), lamin A/C (Abcam, ab238303; 1:1,000), lamin B1 (Proteintech, 12987-1-AP; 1:500), type I fibers (DSHB, BA-D5; 1:50), type IIa fibers (DSHB, SC-71; 1:50), type IIb fibers (DSHB, BF-F3; 1:50), uGIPpolyGly pAb or uN2C pAb (rabbit polyclonal homemade; 1:100) and uGIPpolyAla 1A7 or 3G4, uN2C 4D12, asRIL 2D8 or 4B9 or LOC6 2E8 (mouse monoclonal homemade; 1:100). After washing with PBS, the slides were incubated with donkey antimouse or donkey antirabbit secondary antibodies conjugated with Alexa 488, CY3 or CY5 (Jackson Immunoresearch; 1:500), goat antimouse IgM DyLight 405, IgG2b CY3 and IgG1 CY5 (Jackson Immunoresearch; 1:100) and wheat germ agglutinin conjugated to Alexa 555 (1:300) for 1 h, washed twice with PBS and incubated for 3 min in PBS/DAPI (1:1,000 dilution) for immunofluorescence. Concerning immunochemistry, slides were washed in PBS, incubated with horse antirabbit coupled to peroxidase (Vector, MP-7401) for 30 min, washed in PBS and revealed by DAB EqV substrate (Vector, SK-4103). The reaction was stopped by immersing the slide in PBS. Then, the slides were washed for 15 min in water, hematoxylin and eosin stain after the constructor's instructions (Abcam, ab245880). For succinate dehydrogenase activity, tissue sections were incubated in a succinate dehydrogenase reaction mixture (1.5 mM nitroblue tetrazolium, 5 mM EDTA, 48 mM succinic acid, 750 µM sodium azide, 30 mM methyl−phenylmethlyl sulfate, phosphate buffered to pH 7.6). Slides were immersed in Sub-X and mounted in CV Ultra mounting medium (Leica).

Image acquisition

Slides were imaged by spinning disk Yokogawa CSU W1 mounted with Leica Dmi 8 microscope with ×20 or ×63 objectives or by Zeiss Axioscan 7 scanner with ×20 objective for immunohistochemistry or ×40 objective for immunofluorescence. For live-cell imaging, U2OS Nup50-mCherry was plated in glass-bottom plates and transfected for 5 h before the acquisition started. Images were taken every 30 min for 15 h with the

Yokogawa CSU W1 mounted with Leica DMi 8 microscope with ×63 objectives for videos or for 65 h with the Zeiss Axio Observer Z1 with ×20 objectives for cell death quantification ($n = 3$, >100 cells counted).

Correlative light and EM

Cells were grown to near confluency directly on carbon-coated sapphire disks (3 × 0.05 mm; engineering office M. Wohlwend GMBH). The sapphire disks were then transferred to 300-µm deep flat carriers and subjected to high-pressure freezing with the HPM10 BALTEC apparatus. Automated freeze substitution (AFS) was performed in the chamber of an AFS2 device (Leica Microsystems GmbH). The samples were kept at −90 °C in dry acetone containing 0.1% uranyl acetate for 24 h. The temperature was gradually increased to −45 °C at a rate of 5 °C per hour, followed by 5 h at −45 °C. The samples were washed with pure acetone and infiltrated with graded concentrations of Lowicryl HM20. Polymerization was achieved by ultraviolet light exposure at −25 °C for 48 h, followed by an additional 9 h at room temperature (20 °C). Ultrathin sections were cut 90 nm with Leica Ultracut, picked up on 200 mesh copper grids coated with a carbon film. Sections were viewed on a spinning-disk Yokogawa X1 microscope equipped with a Nikon TI2 microscope to locate the fluorescence signal. Then, after poststained for 10 min in 2% aqueous UA and 5 min in lead citrate, sections were viewed on a Tecnai G2-20 transmission EM operated at 120 kV, and images were acquired on a TVIPS TemCam F416 camera.

Muscle fiber segmentation

Hematoxylin and eosin staining or wheat germ agglutinin channel for fluorescence images were used to segment muscle fibers. The muscle fibers were segmented from Qupath with Cellpose. Then, the fiber classification was calculated with a Python script and reimported into Qupath to display the result.

Transgenic constructs and *Drosophila* strains

The uORFs of human *GIPC1* and antisense *RILPL1* with either 10 or 100 GGN repeats were subcloned into the attB-pUAST vector and integrated into the attP2 site of phiC31 stocks through standard microinjection procedures. Transgenic *Drosophila* lines expressing either GFP vector control, uGIP-GFP, uGIPpolyG-GFP, asRIL-GFP or uRILpolyG-GFP were successfully generated. All Gal4 driver lines were acquired from the Bloomington Drosophila Stock Center. The fly strains were maintained under standard conditions at 25 °C on cornmeal agar medium, with a regulated 12-h light/12-h dark cycle.

Fly EM analysis

Samples were collected and fixed in 2.5% glutaraldehyde at 4 °C overnight. Subsequently, the samples were sectioned using a Leica EM UC6/FC6 Ultramicrotome. To verify the proper orientation and quality of the sections, they were stained with toluidine blue. The selected sections were then transferred to copper grids and counterstained with uranyl acetate and lead citrate to enhance contrast. Finally, the prepared samples were imaged using an EM.

Fly climbing assay and lifespan assay

Flies were separated by sex within 24 h of hatching and transferred to fresh vials every 4 days throughout the experimental period. Vials containing 20–30 flies of the same genotype were used. For climbing assays, flies were gently tapped to the vial bottom, and the number crossing the 5 cm mark within 15 s was recorded. Each trial was repeated five times, with mean and s.e. calculated. Lifespan assays used two sex-matched and age-matched groups. Climbing ability and lifespan were assessed and recorded every 5 days.

Fly porphyrin TMPyP4 administration

Porphyrin TMPyP4 was obtained from Selleck Chemicals and stored as a 1 mM stock solution at −20 °C. For *Drosophila* studies, the compound was dissolved in sterile water to achieve final concentrations of 30 µM, 100 µM and 200 µM. Before each experiment, fresh dilutions were prepared in *Drosophila* cornmeal agar medium. The drug was administered starting from the egg stage, and adult flies were transferred to fresh vials containing porphyrin TMPyP4 every 3 days.

Quantification and statistical analysis

To eliminate bias, image or animal analyses were either completely automated or blinded. All statistical analyses were performed using Excel (Microsoft) and an online web statistical calculator (Astatsa). Experiments are represented as either mean ± s.e.m. or box-and-whisker plots with box upper and lower limits representing the 25th and 75th quartiles, respectively, and the whiskers depicting the lowest and highest data points and the horizontal line through the box represent the median. The statistical tests used are two-tailed unpaired Student *t* test or one-way analysis of variance with post hoc Tukey's honestly significant difference test. Sample sizes were determined by past experiments and to minimize the number of mice used. No statistical method was used to determine whether data meet assumptions of the statistical approach. Detailed statistical information, including the statistical test, measures, number '*n*' of animals, biological replicates and/or assays, are indicated in Figs. 1, 4, 5, 6 and 7 and their respective legends.

Additional methods

Details of RT-qPCR, mass spectrometry and RNA sequencing are presented in Supplementary Note.

Reporting summary

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

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium through the PRIDE[40] partner repository with the dataset identifier PXD071983. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus[41] and are accessible through GEO Series accession GSE314684. Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Nicolas Charlet-Berguerand (ncharlet@igbmc.fr). All unique reagents (for example, polyG DNA constructs, antibodies, etc.) generated in this study are available from the Lead Contact under MTA subject to restrictions from commercial source. Other source data are available from the corresponding author upon reasonable request. Source data are provided with this paper.

References

39. Tabebordbar, M. et al. Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. *Cell* **184**, 4919–4938 (2021).
40. Perez-Riverol, Y. et al. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* **53**, D543–D553 (2025).
41. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

Acknowledgements

This work was supported by the following grants: the National Natural Science Foundation of China (82071409, U20A20356, 82171846, 82422025, 82430059 and 82401635), Beijing Nova Program (20220484017, 20230484403), Beijing Natural Science Foundation (7244421), the National High-Level Hospital Clinical Research Funding (2023HQ03), Association Française contre les Myopathies AFM-Téléthon (28811 to M.B.), Fondation pour la Recherche sur le Cerveau FRC 2023 and Fondation pour la Recherche Médicale (FRM PMT202306017578 and FRM EQU202103012936 to N.C.-B.).

M.B. was supported by a 2-year postdoctoral ITI IMCBio funding. The GenomEast Sequencing platform at IGBMC is a member of the national France Génomique consortium supported by the French National Research Agency (ANR-10-INBS-0009). The Light Microscopy Facility at the IGBMC imaging center is a member of the National Infrastructure France-BioImaging and is supported by the French National Research Agency (ANR-10-INBS-04). Lastly, this work of the Interdisciplinary Thematic Institute IMCBio+, as part of the ITI 2021-2028 program of the University of Strasbourg, National Centre for Scientific Research (CNRS) and Inserm, was supported by IdEx Unistra (ANR-10-IDEX-0002), and by SFRI-STRAT'US project (ANR-20-SFRI-0012) and EUR IMCBio (ANR-17-EURE-0023) under the framework of the France 2030 Program (IGBMC). The authors extend their gratitude to the patients and their families for their invaluable participation in this study. Special thanks to C. Bam'Hamed, H. Ennah and M. Richert (IGBMC) for mouse care; H. Jacob, O. Wendling, B. Weber and L.E. Fertak (IGBMC) for their expertise in histology and mouse tissue analysis; J. Xu (Peking University First Hospital) for expert acquisition of EM images; and J. Liu and Q. Wang (Peking University First Hospital) for their preparation of histopathological sections.

Author contributions

M.B., J.Y., N.E., L.S., D. Pietri, P.G.-R., C.N., A.M., B.M., C.G., E.L., M.P., P.E., A.P., P.R. and M.O.-A. performed the experiments. I.N., K.Y., N.W., Z.W. and J.D. originated the control and OPDM cases. M.B., J.Y., N.E., E.G., D. Plassard, A.M., B.M., C.T., J.D. and N.C.-B. collected and analyzed the data. J.D., M.B. and N.C.-B. designed the study and contributed to coordinating the work and writing the original draft with input from all authors.

Competing interests

All authors declare no competing interests.

Additional information

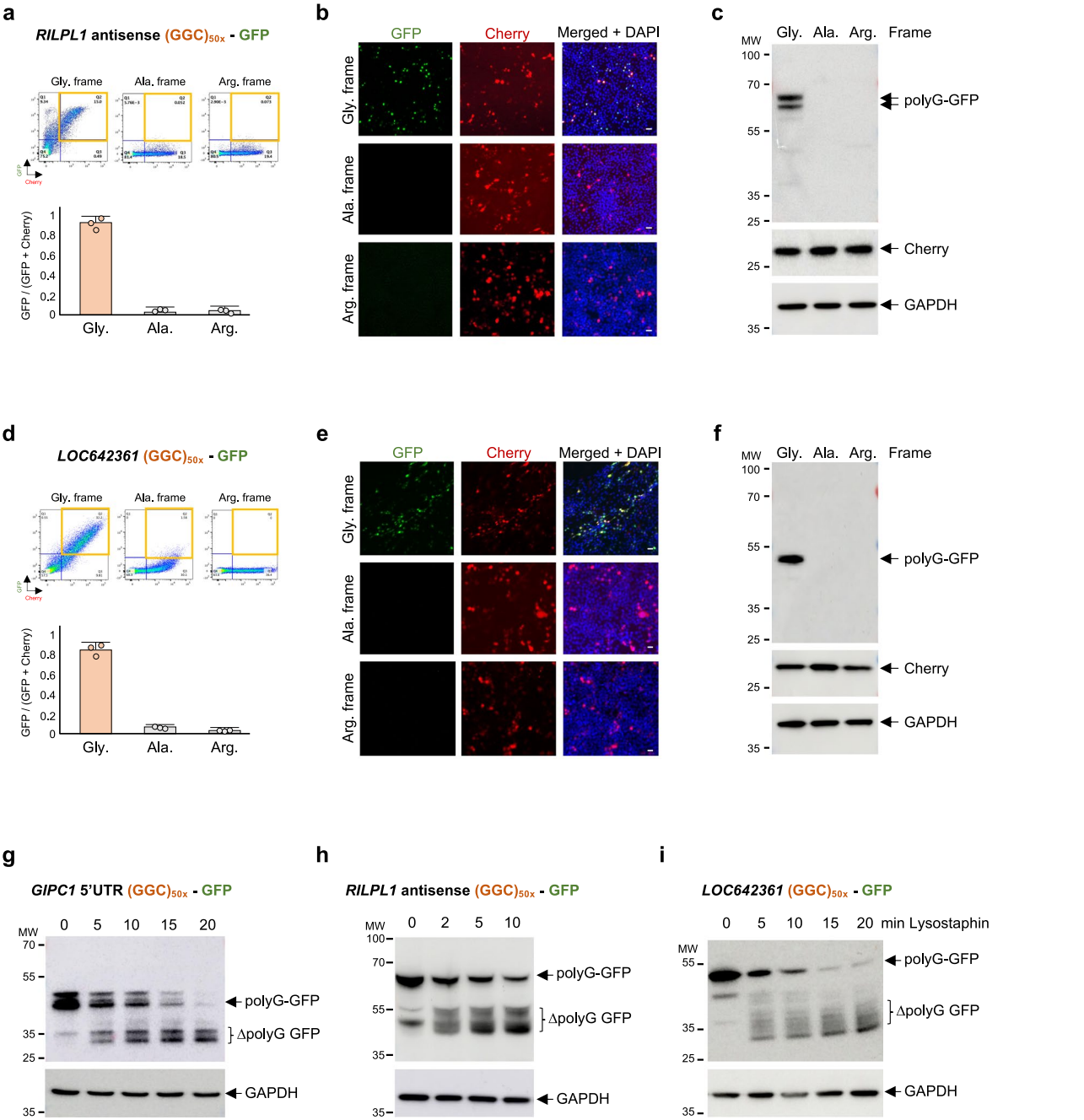
Extended data is available for this paper at https://doi.org/10.1038/s41588-026-02507-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-026-02507-z.

Correspondence and requests for materials should be addressed to Manon Boivin, Jianwen Deng or Nicolas Charlet-Berguerand.

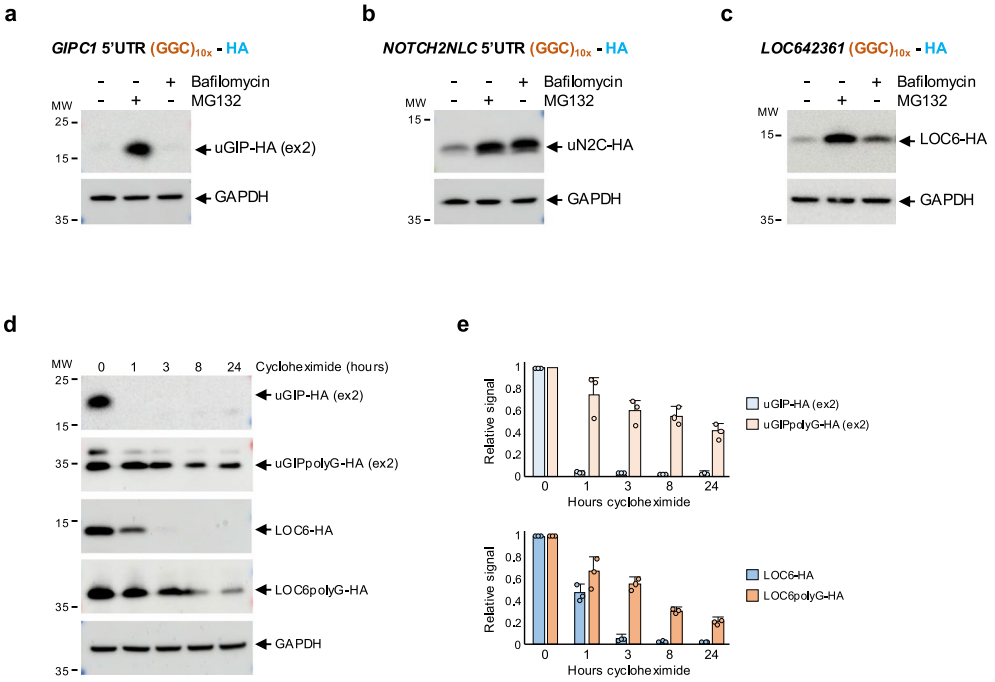
Peer review information *Nature Genetics* thanks Christel Depienne and the other, anonymous, reviewr(s) their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-026-02507-z



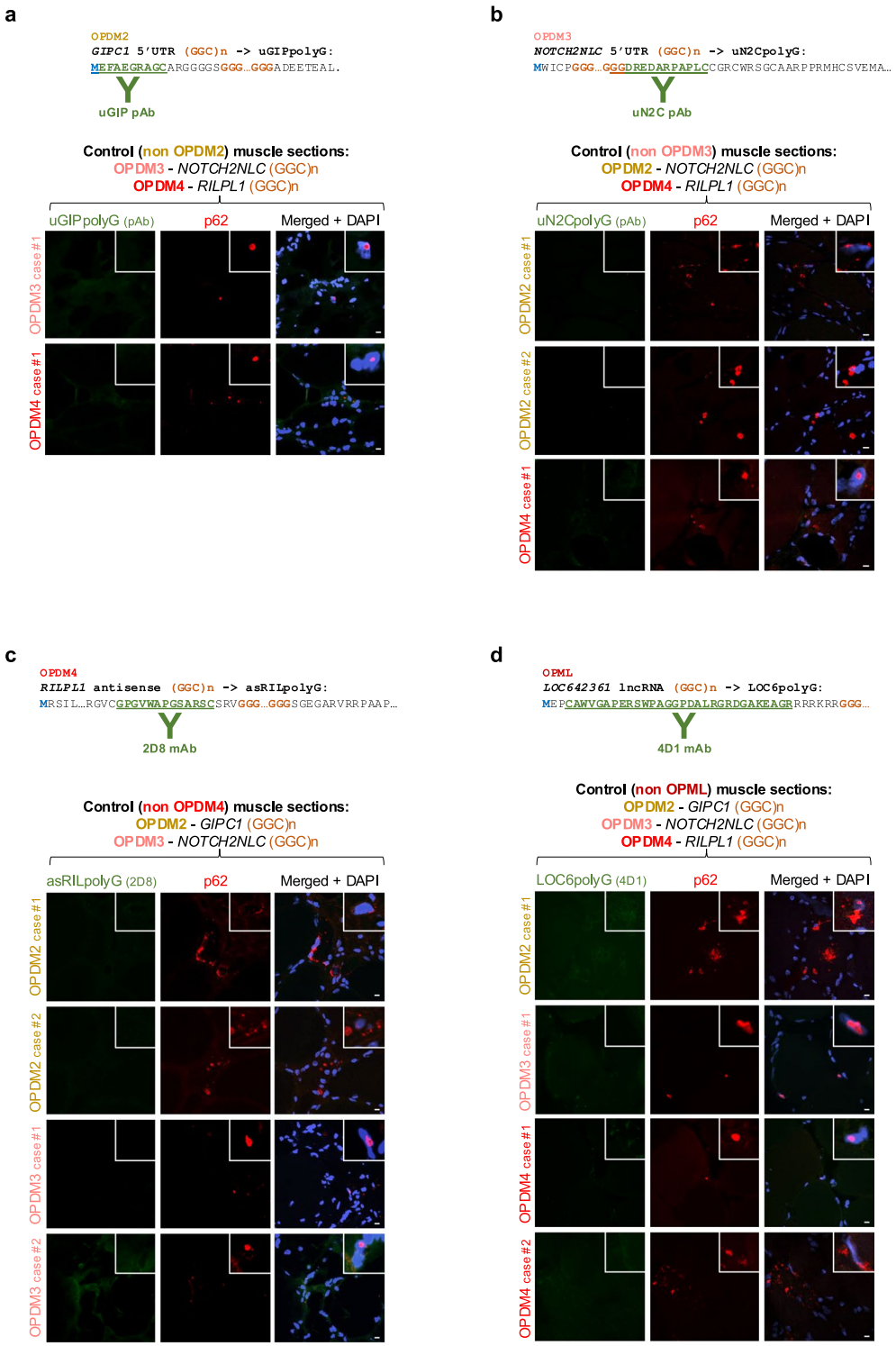
Extended Data Fig. 1 | The *GIPC1*, *RILPL1* and *LOC642361* GGC repeats are translated into polyglycine. a–**c**, GFP and Cherry FACS analysis (higher panel) with its quantification (lower panel), microscopy fluorescence and immunoblotting of HEK293 cells transfected for 24 hours with a plasmid expressing 50 GGC repeats embedded within the *RILPL1* antisense sequence fused to GFP in the glycine, alanine or arginine frames, while Cherry is expressed independently. Scale bars = 10 µm. **d**–**f**, As in **a**–**c** but with 50 GGC repeats embedded in the *LOC642361* sequence. Scale bars = 10 µm. Bar heights represent the mean. Error bars represent the mean values ± s.e.m. Sample size: $n = 3$ independent biological replicates. **g**–**i**, Immunoblots against GFP or GAPDH of lysostaphin-digested proteins extracted from HEK293 cells transfected for 24 hours with GGC repeats embedded within *GIPC1* 5′UTR, *RILPL1* antisense transcript or *LOC642361* fused to GFP in the glycine frame.

https://doi.org/10.1038/s41588-026-02507-z



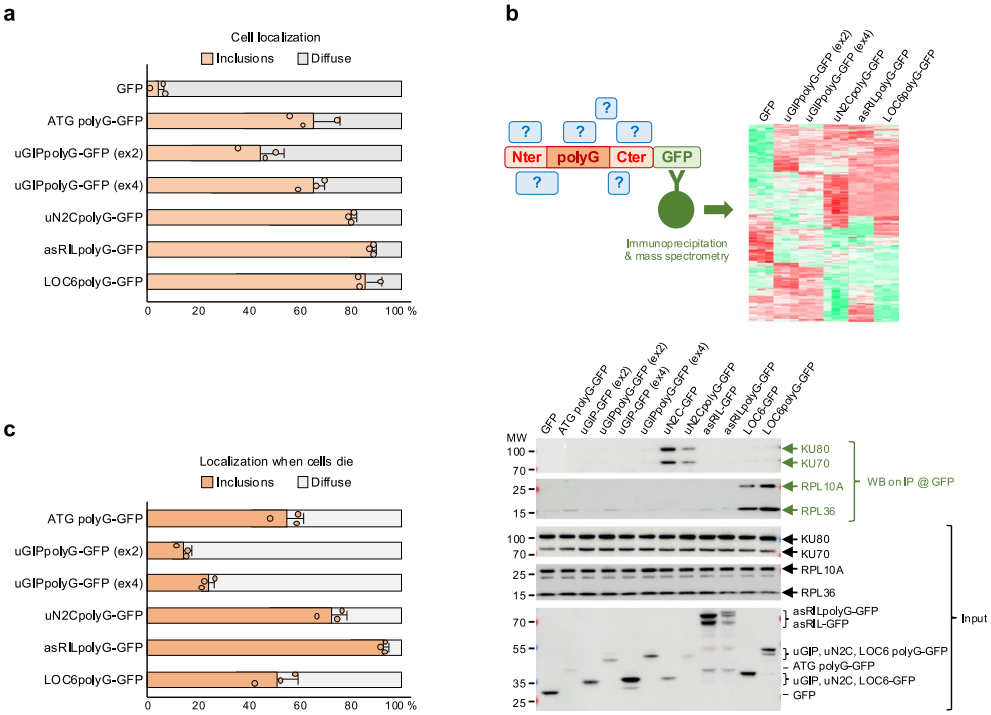
Extended Data Fig. 2 | The *GIPC1*, *NOTCH2NLC* and *LOC642361* GGC repeats are embedded in small ORFs. a–**c**, Immunoblots against HA tag or GAPDH of proteins extracted from HEK293 cells transfected for 24 hours with the *GIPC1*, *NOTCH2NLC* or LOC642361 small ORFs with a control size (10×) of GGC repeats fused to an HA tag and treated with or without MG132 and/or Bafilomycin A1 for 15 hours. **d**,**e**, Half-life of HA-tagged uGIP and LOC6 proteins with 10 or 100 glycine, expressed in HEK293 cells, and determined by cycloheximide chase experiment. Bar heights represent the mean. Error bars represent the mean values ± s.e.m. Sample size: $n = 3$ independent biological replicates.

https://doi.org/10.1038/s41588-026-02507-z



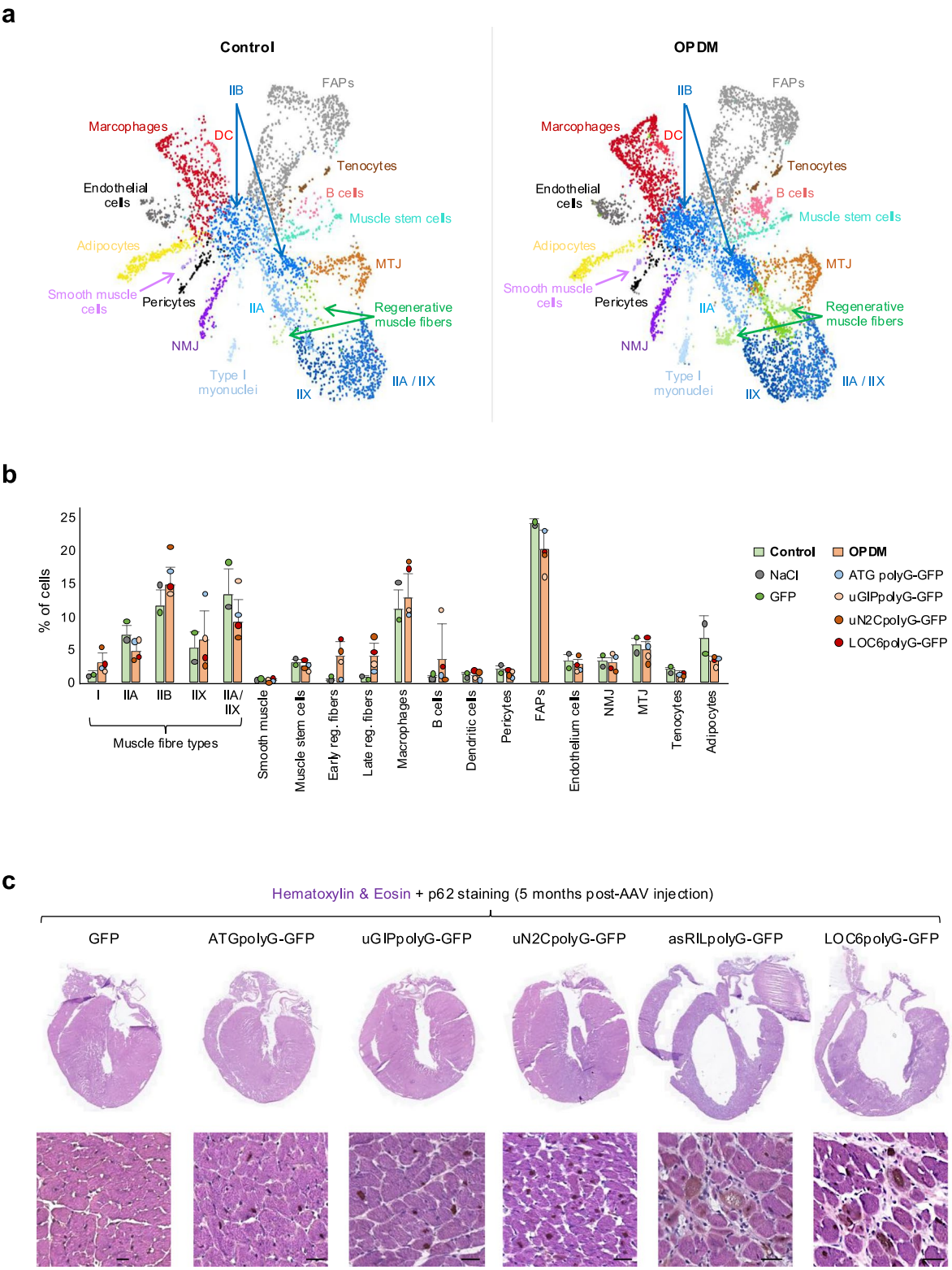
Extended Data Fig. 3 | Polyglycine proteins are present in typical OPDM/OPML p62-positive inclusions. a, Upper panel, partial amino acid sequence of the uGIPpolyG protein encoded by the expanded GGC repeats embedded in the *GIPC1* 5′UTR sequence causing OPDM2. The peptide sequences against which the uGIP antibody is directed are indicated in bold and underlined. Lower panel, immunofluorescence staining against p62 and the uGIPpolyG protein on skeletal muscle sections of individuals with OPDM3 or OPDM4. **b**, As in **a** but with an antibody against the uN2CpolyG protein expressed in NIID/OPDM3 and with staining of skeletal muscle sections of individuals with OPDM2 or OPDM4. **c**, As in **a** but with an antibody against the asRILpolyG protein expressed in OPDM4 and with staining of skeletal muscle sections of individuals with OPDM2 or OPDM3. **d**, As in **a** but with an antibody against the LOC6polyG protein expressed in OPML and with staining of skeletal muscle sections of individuals with OPDM2, OPDM3 or OPDM4. Scale bars, 10 µm.

https://doi.org/10.1038/s41588-026-02507-z



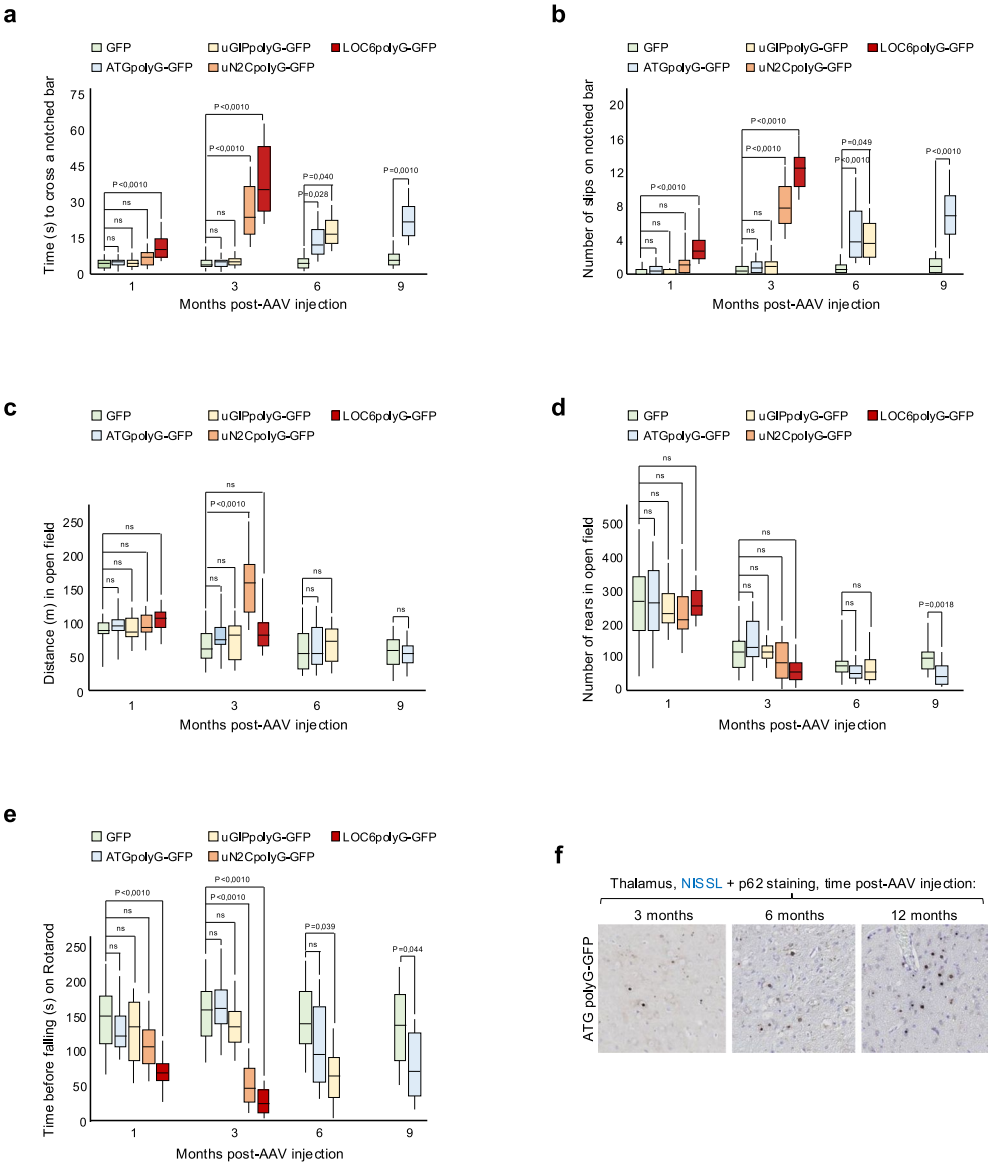
Extended Data Fig. 4 | Expression of polyglycine proteins forms inclusions and is pathogenic in muscle cells. a, Quantification of GFP localization diffuse versus in inclusions of LHCN-M2 muscle cells differentiated for 2 days and expressing either GFP or GFP-tagged ATG polyG, OPDM2 uGIPpolyG, OPDM3 uN2CpolyG, OPDM4 asRILpolyG or OPML LOC6polyG. Bar heights represent the mean. Error bars represent the mean values ± s.e.m. Sample size: $n = 3$ independent biological replicates. **b**, Upper panel, scheme and mass spectrometry heat map of the interactants of GFP immunoprecipitated ATG polyG, uGIPpolyG, uN2CpolyG, asRILpolyG or LOC6polyG-GFP-tagged proteins expressed in 48-hour-differentiated LHCN-M2 muscle cells. Lower panel, immunoblotting of the KU70/80 and RPL10/36 proteins from protein lysates (input) or GFP-immunoprecipitated proteins from 24-hour-transfected HEK293 cells expressing either GFP, GFP-tagged ATG polyG, *GIPC1* 5′UTR, *NOTCH2NLC* 5′UTR, *RILPL1* antisense transcript or *LOC642361* lncRNA ORFs with or without a polyglycine stretch. **c**, Single cell tracking quantification of whether the GFP signal was localized diffusely or in inclusions in Nup50-Cherry U2OS cells that died upon expression of GFP-tagged ATG polyG, OPDM2 uGIPpolyG, OPDM3 uN2CpolyG, OPDM4 asRILpolyG or OPML LOC6polyG. Bar heights represent the mean. Error bars represent the mean values ± s.e.m. Sample size: $n = 3$ independent biological replicates.

https://doi.org/10.1038/s41588-026-02507-z



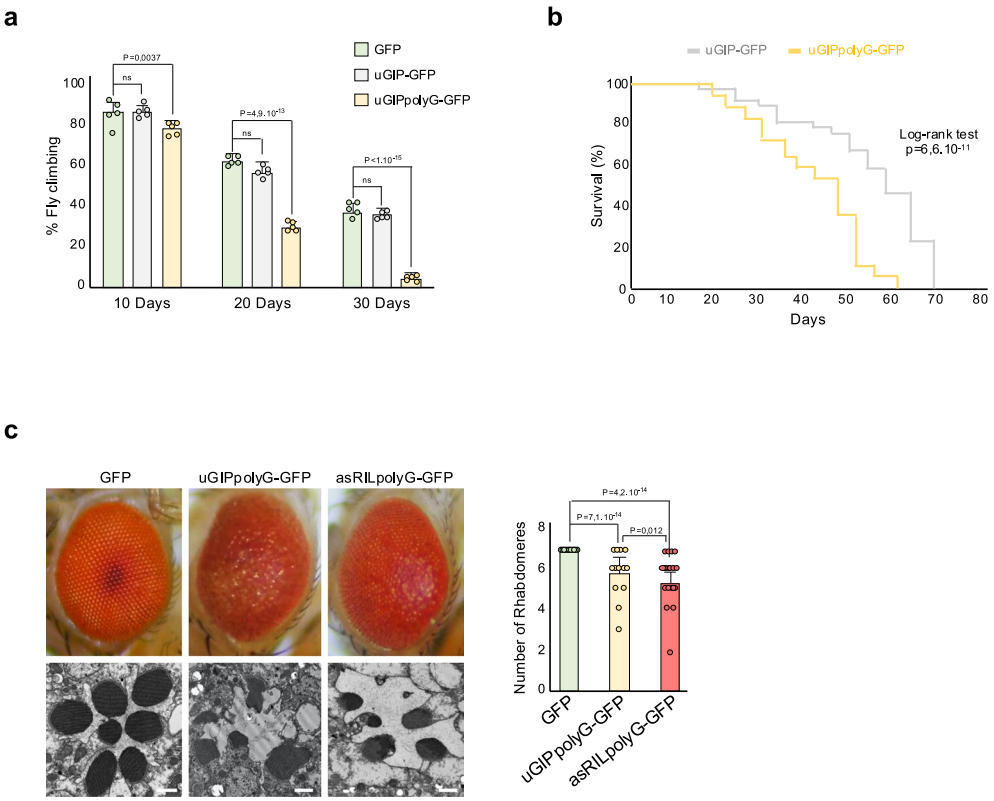
Extended Data Fig. 5 | Expression of polyG proteins forms inclusions and is pathogenic in animal muscles. a, UMAP of single nuclei RNA sequencing of pooled leg muscles (TA, gastrocnemius and quadriceps) from 7-months injected control NaCl or AAV expressing GFP, ATG polyG-GFP, OPDM2 uGIPpolyG-GFP, OPDM3 uN2CpolyG-GFP or OPML LOC6polyG-GFP. **b**, Cell type quantification of **a**. Bar heights represent the mean. Error bars represent the mean values ± s.e.m. Sample size: $n = 2$ pooled mice per condition with controls encompassing NaCl- and GFP-injected mice and OPDM comprising polyglycine-expressing mice. **c**, Upper panel, hematoxylin and eosin (H&E) staining of frozen heart sections of 5-months AAV-injected male mice expressing GFP, ATG polyG-GFP, OPDM2 uGIPpolyG-GFP, OPDM3 uN2CpolyG-GFP, OPML LOC6polyG-GFP or OPDM4 asRILpolyG-GFP. Lower panel, corresponding immunohistochemistry against p62 with H&E counterstaining. Scale bars, 40 µm.

https://doi.org/10.1038/s41588-026-02507-z



Extended Data Fig. 6 | Expression of polyglycine proteins forms inclusions and is pathogenic in the CNS of animals. a,**b**, Time to cross a notched bar and numbers of paw slips of mice 1-, 3-, 6- and 9-months postinjection of AAV expressing GFP, ATG polyG-GFP or GFP-tagged OPDM2 uGIPpolyG, OPDM3 uN2CpolyG or OPML LOC6polyG. **c**,**d**, Maximal distance traveled and number of rears during 30 minutes in an open field of mice 1-, 3-, 6- and 9-months postinjection of AAV expressing GFP, ATG polyG-GFP or GFP-tagged OPDM2 uGIPpolyG, OPDM3 uN2CpolyG or OPML LOC6polyG. **e**, Time before falling from a rotating rod of mice 1-, 3-, 6- and 9-months postinjection of AAV expressing GFP, ATG polyG-GFP or GFP-tagged OPDM2 uGIPpolyG, OPDM3 uN2CpolyG or OPML LOC6polyG. **a**–**e**, Sample size: $n = 10$ mice per condition. Box-and-whisker plot, box upper and lower limits represent 25th and 75th percentiles, whiskers represent minimum and maximum values and the horizontal line across the box represents the median. One-way ANOVA with Tukey post hoc test. **f**, Immunohistochemistry against p62 with cresyl violet (Nissl) counterstaining of mouse brain thalamus area 3-, 6- or 12-months postinjection of AAV expressing the ATG polyG-GFP protein.

https://doi.org/10.1038/s41588-026-02507-z



Extended Data Fig. 7 | The porphyrin TMPYP4 alleviates aggregation and toxicity of polyglycine proteins. a, Adult male flies expressing uGIPpolyG-GFP exhibited progressive locomotor deficits compared to GFP- or uGIP-GFP-expressing controls. Bar heights represent the mean. Error bars represent the mean values ± s.e.m. Sample size: $n = 5$ independent experiments with >100 flies per condition. Unpaired two-tailed t-test compared to the GFP control condition. **b**, Survival analysis showed a decreased median lifespan in uGIPpolyG-GFP-expressing male flies compared to uGIP-GFP-expressing controls. Genotypes: Actin5C-Gal4>UAS-GFP (control), Actin5C-Gal4>UAS-uGIP-GFP, Actin5C-Gal4>UAS-uGIPpolyG-GFP. Log-rank Mantel–Cox test; sample size: $n = 5$ independent experiments with >100 flies per genotype. **c**, Left panel, representative light microscopy (upper panel) and electron microscopy (lower panel) images of fly eyes expressing GFP (control), uGIPpolyG-GFP or asRILpolyG-GFP. Scale bars, 1 µm. Right panel, quantification of intact rhabdomeres per ommatidium. Bar heights represent the mean. Error bars represent the mean values ± s.e.m. All quantitative data are given as the number of rhabdomeres per ommatidium. Sample size: $n = 42$ ommatidia from 3 flies per condition. One-way ANOVA with Bonferroni post hoc test.

nature portfolio

Corresponding author(s): Nicolas CHARLET-BERGUERAND

Last updated by author(s): 10 November 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No custom code used, standard analysis software comprise : FloJo, Leica Application Suite, ImageJ, QuPath.
Data analysis	ImageJ, QuPath, GraphPad prism.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Source data are provided in the SourceData file, in the Supplementary Information file and in the Supplementary Tables 1 to 5. Additional data are available from the corresponding authors upon request.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were either determined by G*Power using an ANOVA test with a power (1-ß) of 0.8 and a significance level (alpha) of 0.05, or based on previous published studies and our own experimental experience (Boivin et al., 2021).
Data exclusions	No data were excluded.
Replication	At least 3 independent biologically samples were analyzed in all experiments. Exact number of replicates are indicated in the figure legends and Methods sections.
Randomization	Animal were randomly assigned into their experimental groups. Randomization is not relevant for other experiments
Blinding	Image analysis and data collection were blinded to group assignment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Western blotting: anti-GFP (Abcam, ab290, 1/10000), GFP (Abcam, ab1218, 1/10000), mCherry (Abcam, ab167453, 1/5000), GAPDH (Abcam, ab8245,

1/10000), Ku70 (SantaCruz, sc-56129, 1/5000), Ku80 (Abcam, ab119935, 1/10000), RPL10A (Thermofisher, MA5-27171, 1/3000), RPL36 (Abcam, ab241584, 1/10000), HA (Abcam, ab130275, 1/5000), uGIPpolyG pAb or uN2CpolyG pAb (rabbit polyclonal homemade, 1/1000), uN2CpolyG (4D12, mouse monoclonal homemade, 1/1000), asRILpolyG (2D8 or 4B9 mouse monoclonal homemade, 1/1000) and LOC6polyG (2E8, mouse monoclonal homemade, 1/1000).

Immunofluoresence on PFA-fixed cells
p62 (Abcam, ab56416, 1/1000), p62 (Abcam, ab109012, 1/1000), Desmin (Abcam, ab32362, 1/500), Lamin A/C (Abcam, ab238303, 1/1000), uGIPpolyG pAb or uN2CpolyG pAb (rabbit polyclonal homemade, 1/100), uN2CpolyG (4D12, mouse monoclonal homemade, 1/100), asRILpolyG (2D8 or 4B9 mouse monoclonal homemade, 1/100) and LOC6polyG (2E8, mouse monoclonal homemade, 1/100).

Immunofluoresence or immunochemistry on PFA-fixed tissue sections
Calbindin (CST, 13176S, 1/800), GFAP (Abcam, ab68428, 1/10000), p62 (Abcam, ab56416, 1/1000), p62 (CST, 23214S, 1/500), Tyrosine Hydroxylase (Abcam, ab112, 1/2000).

Immunofluoresence or immunochemistry on isopentane-frozen sections
p62 (Abcam, ab109012, 1/1000), p62 (CST, 23214S, 1/500), Lamin A/C (Abcam, ab238303, 1/1000), Lamin B1 (Proteintech, 12987-1-AP, 1/500), Muscle type I fibers (DSHB, BA-D5, 1/50), Muscle type IIa fibers (DSHB, SC-71, 1/50), Muscle type IIb fibers (DSHB, BF-F3, 1/50), uGIPpolyGly pAb or uN2CpolyG pAb (rabbit polyclonal homemade, 1/100), uGIPpolyAla (1A7 or 3G4, mouse monoclonal homemade, 1/100) uN2CpolyG (4D12, mouse monoclonal homemade, 1/100), asRILpolyG (2D8 or 4B9, mouse monoclonal homemade, 1/100) or LOC6polyG (2E8, mouse monoclonal homemade, 1/100).

Validation

Validation of custom-made antibodies are provided in the Supplementary Information. Commercial antibodies validations are provided on their corresponding manufacturer's website and in various published studies.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK-293 (CRL-1573), U-2 OS (HTB-96) cell lines originate from ATCC, LHCN-M2 originates from the University of Texas Southwestern Medical Center (PMID: 17559502),
Authentication	Cell morphology was confirmed by microscopy, no further authentication was performed.
Mycoplasma contamination	Cells were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	n/a. No commonly misidentified cell lines were used.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	C57BL/6, 2 months old, male, ordered at Charles River Laboratories
Wild animals	n/a. No wild animals were used in this study.
Reporting on sex	Only male mice were used.
Field-collected samples	n/a. No field collected samples were used in this study.
Ethics oversight	All animal work was performed with approval from the IGBMC/ICS Animal Care Committee and of the French agency for research on animal (DGRI) authorization number APAFIS#33864-2021111217327782.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a