Article

m^{1}A in CAG repeat RNA binds to TDP-43 and induces neurodegeneration

https://doi.org/10.1038/s41586-023-06701-5

Received: 18 April 2022

Accepted: 2 October 2023

Published online: 8 November 2023

Open access

Check for updates

Yuxiang Sun[1,8], Hui Dai[1,2,8], Xiaoxia Dai[1], Jiekai Yin[3], Yuxiang Cui[3], Xiaochuan Liu[1], Gwendolyn Gonzalez[3], Jun Yuan[3], Feng Tang[1], Nan Wang[4], Alexandra E. Perlegos[5], Nancy M. Bonini[5,6], X. William Yang[4,7], Weifeng Gu[2] & Yinsheng Wang[1,3]✉

Microsatellite repeat expansions within genes contribute to a number of neurological diseases[1,2]. The accumulation of toxic proteins and RNA molecules with repetitive sequences, and/or sequestration of RNA-binding proteins by RNA molecules containing expanded repeats are thought to be important contributors to disease aetiology[3–9]. Here we reveal that the adenosine in CAG repeat RNA can be methylated to N^1-methyladenosine (m^{1}A) by TRMT61A, and that m^{1}A can be demethylated by ALKBH3. We also observed that the m^{1}A/adenosine ratio in CAG repeat RNA increases with repeat length, which is attributed to diminished expression of ALKBH3 elicited by the repeat RNA. Additionally, TDP-43 binds directly and strongly with m^{1}A in RNA, which stimulates the cytoplasmic mis-localization and formation of gel-like aggregates of TDP-43, resembling the observations made for the protein in neurological diseases. Moreover, m^{1}A in CAG repeat RNA contributes to CAG repeat expansion-induced neurodegeneration in *Caenorhabditis elegans* and *Drosophila*. In sum, our study offers a new paradigm of the mechanism through which nucleotide repeat expansion contributes to neurological diseases and reveals a novel pathological function of m^{1}A in RNA. These findings may provide an important mechanistic basis for therapeutic intervention in neurodegenerative diseases emanating from CAG repeat expansion.

Nucleotide repeat expansions contribute to a number of neurological diseases[1,2]. For instance, expansion of a GGGGCC hexanucleotide repeat in the *C9ORF72* gene is the major genetic cause of amyotrophic lateral sclerosis[10] (ALS), and CAG repeat expansions contribute to the development of multiple neurodegenerative disorders including Huntington's disease and various forms of spinocerebellar ataxia[2] (SCA). Several mechanisms have been proposed for neurological diseases elicited by repeat expansions; these include neurotoxicity arising from formation of neuronal intranuclear inclusions and amyloid-like aggregates of proteins translated from repeat RNAs[3,4], and the ensuing impairment of the ubiquitin–proteasome system[11]. Additionally, accumulation and aberrant phase separation of toxic RNAs, and their sequestration of RNA-binding proteins are also important contributors to disease aetiology[5–9].

Cytoplasmic mis-localization and aggregation of TDP-43 in the degenerated regions of the brain are clinical hallmarks of many neurological diseases, including ALS, frontotemporal lobar degeneration[12,13] (FTLD), and CAG repeat expansion disorders, such as Huntington's disease[14] and SCA[15]. Additionally, intermediate-length CAG repeat expansions in the *ATXN2* gene are significantly associated with ALS[16], in which genetic depletion of *ATXN2* extends lifespan and mitigates pathological aggregates in TDP-43 transgenic mice[15]. TDP-43 contains two RNA-recognition motifs (RRMs) and a C-terminal low-complexity domain[13] (LCD), through which the protein can undergo intermolecular interactions and assemble into phase-separated liquid condensates[17]. Furthermore, phase-separated droplets of RNA–protein complexes can become less dynamic and aggregate over time, suggesting that aberrant phase transition contributes to the aggregation of RNA-binding proteins[18].

Dynamic post-transcriptional modifications in RNA, especially methylation at the N^6 position of adenosine, constitute an important mechanism regulating the stability and translation efficiency of mRNAs[19,20], and dysregulation of this methylation in mRNAs are linked with human diseases[21]. Here we show that CAG repeat expansions lead to increased levels of m^{1}A, which binds to TDP-43, alters its subcellular distribution and phase separation behaviour, and contributes to neurodegeneration.

m^{1}A increases with CAG repeat length

Previous studies documented that the age of onset of neurological disorders arising from CAG repeat expansion is negatively correlated with repeat length, whereas disease severity is positively associated with repeat length[2]. We set out to test whether methylated adenosines—especially N^6-methyladenosine (m^{6}A) and m^{1}A—are present in CAG repeat RNA and if so, how their modification frequencies may be influenced by repeat length and how they modulate RNA–protein interactions.

[1]Department of Chemistry, University of California Riverside, Riverside, CA, USA. [2]Department of Molecular, Cell and Systems Biology, University of California Riverside, Riverside, CA, USA. [3]Environmental Toxicology Graduate Program, University of California Riverside, Riverside, CA, USA. [4]Center for Neurobehavioral Genetics, The Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, USA. [5]Neurosciences Graduate Group, University of Pennsylvania, Philadelphia, PA, USA. [6]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA. [7]Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [8]These authors contributed equally: Yuxiang Sun, Hui Dai. ✉e-mail: Yinsheng.Wang@ucr.edu

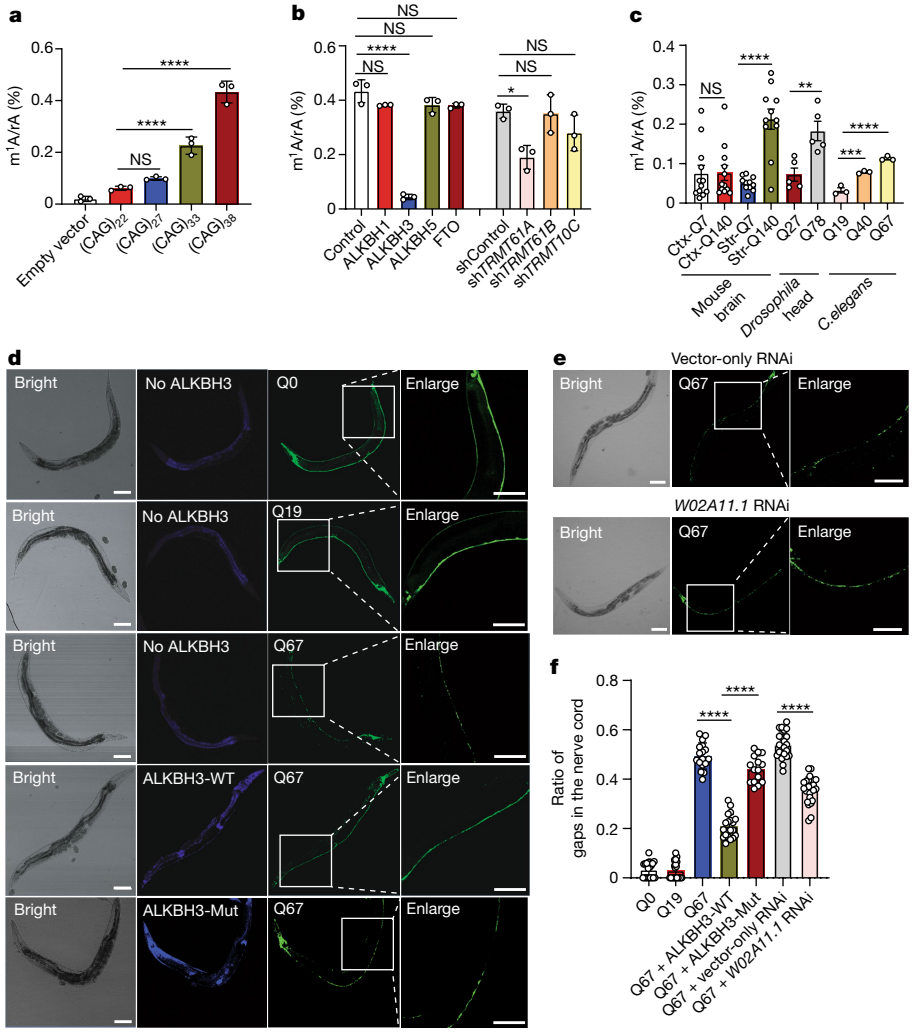


Fig. 1 | m¹A in CAG repeat RNA increases with repeat length and contributes to neurodegeneration in *C. elegans*. **a**–**c**, The m¹A/rA ratios in ectopically expressed CAG repeat RNAs isolated from HEK293T cells (**a**; $n = 3$ biologically independent experiments), ectopically expressed $(CAG)_{38}$ RNA isolated from HEK293T cells with ectopic co-expression of the indicated RNA demethylases or stable knockdown of m¹A methyltransferases (**b**; $n = 3$ biologically independent experiments), and CAG repeat RNA from mouse brain ($n = 11$ and 12 biologically independent samples for striatum (str) and cortex (ctx) tissues, respectively), *Drosophila* head ($n = 5$ biologically independent samples, 10-day old males, 30 heads per sample) and *C. elegans* ($n = 3$ biologically independent experiments) (**c**). **d**,**e**, Fluorescence images of representative Q0, Q19 and Q67 nematodes, in Q67 nematodes with ectopic expression of wild-type human ALKBH3 (ALKBH3-WT) or its catalytically inactive mutant (ALKBH3-Mut) (**d**), and in Q67 nematodes with or without knockdown of *W02A11.1* gene (**e**). **f**, Ratios of gaps in the nerve cord in Q67 worms with or without expression of human ALKBH3-WT and ALKBH3-Mut (Q0: $n = 20$ worms; Q19: $n = 20$ worms; Q67: $n = 17$ worms; Q67 + ALKBH3-WT: $n = 20$ worms; Q67 + ALKBH3-Mut: $n = 15$ worms), and in Q67 nematodes with or without knockdown of *W02A11.1* gene (Q67 + vector RNAi: $n = 22$ worms; Q67 + *W02A11.1* RNAi: $n = 23$ worms). The ratios of gaps were calculated by dividing the total length of gaps over the entire length of the nerve cord. Data are mean ± s.d. *P* values for the data of mouse brain and *Drosophila* head samples in **c**, and between Q67 + vector RNAi and Q67 + *W02A11.1* RNAi in **f** were determined using two-tailed Student's *t*-test; all other *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. NS, not significant ($P > 0.05$). **b**, $^{*}P = 0.021$. **c**, $^{**}P = 0.0058$ and $^{***}P = 0.0007$. $^{****}P < 0.0001$. Scale bars, 100 µm.

To address these questions, we ectopically expressed mRNAs with different CAG repeat lengths in HEK293T cells, isolated the CAG repeat RNA and quantified the levels of m¹A and m⁶A in the resulting RNA samples using liquid chromatography–tandem mass spectrometry[22] (LC–MS/MS). The results revealed a progressive increase in the frequency of m¹A (relative to adenosine (rA)) with repeat length, which is accompanied with a gradual diminution in m⁶A level (Fig. 1a and Supplementary Figs. 2, 3 and 4a). Of note, the level of m¹A in $(CAG)_{38}$ RNA was even higher than that of m⁶A (Fig. 1a and Supplementary Fig. 4a). We also annealed the purified CAG repeat RNA with two oligodeoxynucleotides whose sequences are complementary to the 5′ and 3′ flanking sequences of the $(CAG)_{38}$ RNA, and removed the RNA in the ensuing RNA/DNA hybrid with RNase H. The results showed that such removal led to an increased level of m¹A, but not m⁶A, underscoring a higher frequency of m¹A in $(CAG)_{38}$ repeat than in flanking sequences (Supplementary Fig. 5a,b). Notably, $(CAG)_{22}$ and $(CAG)_{38}$ RNAs were expressed at similar levels (Supplementary Fig. 5c).

m¹A can be demethylated by ALKBH1 and ALKBH3[23,24], and m⁶A can be demethylated by FTO and ALKBH5[25,26]. We tested whether these demethylases could also act on m⁶A and m¹A in CAG repeat RNAs. We observed a marked attenuation (by approximately tenfold) in the levels of m¹A in $(CAG)_{38}$ after ectopic expression of ALKBH3, but not ALKBH1, FTO or ALKBH5 (Fig. 1b). In addition, we found that m⁶A in $(CAG)_{38}$ could be demethylated by FTO (Supplementary Fig. 4b), but not by ALKBH1 or ALKBH5; ectopic expression of ALKBH3 also led to a diminished level of m⁶A in $(CAG)_{38}$ RNA (Supplementary Fig. 4b).

Several m¹A methyltransferases have been identified, including TRMT6–TRMT61A[27], TRMT61B[28] and TRMT10C[29]. Although TRMT61B

and TRMT10C are involved in m^1A formation in mitochondrial tRNAs[28,29], the TRMT6–TRMT61A heterodimer can install m^1A in both tRNAs and mRNAs[27,30]. TRMT61A is the catalytic subunit of the methyltransferase complex and it is highly conserved from invertebrates to humans[27]. We therefore used short hairpin RNA (shRNA) to knock down TRMT61A expression in human cells (Supplementary Fig. 6a,d) and measured the levels of m^1A in $(CAG)_{38}$ RNA. TRMT61A depletion led to a significant decrease in m^1A level in $(CAG)_{38}$ RNA, whereas we observed no change in m^6A level (Fig. 1b and Supplementary Figs. 4b and 6e,f). We also observed that recombinant TRMT6–TRMT61A proteins could catalyse the formation of m^1A in $(CAG)_{16}$ RNA in vitro, whereas the heat-inactivated enzyme did not (Supplementary Fig. 7). Knockdown of TRMT61B or TRMT10C in HEK293T cells, however, did not alter the level of m^1A in $(CAG)_{38}$ RNA (Fig. 1b).

We next examined whether the frequency of m^1A also increases with CAG repeat length in vivo. We used affinity purification to isolate CAG repeat RNA from the striatum and cortex tissues of wild-type mice that carry 7 CAG repeats (Q7) or transgenic mice expressing one wild-type endogenous *Htt* allele and a second *Htt* allele with knock-in of human mutant *HTT* (*mHTT*) exon 1, which contains 140 CAG repeats[31] (Q140). LC–MS/MS analysis showed that the m^1A level was significantly higher in Q140 mRNA than in Q7 mRNA isolated from striatal tissues (Fig. 1c), but there was no significant difference in m^6A level in Q140 and Q7 mRNA (Supplementary Fig. 4c). In addition, no apparent difference was detected for the level of m^6A or m^1A in the cortical tissues of mice expressing the two different lengths of CAG repeat. These observations parallel the previous finding that repeat length-dependent transcriptional signatures are more prominent in the striatum than in the cortex tissues of these mice[31]. Similarly, m^1A, but not m^6A, was present at a higher level in CAG repeat mRNA isolated from *Drosophila* heads with neuronal-specific expression of an *SCA3* transgene with 78 repeats of CAG (Q78), than those with a short Q27 CAG repeat[6] (Fig. 1c and Supplementary Fig. 4c). We also detected markedly increased levels of m^1A in CAG repeat RNA isolated from *C. elegans* with pan-neuronal expression of $(CAG)_{40}$ and $(CAG)_{67}$ mRNA (Q40 and Q67) than in worms expressing $(CAG)_{19}$ mRNA[32] (Q19), although no difference was detected in the levels of m^6A in these mRNA samples (Fig. 1c and Supplementary Fig. 4c).

We next examined the role of TRMT61A in modulating m^1A level in CAG repeat RNA in *C. elegans*. We performed RNA-mediated interference (RNAi) targeting the *W02A11.1* gene, which is the nematode orthologue of the human *TRMT61A* gene. Knockdown of *W02A11.1* led to a significantly diminished level of m^1A in CAG repeat RNA extracted from the nematodes, whereas there was no significant difference in m^6A level (Supplementary Fig. 8a,b). Together, these results indicate that the levels of m^1A in CAG repeat RNA increase with repeat length, with the m^1A in CAG repeat RNA being installed and removed by TRMT61A and ALKBH3, respectively.

Diminished ALKBH3 leads to increased m^1A

We next examined the origin of the repeat length-dependent increase in m^1A level in CAG repeat RNA. We tested whether expression of CAG repeat RNA modulates the expression levels of the aforementioned m^1A writer and eraser—TRMT61A and ALKBH3. respectively. We found no apparent change in expression level of TRMT61A protein in HEK293T cells or mouse tissues expressing different lengths of CAG repeat RNA (Extended Data Fig. 1a–c). By contrast, the levels of ALKBH3 mRNA and protein were diminished in HEK293T cells expressing $(CAG)_{22}$ and $(CAG)_{38}$, with a more pronounced diminution being observed for the latter (Extended Data Fig. 1d–f). In addition, we found that ALKBH3 protein was expressed at a lower level in striatal tissues of Q140 mice than Q7 mice; the expression level of this protein, however, was similar in the cortex tissues of Q140 and Q7 mice (Extended Data Fig. 1g–j). This result is in agreement with the levels of m^1A detected in the corresponding cells and tissues. Together, these results support that diminished expression of ALKBH3 contributes to elevated frequencies of m^1A in CAG repeat RNA.

m^1A results in neurodegeneration in vivo

We next investigated whether m^1A in CAG repeat expansion mRNA contributes to neurodegeneration in vivo. We first examined the role of m^1A in modulating neurodegeneration in *C. elegans* with pan-neuronal expression of 67 repeats of CAG—that is, $P_{F25B3.3}$::Q67::CFP, where CFP is fused to the C-terminus of Q67[32]. On this genetic background, we generated transgenic nematodes expressing BFP-tagged human ALKBH3. In line with previous observations[32], Q67 worms exhibited pronounced degeneration of the neuron network, as manifested by the loss of neuronal cell bodies and expansion of gaps in neurites of the ventral and dorsal nerve cords (Fig. 1d,f). Expression of wild-type human ALKBH3 (ALKBH3-WT), but not its catalytically inactive mutant (ALKBH3-Mut), diminished the length of the gaps and improved the continuity of dorsal and ventral nerve cords (Fig. 1d,f). Similarly, genetic depletion of *W02A11.1* alleviated neurodegeneration in *C. elegans* (Fig. 1e,f).

We also observed an extension of the lifespan in Q78 *Drosophila* with neuronal expression ALKBH3-WT, but not in those expressing ALKBH3-Mut (Extended Data Fig. 2a,b), indicating mitigated neurotoxicity. Our results showed an attenuated level of m^1A in CAG repeat RNA isolated from Q78 *Drosophila* expressing wild-type human ALKBH3 relative to those expressing its catalytically inactive mutant, with no apparent difference being detected for m^6A level (Extended Data Fig. 2c,d). These data indicate that m^1A contributes to neurotoxicity of CAG repeat RNA in *Drosophila*.

TDP-43 is an m^1A reader protein

Previously, stable isotope labelling by amino acids in cell culture (SILAC)-based affinity screening led to the identification of several candidate m^1A-binding proteins, including TDP-43, YTHDF1-3 and DDX56, with TDP-43 exhibiting the highest SILAC protein ratio for m^1A- over rA-containing probe in proteomic samples prepared from HeLa cells[33]. The mass spectrometry results for QSQDEPLR, a tryptic peptide derived from TDP-43, are shown in Supplementary Fig. 9.

We examined whether TDP-43 can bind directly to m^1A-containing RNA. Electrophoretic mobility shift assay (EMSA) showed that recombinant TDP-43 binds more strongly to the m^1A-carrying RNA substrate used in the quantitative proteomic experiment than its unmethylated or m^6A-containing counterpart, with the dissociation constants (K_d values) of 18, 38 and 46 nM, respectively (Extended Data Fig. 3b,c). Moreover, we found that recombinant TDP-43 exhibited stronger binding to an m^1A-carrying CAG repeat RNA than the corresponding unmodified RNA, with K_d values of 70 and 146 nM, respectively (Extended Data Fig. 3d,e).

TDP-43 contains two RRMs, in which highly conserved aromatic amino acid residues are necessary for the interaction of the protein with RNA, and mutations of 5 Phe residues to Leu in these domains (Extended Data Fig. 3a) abolish the ability of TDP-43 to bind RNA[16]. We observed similar binding selectivity for a truncated version of TDP-43—which contains only the two RRM domains—towards synthetic RNAs with m^1A, rA and m^6A, with K_d values of 31, 73 and 81 nM, respectively (Extended Data Fig. 3f,g). Additionally, we found that the RRM domain exhibits stronger binding to m^1A-containing CAG repeat RNA than the corresponding unmodified CAG repeat RNA, with the K_d values of 0.18 and 0.38 μM, respectively, whereas TDP-43-5FL is incapable of binding m^1A-containing RNA (Extended Data Fig. 3h–k). The purities of recombinant TDP-43 proteins were confirmed by SDS–PAGE analysis (Supplementary Fig. 10).

We next conducted a cross-linking immunoprecipitation–PCR experiment to determine whether TDP-43 binds preferentially to m^1A in CAG

repeat RNA in cells. Our results showed that the ability of TDP-43 to bind with $(CAG)_{38}$ RNA was significantly attenuated in cells with ectopic expression of wild-type ALKBH3, but not with a catalytically inactive mutant (Supplementary Fig. 11a). We also expressed Flag-tagged TDP-43 in HEK293T cells, immunoprecipitated the protein, and quantified the levels of m^1A in the mRNA samples isolated from the immunoprecipitates by LC–MS/MS. Our results showed that the level of m^1A was approximately 2.5-fold higher in the anti-Flag pull-down mRNA samples from cells expressing TDP-43–Flag than those expressing the control empty vector (Supplementary Fig. 11b). In addition, the enrichment of m^1A-bearing mRNA was significantly diminished in the pull-down sample of the corresponding RRM domain mutant TDP-43-5FL–Flag (Supplementary Fig. 11b). We also found that the m^1A level in $(CAG)_{38}$ RNA was significantly decreased in HEK293T cells upon ectopic co-expression of wild-type ALKBH3, but not its catalytically inactive mutant H257A[34] (Supplementary Fig. 11c). Immunoblotting experiments results showed similar levels of expression of wild-type and mutant TDP-43, and of ALKBH3 (Supplementary Fig. 11d,e).

We next isolated CAG repeat RNA using affinity pull-down and monitored the level of TDP-43 protein in the pull-down mixture by immunoblot. Our results showed that TDP-43 was present at a higher level in the pull-down mixture of $(CAG)_{38}$ than that of $(CAG)_{22}$, and the amount of TDP-43 in the pull-down mixture was diminished upon ectopic co-expression of ALKBH3 (Supplementary Fig. 11f,g). By contrast, overexpression of ALKBH1 or FTO, or siRNA-mediated knockdown of METTL3—the catalytic subunit of the major m^6A methyltransferase complex[35]—did not appreciably alter the binding of TDP-43 to $(CAG)_{38}$ (Supplementary Fig. 11h,i). Moreover, $(CAG)_{38}$ RNA did not pull down TDP-43-5FL (Supplementary Fig. 11j). Together, these results show that TDP-43 interacts directly with CAG repeat RNA, and that this interaction requires the RRM domains of TDP-43 and is markedly enhanced by m^1A in RNA.

m^1A induces aberrant TDP-43 behaviour

Truncation, cytoplasmic redistribution and aggregation of TDP-43 are histopathological hallmarks of ALS and FTLD[13], which, along with our observations of the interaction between this protein and m^1A-containing RNA, prompted us to examine the effect of CAG repeat RNA on the cleavage and subcellular distribution of TDP-43.

We were able to detect truncated endogenous TDP-43 in a detergent-insoluble fraction isolated from cells with ectopic expression of $(CAG)_{22}$ and $(CAG)_{38}$ RNA, with $(CAG)_{38}$ RNA being more effective in eliciting truncated TDP-43 protein (Extended Data Fig. 4a,b). Furthermore, ectopic expression of ALKBH3, but not its catalytically inactive mutant, led to diminished levels of truncated TDP-43 in detergent-insoluble fraction isolated from cells expressing $(CAG)_{38}$ RNA (Extended Data Fig. 4c,d). To substantiate the role of m^1A in CAG repeat RNA in inducing truncated TDP-43 in cells, we used synthetic $(CAG)_7$ and $(CAG)_{16}$ RNA carrying zero or three m^1A residues. Our results showed more truncated TDP-43 in cells transfected with $(CAG)_{16}$-3m^1A RNA than with the corresponding unmodified RNA, though no apparent difference was detected for transfections with $(CAG)_7$-0m^1A and $(CAG)_7$-3m^1A RNAs (Extended Data Fig. 4e,f). Similarly, we detected a significantly higher level of truncated TDP-43 in the striatal tissues of mice expressing Q140 RNA than those expressing Q7 RNA, but found no such difference in the cortical tissues (Extended Data Fig. 4g,h). These observations are in keeping with the m^1A levels detected in CAG repeat RNAs isolated from the two types of mouse tissues (Fig. 1c), suggesting that m^1A in CAG repeat RNA stimulates the generation of truncated TDP-43 protein.

We next examined how expression of CAG repeat RNA modulates the intracellular distribution of TDP-43 protein. Immunostaining for endogenous TDP-43 in U2OS cells showed that whereas TDP-43 is distributed primarily in the nucleus in control cells, ectopic expression of $(CAG)_{22}$ or $(CAG)_{38}$ resulted in cytoplasmic redistribution of TDP-43 (Fig. 2a). In addition, endogenous TDP-43 displayed larger foci in cells expressing $(CAG)_{38}$ than in those expressing $(CAG)_{22}$, where the foci size was decreased upon ectopic expression of ALKBH3 (Fig. 2a,b) or upon genetic depletion of TRMT61A (Fig. 2a,b). Consistent with the results of intracellular distribution of endogenous TDP-43, we observed cytoplasmic mis-localization and aggregation of ectopically expressed TDP-43 in cells expressing $(CAG)_{38}$, where the extent of TDP-43 aggregation was more pronounced than that observed for the endogenous protein (Extended Data Fig. 5a,b).

Next we assessed whether TDP-43 co-localized with CAG repeat RNAs in cells. We combined RNA fluorescence in situ hybridization (FISH) and immunofluorescence assays to monitor subcellular localizations of CAG repeat RNA and TDP-43 protein, respectively. The results showed more extensive co-localization between endogenous TDP-43 and $(CAG)_{38}$ RNA than with $(CAG)_{22}$ RNA (Fig. 2a,c). Additionally, the percentage of this co-localization was significantly attenuated upon ectopic expression of ALKBH3 (Fig. 2a,c) or upon shRNA-mediated knockdown of TRMT61A (Fig. 2a,c). Similar observations were made for ectopically expressed GFP–TDP-43 upon ectopic expression of ALKBH3 (Extended Data Fig. 5c,d); however, ectopically expressed GFP–TDP-43-5FL exhibited impaired co-localization with CAG repeat RNA (Extended Data Fig. 5c,d). This is in keeping with the inability of TDP-43-5FL to bind with CAG repeat RNA (Extended Data Fig. 3j,k).

Together, these results substantiate that CAG repeat RNA binds to TDP-43 in cells, which is driven by m^1A in the repeat RNA and requires the functional RRM domains of TDP-43. Additionally, CAG repeat RNAs promote the cytoplasmic redistribution and truncation of TDP-43, which again require m^1A in CAG repeat RNA and its interaction with TDP-43.

m^1A drives TDP-43 to stress granules

Stress granules are phase-separated compartments in the cytosol[18,36]; we next examined whether CAG repeat RNA and endogenous TDP-43 are localized to stress granules. We observed cytoplasmic foci for endogenous G3BP1 and its co-localization with endogenous TDP-43 in U2OS cells upon expression of $(CAG)_{22}$ and $(CAG)_{38}$, with the extent of co-localization being more pronounced for cells expressing $(CAG)_{38}$ than $(CAG)_{22}$ (Fig. 3a,c). Furthermore, endogenous TDP-43 exhibited diminished co-localization with stress granules upon overexpression of ALKBH3 or genetic depletion of TRMT61A (Fig. 3a,c).

We next determined whether m^1A-marked RNA was also sequestered into stress granules. Our RNA-FISH and immunofluorescence assay results showed that endogenous G3BP1 granules were co-localized with $(CAG)_{38}$ RNA (Fig. 3b,d). Moreover, the extent of this co-localization was substantially attenuated in cells upon ectopic expression of ALKBH3 or upon shRNA-mediated knockdown of TRMT61A (Fig. 3b,d). Together, these results demonstrated that expression of RNA with expanded CAG repeat is capable of inducing stress granules in cells without heat shock or arsenite treatment. In addition, m^1A in CAG repeat RNA promotes the sequestration of TDP-43 into stress granules.

To substantiate the role of m^1A in CAG repeat RNA in influencing the subcellular redistribution and biophysical properties of TDP-43, we used synthetic $(CAG)_7$ and $(CAG)_{16}$ RNA carrying zero or three m^1A residues. Transfection of cells with the synthetic CAG repeat RNAs led to cytoplasmic redistribution of TDP-43 in a sub-population of cells, and transfection with $(CAG)_7$-3m^1A or $(CAG)_{16}$-3m^1A RNA conferred larger TDP-43 foci in the cytosol than with the same amount of the corresponding unmodified RNAs (Fig. 4a,b). The observation of cytoplasmic redistribution of TDP-43 in cells transfected with the unmodified CAG repeat RNAs may be owing to the modification of some of the rA in the repeat RNAs to m^1A in cells after transfection.

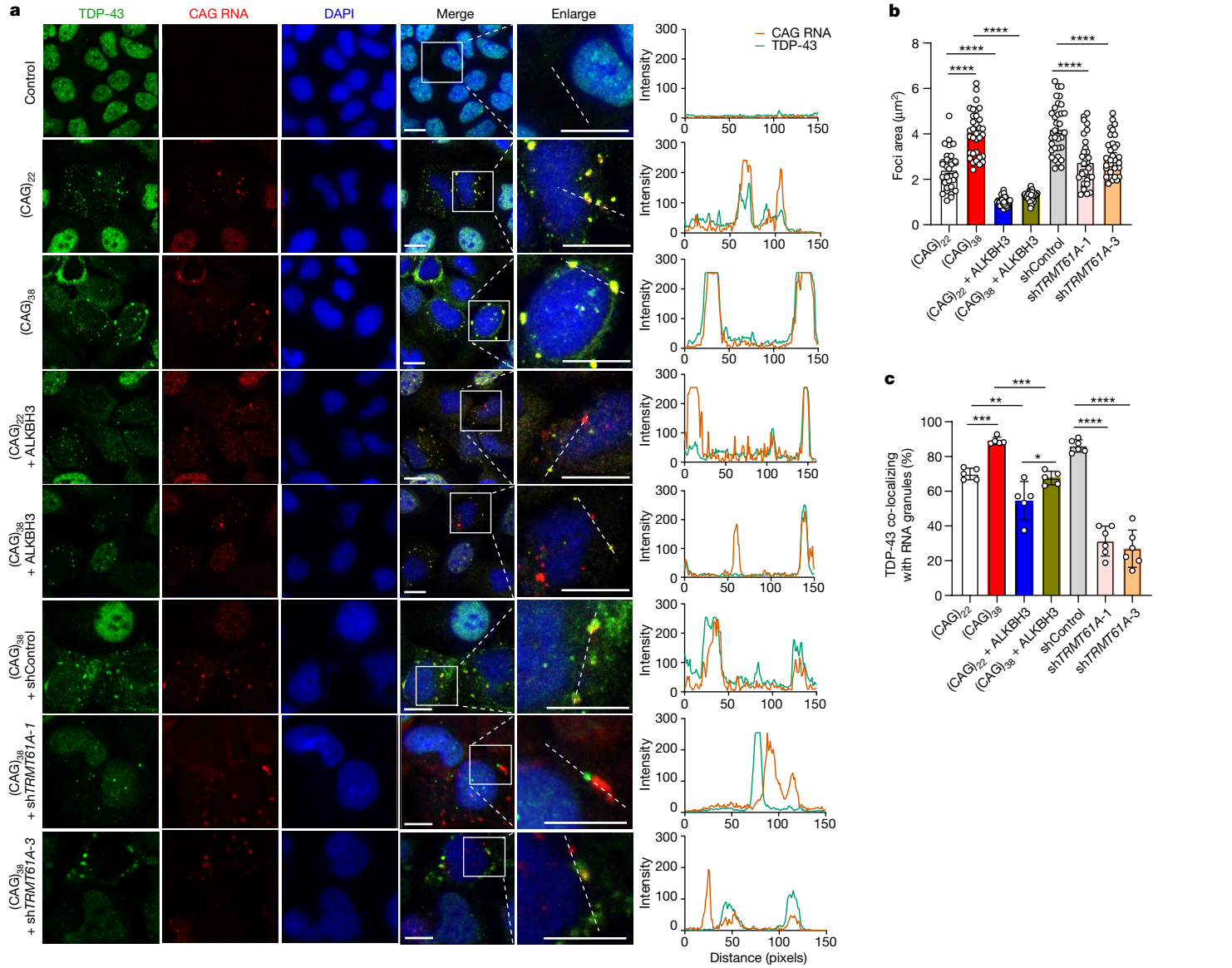


Fig. 2 | m^{1}A induces cytoplasmic mis-localization and aggregation of endogenous TDP-43. a, Representative images of U2OS cells expressing CAG repeat RNA along with ectopic expression of ALKBH3 or shRNA-mediated knockdown of TRMT61A. Scale bars, 10 μm. **b**, Quantification of area of TDP-43 foci in **a**. $(CAG)_{22}$: $n = 31$; $(CAG)_{38}$: $n = 33$; $(CAG)_{22}$ + ALKBH3: $n = 32$; $(CAG)_{38}$ + ALKBH3: $n = 30$; shControl: $n = 33$; sh*TRMT61A-1*: $n = 31$; sh*TRMT61A-3*: $n = 32$. **c**, Percentage of TDP-43 co-localized with CAG repeat RNAs. Images are representative of five or six independent frames for each condition. $n = 5$ for $(CAG)_{22}$, $(CAG)_{38}$, $(CAG)_{22}$ + ALKBH3 and $(CAG)_{38}$ + ALKBH3; $n = 6$ for shControl, sh*TRMT61A-1* and sh*TRMT61A-3*. Data are mean ± s.d. and represent three biologically independent experiments. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. $^*P = 0.020$, $^{**}P = 0.0059$ and $^{***}P = 0.0009$ between $(CAG)_{22}$ and $(CAG)_{38}$; $^{***}P = 0.0003$ between $(CAG)_{38}$ and $(CAG)_{38}$ + ALKBH3.

We also observed co-localization between endogenous TDP-43 and G3BP1 after transfecting cells with the synthetic CAG repeat RNAs, with a higher degree of co-localization being observed in cells transfected with $(CAG)_{16}$-3m^{1}A than in those with the corresponding unmodified RNA (Fig. 4a). Together, these results support the proposition that m^{1}A is a crucial molecular determinant in inducing cytoplasmic mis-localization of TDP-43 and its co-localization with stress granules.

m^{1}A alters phase separation of TDP-43

We next explored whether TDP-43 exhibits phase-separated liquid-like properties in cells and how these properties may be modulated by the interaction of TDP-43 with m^{1}A in CAG repeat RNA. We first examined the physical properties of GFP–TDP-43 in cells expressing RNA with expanded CAG repeats. Our results showed that after photobleaching, the fluorescence of GFP–TDP-43 was rapidly recovered in cells with ectopic expression of $(CAG)_{22}$ RNA (Extended Data Fig. 6a,b), suggesting that a large fraction of TDP-43 protein is mobile in cells expressing $(CAG)_{22}$ RNA. However, we observed little recovery of GFP–TDP-43 fluorescence after photobleaching in cells expressing $(CAG)_{38}$ RNA, indicating that TDP-43 protein is immobile in these cells (Extended Data Fig. 6a,b). Furthermore, ectopic co-expression of ALKBH3 with $(CAG)_{38}$ restored rapid fluorescence recovery of GFP–TDP-43 after photobleaching (Extended Data Fig. 6a,b).

To further investigate the biophysical properties of TDP-43 granules, we treated cells that ectopically co-expressed TDP-43 and $(CAG)_{38}$ or $(CAG)_{22}$ with 1,6-hexanediol (1,6-HD), which disrupts liquid-like condensates[37]. Upon treatment with 1,6-HD, there were fewer GFP–TDP-43 granules in cells expressing $(CAG)_{22}$ (Extended Data Fig. 6c), again substantiating that TDP-43 is mobile and exhibits liquid-like properties

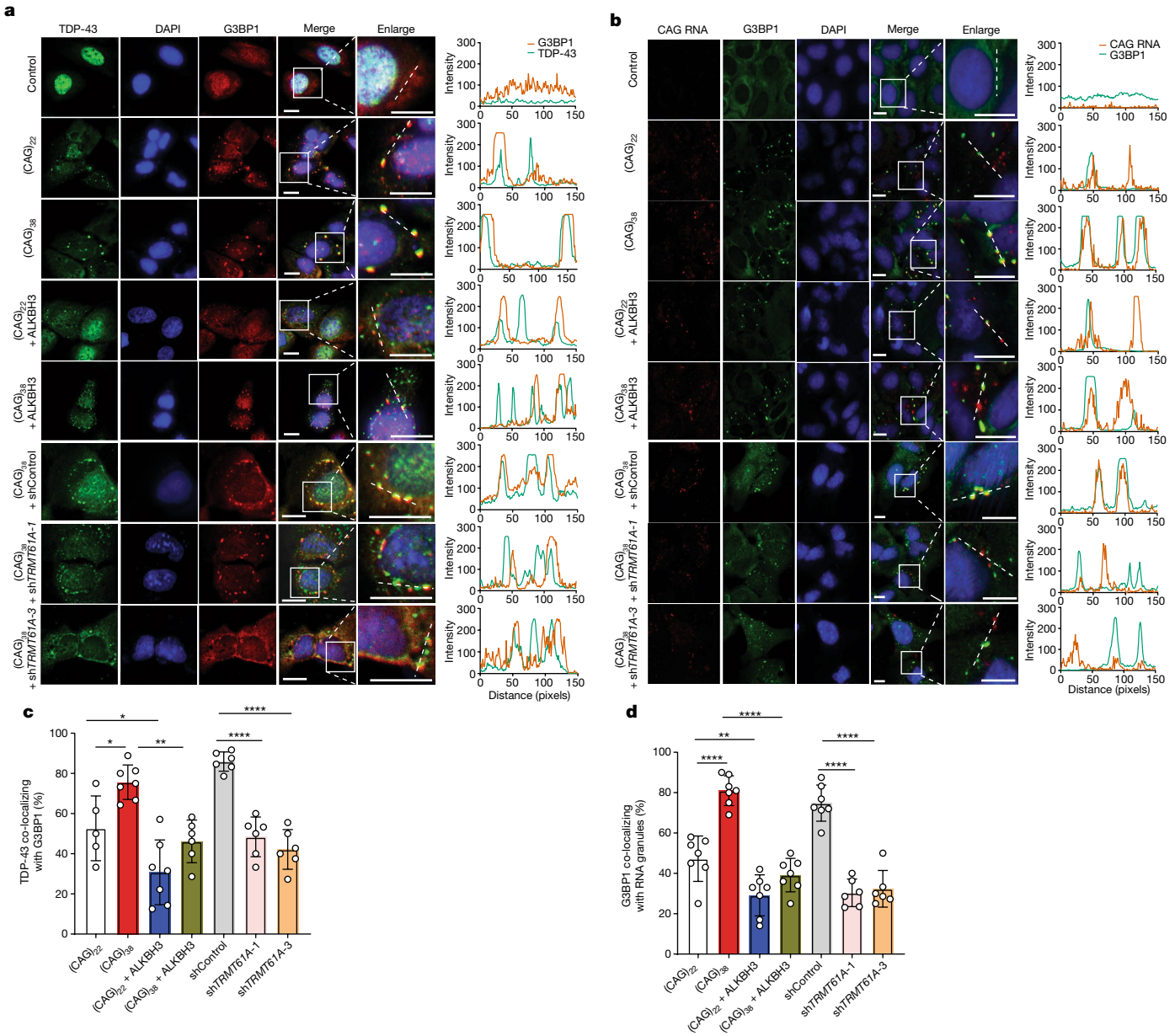


Fig. 3 | m^{1}A enhances the ability of endogenous TDP-43 protein to partition into stress granules. a, CAG repeat RNA-mediated localization of endogenous TDP-43 into stress granules in U2OS cells with ectopic expression of ALKBH3 or knockdown of TRMT61A. **b**, FISH and immunofluorescence microscopy were performed to assess the co-localization between G3BP1 and CAG repeat RNAs in U2OS cells with ectopic expression of ALKBH3 or knockdown of TRMT61A. **c**, Percentages of TDP-43 co-localized with G3BP1. Images are representative of 5 to 7 independent frames for each condition: $n = 5$ for $(CAG)_{22}$; $n = 7$ for $(CAG)_{38}$ and $(CAG)_{22}$ + ALKBH3; $n = 6$ for $(CAG)_{38}$ + ALKBH3, shControl, sh*TRMT61A-1* and sh*TRMT61A-3*. **d**, Percentages of G3BP1 co-localized with CAG repeat RNAs. Images are representative of 6 or 7 independent frames for each condition: $n = 7$ for $(CAG)_{22}$, $(CAG)_{38}$, $(CAG)_{22}$ + ALKBH3, $(CAG)_{38}$ + ALKBH3 and shControl; $n = 6$ for sh*TRMT61A-1* and sh*TRMT61A-3*. Data are mean ± s.d. and represent three biologically independent experiments. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. **c**, $^*P = 0.033$ between $(CAG)_{22}$ and $(CAG)_{38}$; $^*P = 0.043$ between $(CAG)_{22}$ and $(CAG)_{22}$ + ALKBH3; $^{**}P = 0.0030$. **d**, $^{**}P = 0.0072$. Scale bars, 10 µm.

in these cells. The granules in cells expressing $(CAG)_{38}$, however, displayed markedly reduced sensitivity toward 1,6-HD (Extended Data Fig. 6c,d). Moreover, we found that truncated TDP-43 lacking the LCD (TDP-43-ΔLCD), along with $(CAG)_{38}$ RNAs, was located mainly in the nucleus and exhibited a diffused pattern (Extended Data Fig. 6e). Together, this suggests that long CAG repeat RNA expansions facilitate a liquid-to-gel-like transition of TDP-43, which entails m^{1}A in the repeat RNA and the LCD of TDP-43.

We next tested whether m^{1}A regulates phase separation of TDP-43 protein in vitro using synthetic $(CAG)_7$ RNA containing 0, 1 or 3 m^{1}A residues. Although the unmodified $(CAG)_7$ RNA did not alter the phase separation of TDP-43 (Extended Data Fig. 7a), the corresponding RNAs carrying one or three m^{1}A residues triggerred the formation of large droplets and increased the partition coefficient of TDP-43 in the droplets (Extended Data Fig. 7a,b). Unlike wild-type TDP-43, the phase separation of TDP-43-5FL was not influenced by the presence of CAG repeat RNAs (Extended Data Fig. 7a,b). Moreover, TDP-43-ΔLCD did not form droplets in vitro; this effect was independent of the presence of CAG repeat RNAs (Extended Data Fig. 7c).

We also examined whether synthetic m^{1}A-containing RNAs influence the biophysical properties of the TDP-43 droplets in vitro. We observed that the protein droplets fused to form larger droplets within 2.5 s in the

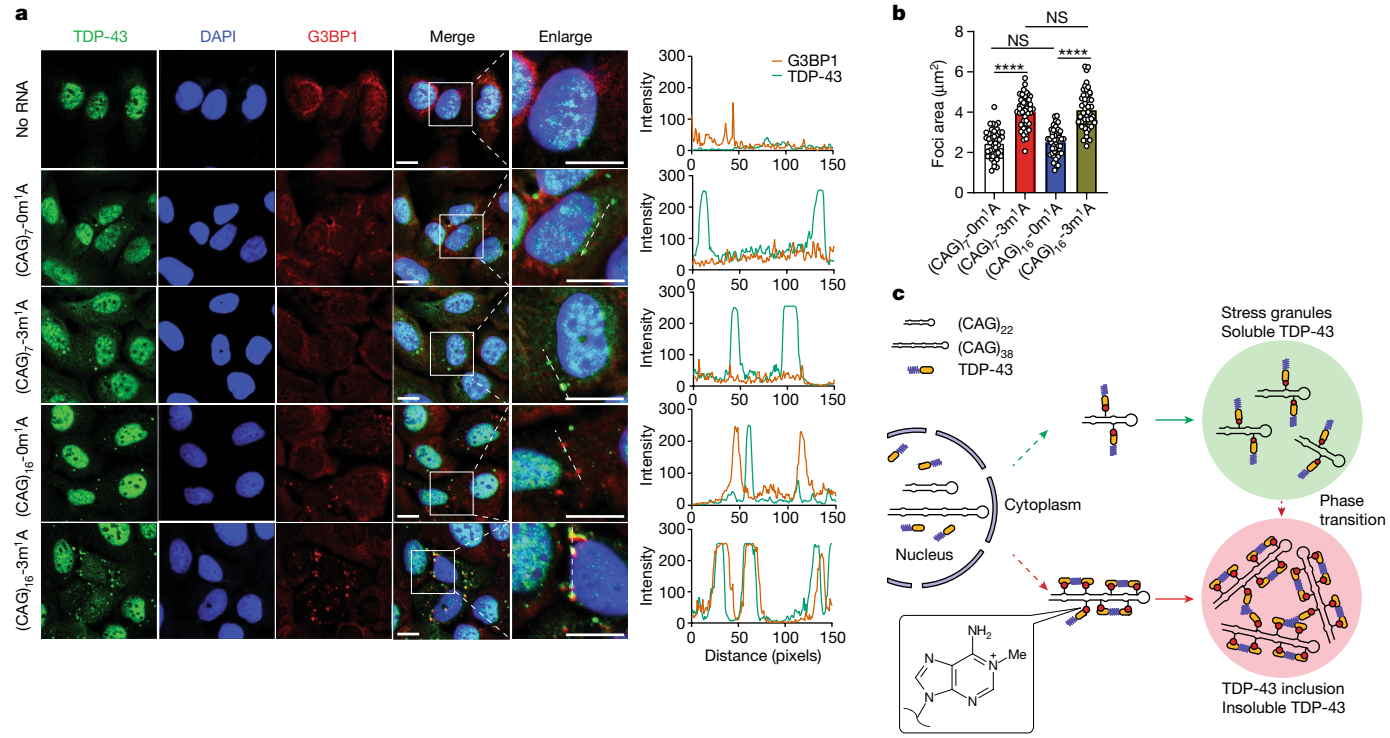


Fig. 4 | Synthetic m^{1}A-containing CAG repeat RNA triggers cytoplasmic redistribution of endogenous TDP-43 and its co-localization with stress granules, and a proposed model illustrating a role of m^{1}A in CAG repeat length-dependent modulation of biophysical properties of TDP-43. **a**, Representative images showing the localization of TDP-43 and G3BP1 in U2OS cells with or without transfection with synthetic CAG repeat RNAs containing zero or three m^{1}A. Scale bars, 10 μm. **b**, Sizes of TDP-43 foci induced by synthetic CAG repeat RNA containing zero or three m^{1}A. $n = 45$ for $(CAG)_7$-0m^{1}A and $(CAG)_7$-3m^{1}A; $n = 46$ for $(CAG)_{16}$-0m^{1}A and $(CAG)_{16}$-3m^{1}A. Data are mean ± s.d. and represent three biological replicates. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. **c**, A model illustrating expanded CAG repeat RNA-induced aberrant phase separation and cytoplasmic redistribution of TDP-43. TDP-43 bound to $(CAG)_{22}$ RNA maintains liquid-like state and is soluble. $(CAG)_{38}$ RNA triggers aberrant phase transition of TDP-43 into insoluble inclusions, in which m^{1}A modification in CAG repeat RNA has an important role in stimulating the liquid-to-gel-like transition of TDP-43 by increasing the local concentration of the protein, thereby promoting its aggregation through its LCD.

presence of $(CAG)_7$ RNA with 0 or 1 m^{1}A, indicating that TDP-43 protein exhibits LLPS (Extended Data Fig. 7d and Supplementary Videos 1–3). The TDP-43 droplets, however, coalesced more slowly upon incubation with $(CAG)_7$-3m^{1}A (Extended Data Fig. 7d and Supplementary Video 4). Photobleaching experiments revealed that TDP-43 assembles into gel-like aggregates upon incubation with $(CAG)_7$-3m^{1}A (Extended Data Fig. 7e,f and Supplementary Videos 5–8). These in vitro results corroborate our findings from cellular experiments and demonstrate that m^{1}A in CAG repeat RNA perturbs the phase separation of TDP-43 protein by rendering the protein much less mobile.

Our findings parallel a recent observation that multivalent m^{6}A-containing RNAs can enhance the phase separation potential of YTHDF proteins in vitro and in cells[38,39]. Notably, our results revealed a repeat length-dependent accumulation of m^{1}A in CAG repeat RNA, and a novel pathological function of m^{1}A—that is, in binding with TDP-43 and eliciting its aberrant biochemical and biophysical properties that recapitulate the observations made for the protein in many neurological diseases (Fig. 4c). It is important to note that neurotoxicity observed in CAG repeat expansion disorders may also arise from polyglutamine-containing proteins translated from the repeat-bearing RNAs[3,4], and the ensuing perturbation of the ubiquitin–proteasome system[11]. Thus, multiple mechanisms probably contribute to neurodegenerative diseases caused by nucleotide repeat expansions.

In sum, our work shifts the current paradigm of TDP-43 proteinopathy and suggests new ways to treat neurological diseases. Our identification of the methyltransferase (TRMT61A) and demethylase (ALKBH3) for m^{1}A in CAG repeat RNA provide a strong foundation for such future efforts.

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/s41586-023-06701-5.

1. 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).
2. Zoghbi, H. Y. & Orr, H. T. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* **23**, 217–247 (2000).
3. DiFiglia, M. et al. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277**, 1990–1993 (1997).
4. Scherzinger, E. et al. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell* **90**, 549–558 (1997).
5. Lin, X. et al. Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.* **15**, 2087–2097 (2006).
6. Li, L. B., Yu, Z., Teng, X. & Bonini, N. M. RNA toxicity is a component of ataxin-3 degeneration in *Drosophila*. *Nature* **453**, 1107–1111 (2008).
7. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
8. Wojciechowska, M. & Krzyzosiak, W. J. Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum. Mol. Genet.* **20**, 3811–3821 (2011).
9. Swinnen, B., Robberecht, W. & Van Den Bosch, L. RNA toxicity in non-coding repeat expansion disorders. *EMBO J.* **39**, e101112 (2020).
10. DeJesus-Hernandez, M. et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of *C9ORF72* causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
11. Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**, 1552–1555 (2001).
12. Neumann, M. et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).

13. Lee, E. B., Lee, V. M. & Trojanowski, J. Q. Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nat. Rev. Neurosci.* **13**, 38–50 (2011).
14. St-Amour, I., Turgeon, A., Goupil, C., Planel, E. & Hebert, S. S. Co-occurrence of mixed proteinopathies in late-stage Huntington's disease. *Acta Neuropathol.* **135**, 249–265 (2018).
15. Becker, L. A. et al. Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice. *Nature* **544**, 367–371 (2017).
16. Elden, A. C. et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466**, 1069–U1077 (2010).
17. Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **24**, 1537–1549 (2016).
18. Ramaswami, M., Taylor, J. P. & Parker, R. Altered ribostasis: RNA–protein granules in degenerative disorders. *Cell* **154**, 727–736 (2013).
19. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
20. Zaccara, S., Ries, R. J. & Jaffrey, S. R. Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* **20**, 608–624 (2019).
21. Chen, X. Y., Zhang, J. & Zhu, J. S. The role of m^6A RNA methylation in human cancer. *Mol. Cancer* **18**, 103 (2019).
22. Fu, L. et al. Simultaneous quantification of methylated cytidine and adenosine in cellular and tissue RNA by nano-flow liquid chromatography-tandem mass spectrometry coupled with the stable isotope-dilution method. *Anal. Chem.* **87**, 7653–7659 (2015).
23. Li, X. et al. Transcriptome-wide mapping reveals reversible and dynamic N^1-methyladenosine methylome. *Nat. Chem. Biol.* **12**, 311–316 (2016).
24. Liu, F. G. et al. ALKBH1-mediated tRNA demethylation regulates translation. *Cell* **167**, 816–828 (2016).
25. Jia, G. et al. N^6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
26. Zheng, G. et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
27. Ozanick, S., Krecic, A., Andersland, J. & Anderson, J. T. The bipartite structure of the tRNA m^1A58 methyltransferase from *S. cerevisiae* is conserved in humans. *RNA* **11**, 1281–1290 (2005).
28. Chujo, T. & Suzuki, T. Trmt61B is a methyltransferase responsible for 1-methyladenosine at position 58 of human mitochondrial tRNAs. *RNA* **18**, 2269–2276 (2012).
29. Vilardo, E. et al. A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase-extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic Acids Res.* **40**, 11583–11593 (2012).
30. Li, X. et al. Base-resolution mapping reveals distinct m^1A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol. Cell* **68**, 993–1005 (2017).
31. Langfelder, P. et al. Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nat. Neurosci.* **19**, 623–633 (2016).
32. Brignull, H. R., Moore, F. E., Tang, S. J. & Morimoto, R. I. Polyglutamine proteins at the pathogenic threshold display neuron-specific aggregation in a pan-neuronal *Caenorhabditis elegans* model. *J. Neurosci.* **26**, 7597–7606 (2006).
33. Dai, X., Wang, T., Gonzalez, G. & Wang, Y. Identification of YTH domain-containing proteins as the readers for N^1-methyladenosine in RNA. *Anal. Chem.* **90**, 6380–6384 (2018).
34. Aas, P. A. et al. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421**, 859–863 (2003).
35. Liu, J. et al. A METTL3–METTL14 complex mediates mammalian nuclear RNA N^6-adenosine methylation. *Nat. Chem. Biol.* **10**, 93–95 (2014).
36. Protter, D. S. W. & Parker, R. Principles and properties of stress granules. *Trends Cell Biol.* **26**, 668–679 (2016).
37. Kroschwald, S., Maharana, S. & Simon, A. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters* https://doi.org/10.19185/matters.201702000010 (2017).
38. Ries, R. J. et al. m^6A enhances the phase separation potential of mRNA. *Nature* **571**, 424–428 (2019).
39. Gao, Y. F. et al. Multivalent m^6A motifs promote phase separation of YTHDF proteins. *Cell Res.* **29**, 767–769 (2019).

Methods

Cell culture and shRNA knockdown

HEK293T cells and U2OS cells were obtained from ATCC. The cells were maintained in DMEM (11995-065, Gibco) with 10% FBS, 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin under standard cell culture conditions (37 °C, 5.0% CO_2). The cell lines were authenticated by ATCC using short tandem repeat (STR) analysis as described in 2012 in ANSI Standard (ASN-0002) Authentication of Human Cell Lines. The cells were confirmed to be free of mycoplasma contamination by using LookOut Mycoplasma PCR Detection Kit (MP0035, Sigma-Aldrich).

shRNAs against *TRMT61A*, *TRMT61B* and *TRMT10C* were designed and their sequences are listed in Supplementary Table 3. All shRNAs and control non-targeting shRNAs were cloned into the AgeI/EcoRI site of the pLKO.1 vector (Addgene, plasmid #10878) and confirmed by Sanger sequencing. Cells were transfected with pLKO.1/puro-shRNAs together with pLTR-G (Addgene plasmid #17532) envelope plasmid and pCMV-dR8.2 dvpr (Addgene plasmid #8455) package plasmid using PolyFect transfection reagent (QIAGEN). Viral particles were collected 48 h later and filtered through a 0.45-μm sterile filter. After that, cells were transfected with lentiviral constructs expressing shRNA for 24 h, and selected with puromycin for 7 days.

Extraction of CAG repeat mRNA and LC–MS/MS measurements

Total RNA was extracted from HEK293T cells using TRI reagent (Sigma). The RNA sample (~20 μg) was subsequently incubated with 200 μM of 5′-biotinylated 5×CTG oligodeoxyribonucleotides in a 2-volume hybridization buffer (50 mM Tris-HCl, pH 7.0, 750 mM NaCl, 1 mM EDTA, 1% SDS and 15% formamide) at room temperature for 4 h with gentle mixing. To the resulting mixture were then added streptavidin-conjugated agarose beads (Thermo Scientific) and the suspension was incubated at 4 °C for 2 h. After the incubation, the oligodeoxyribonucleotide-bound beads were washed for four times with 2× SSC buffer (0.30 M NaCl, 15 mM sodium citrate, pH 7.0) and the beads was resuspended in RNase-free water for digestion.

For the cellular pull-down experiment, HEK293T cells were transfected with Flag-tagged TDP-43 or 5FL mutant plasmid. After a 24-h incubation, the cells were washed with PBS and lysed in a buffer containing 10 mM HEPES (pH 7.5), 150 mM KCl, 2 mM EDTA, 0.5% IGEPAL CA-630, 0.5 mM DTT, protease inhibitor (Sigma), and 40 units ml^{-1} RNase inhibitor. The supernatant was incubated with anti-Flag M2 beads (Sigma) at 4 °C overnight. The beads were washed for three times with a buffer containing 50 mM HEPES (pH 7.5), 200 mM NaCl, 2 mM EDTA, 0.05% IGEPAL CA-630, and 0.5 mM DTT. The TDP-43-bound RNA was extracted from the beads using TRI reagent, and mRNA was isolated and purified by using PolyATtract mRNA Isolation System IV (Promega) according to the manufacturer's instructions.

The above affinity-purified RNA (400 ng) was digested with 1 unit of nuclease P1 in 25 μl buffer containing 25 mM NaCl and 2.5 mM $ZnCl_2$. The mixture was then incubated at 37 °C for 2 h, and to the mixture were added 0.5 unit of Antarctic phosphatase and 3 μl of 1.0 M NH_4HCO_3. After incubation at 37 °C for an additional 2 h, the digestion mixture was dried and reconstituted in 100 μl ddH_2O. [$^{13}C_5$]-adenosine, [D_3]-m^6A, and [D_3]-m^1A were used as internal standards for the quantifications of rA, m^6A, and m^1A, respectively. In parallel, total RNA (400 ng as input control RNA) was digested in the same way. The enzymes in the digestion mixture were removed by extraction using chloroform:isoamyl alcohol (24:1), and salt in the samples was removed by acetonitrile precipitation. The resulting supernatant was dried, and the dried residues were reconstituted in 10 μl ddH_2O and injected for LC–MS/MS analysis on a TSQ Altis triple-quadrupole mass spectrometer (Thermo). For m^1A and m^6A quantifications, the pre-column and analytical column were packed with porous graphitic carbon and Zorbax SB-C18 stationary phase materials, respectively, where a gradient of 0–15% B in 10 min, 15–95% B in 30 min, and 95% B in 10 min was used. The mass spectrometer was operated in the multiple-reaction monitoring (MRM) mode, where the loss of a ribose from the $[M + H]^+$ ions of rA, m^6A, m^1A and their stable isotope-labelled counterparts were monitored.

RNase H was used to remove the 5′ and 3′ flanking RNA sequences of the $(CAG)_{38}$ repeat. In particular, the isolated RNA was annealed in the RNase H reaction buffer (New England Biolabs) by heating to 95 °C and cooling slowly to 70 °C. The temperature was held at 70 °C for 10 min to melt nonspecific RNA structures, before cooling slowly to 37 °C. RNase H was added to the mixture, incubated at 37 °C for 2 h, and the reaction was terminated with 0.5 M EDTA. $(CAG)_{38}$ RNA was subsequently isolated by affinity purification, digested with enzymes, and analysed by LC–MS/MS, as described above. Unless specifically noted, the reported levels for m^1A and m^6A in CAG repeat RNA were without RNase H treatment.

C57BL/6 mice expressing one wild-type endogenous Htt allele and a second Htt allele with knock-in of human *mHTT* exon 1 containing 140 CAG repeats were described previously[31,40]. All mice were maintained and bred under standard conditions consistent with National Institutes of Health guidelines and approved by the University of California, Los Angeles Institutional Animal Care and Use Committee. The cages were maintained on a 12:12 light:dark cycle, with food and water ad libitum. Striatum and cortex tissues were harvested from wild-type and heterozygous HD knock-in Q140 mice at six months of age and immediately frozen on dry ice, where tissues from both male and female mice were used. Q19 (AM49, rmls172), Q40 (AM101, rmls110) and Q67 worm strains (AM44, rmls190) were purchased from *Caenorhabditis* Genetics Center (CGC). *Drosophila* expressing a SCA3 transgene with 78 repeats of CAG (UAS-MJDtrQ78) and those with a short (UAS-HA-MJDtrQ27 line c19.1) CAG repeat (Bloomington *Drosophila* Stock Center line 8149) have been described[6,41], where the transgene was expressed in neurons using ElavGal4 (Bloomington line 458). Coding sequences of wild-type and catalytically inactive mutant human *ALKBH3* were subcloned into a pUAST transformation vector, and transgenic fly lines were subsequently generated (chrIII, attP2). CAG repeat RNA was isolated from total RNA extracted from mouse striatum and cortex tissues, *Drosophila* heads and *C. elegans*, digested, and analysed in a similar way as described above for cellular RNA.

The levels of rA, m^6A and m^1A were quantified by employing calibration curves obtained from pure nucleoside standards analysed under the same instrument conditions (Supplementary Figs. 2 and 3). We also assessed the recovery rate of the analytical workflow in measuring m^1A and m^6A in CAG repeat RNA by mixing $(CAG)_7$-0m^1A with $(CAG)_7$-1m^1A and $(CAG)_7$-1m^6A at defined m^1A/rA and m^6A/rA molar ratios, and quantified the levels of m^1A and m^6A by following the same procedures as described above for CAG repeat RNA isolated from cells and tissues. The recoveries of m^1A and m^6A were determined to be approximately 85% and 87%, respectively (Supplementary Fig. 12). Given the relatively high recovery rates, the reported measurement results for m^1A and m^6A were without correction for the recovery rates.

Pull-down of TDP-43 with CAG repeat RNA and western blot analysis

HEK293T cells were transfected with a plasmid encoding Flag–TDP-43 or Flag–TDP-43-5FL together with the indicated CAG repeat plasmid. The cells were harvested at 70–80% confluence, washed with PBS, and lysed in CelLytic M cell lysis reagent (Sigma). The lysates were centrifuged at 13,000 rpm for 10 min at 4 °C. The supernatant was pre-cleared by incubating with streptavidin-conjugated agarose beads (Thermo Scientific) at 4 °C for 1 h, and 20 μl samples were taken as inputs. Biotinylated RNA baits (3 μg) were incubated, at room temperature for 4 h, with pre-cleared cell lysates in a hybridization buffer containing protease and RNase inhibitors. Streptavidin-conjugated agarose beads were then added to the mixture, and the mixture was kept in a shaker at 4 °C for 2 h. The beads were washed extensively and boiled for 10 min, and the supernatant collected. The samples were subsequently resolved

by 10% SDS–PAGE and transferred to a PVDF membrane (Durapore membrane filter, 0.45 µm, Millipore). Membranes were blocked in 5% milk at room temperature for 1 h, and then incubated with primary antibodies in 5% BSA at 4 °C overnight. Primary antibodies for the following proteins or epitope were used: TDP-43 (Proteintech, 10782-2-AP, 1:1,000), G3BP1 (Proteintech, 66486-1-lg, 1:1,000), TRMT61A (Thermo Fisher, A305-858A-T, 1:1,000), Flag epitope (Cell Signaling Technology, 14793 S, 1:2,000), ALKBH3 (Cell Signaling Technology, 87620, 1:1,000), GAPDH (Santa Cruz Biotechnology, sc-32233, immunoblot, 1:5,000), and α-tubulin (Santa Cruz Biotechnology, sc-32293, 1:5,000). After washing with PBS containing 0.05% Tween (PBS-T), the membranes were incubated with secondary antibodies at room temperature for 1 h. Secondary antibodies used for western blotting included anti-rabbit IgG (whole molecule)-Peroxidase antibody produced in goat (Sigma, A0545, 1:2500) and m-IgGκ BP-HRP (Santa Cruz Biotechnology, sc-516102, 1:2500). Membranes were then washed with PBS-T and bands were visualized using a LI-COR imaging system and quantified by using ImageJ. All western blots shown are representative results obtained from at least three biological replicates.

Detergent solubility fractionation assay of TDP-43

Detergent solubility fractionation of TDP-43 was performed as described[42]. In brief, cells were washed once with ice-cold PBS, lysed with CelLytic M cell lysis reagent (Sigma), and incubated on ice for 10 min. Following a brief sonication on ice, lysates were centrifuged for 40 min at 15,000*g* at 4 °C. Supernatants were collected as the detergent-soluble fraction. The pellets were subsequently resuspended in a urea buffer (30 mM Tris pH 7.5, 7 M urea, and protease inhibitor cocktail), briefly sonicated on ice, and lysates were centrifuged at 15,000*g* at room temperature for 1 h, where supernatant was collected as the detergent-insoluble, urea-soluble fraction. Proteins in the detergent-soluble and insoluble fractions were then resolved by SDS–PAGE and analysed by western blot.

Expression and purification of recombinant proteins

Recombinant TDP-43 proteins were purified by following previously described procedures[43]. For purification of full-length TDP-43–MBP–6×His, TDP-43-5FL–MBP–His, TDP-43-5FL–eGFP–MBP–His and TDP-43-ΔLCD–eGFP–MBP–His proteins, BL21-DE3 *Escherichia coli* cells were cultured at 37 °C to an OD at 600 nm of 0.6–0.9, and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18 °C for 16 h. The cells were subsequently collected, pelleted and resuspended in a binding buffer containing 20 mM Tris-HCl (pH 8.0), 1 M NaCl, 10% (v/v) glycerol, 1 mM DTT, and EDTA-free protease inhibitor cocktail according to the manufacturer's instructions. The cells were lysed by sonication and centrifuged at 10,000*g* for 20 min. The proteins were purified over pre-packed Dextrin Sepharose High Performance MBPTrap HP (MBPTrap HP columns, GE) and eluted by using the binding buffer containing 10 mM maltose. Proteins were further purified over pre-packed Ni Sepharose High Performance HisTrap HP (GE) and eluted with a buffer containing 50 mM Tris (pH 7.4), 500 mM NaCl and 400 mM imidazole. Protein concentration was determined by Bradford assay (Bio-Rad). The eluted fractions were analysed by using denaturing SDS–PAGE, and the purities of the recombinant proteins were assessed by SDS–PAGE analysis.

Electrophoretic mobility shift assay

RNA probe, 5′-biotin-CCGUUCCGCCCXGGCCGCGCCCAGCUGGAAUG CA-3′ (where X = m^1A, A or m^6A; Supplementary Table 2), was radiolabelled at the 3′ terminus, following previously described procedures[44]. In brief, [5′-^{32}P]cytidine-3′,5′-bis(phosphate) (pCp) was prepared by incubating 0.5 µl 0.5 mM cytidine-3′-monophosphate (Cp, Carbosynth) with 1.0 µl 10× T4 PNK buffer (NEB), 1 µl T4 PNK (3′ phosphatase minus, NEB), and 8.25 µl [γ-^{32}P]ATP at 37 °C for 1 h. The mixture was incubated at 65 °C for 10 min. RNA probe was 3′-end labelled in a 30-µl buffer containing 3 µl 10× T4 RNA ligase buffer (NEB), 3 µl 10 mM ATP (NEB), 1 µl 100 mM DTT (Promega), 3 µl DMSO (Sigma), 9.1 µl [5′-^{32}P]pCp, and 3 µl T4 RNA ligase I (NEB) at 4 °C for 24 h. The probes were then purified by using micro bio-spin P-30 columns (Bio-Rad). In addition, synthetic 21-mer $(CAG)_7$ RNA with or without m^1A (Supplementary Table 2) were labelled with ^{32}P on the 5′ termini by following standard procedures and used for EMSA experiments. In brief, the probe (20 fmol) was incubated with increasing concentrations of TDP-43 or its truncated or mutated variants in a binding buffer (10 mM HEPES, pH 8.0, 50 mM KCl, 1 mM EDTA, 5% glycerol, 1 mM DTT, 40 units ml^{-1} RNase inhibitor) at 4 °C for 1 h. The entire 20-µl sample was subjected to electrophoresis on an 8% native polyacrylamide gel and the gel band intensities were quantified using phosphorimager analysis with a Typhoon 9410 Variable Mode Imager. The dissociation constant (K_d) was calculated using Prism software (version 8.0.1).

Quantification of TDP-43 granule size

For quantification of TDP-43 granule size in U2OS cells expressing RNA with different lengths of CAG repeat, the cells were fixed using methanol and 20–35 granules were analysed per condition. RNA granules were enumerated, and the percentages of granules co-localized with the stress granule marker (G3BP1) were calculated. The intensity and area of co-localized granules were calculated using plot profile tool in Fiji.

Fluorescence recovery after photobleaching

FRAP experiments were performed using an LSM 880 laser scanning confocal microscope (Zeiss 880 Inverted Airyscan Fast) coupled with a temperature-, humidity- and CO_2-controlled top-stage incubator for live-cell imaging. Photobleaching was conducted by a 488-nm line from an Argon laser at 50% power. Three regions of interest (ROIs) were defined for these experiments: ROI-1 was the indicated circular region in the droplet; ROI-2 was a circular, un-photobleached region of similar size in the same droplet; and ROI-3 was defined as background and drawn outside of the droplet, where its signal was subtracted from those of ROI-1 and ROI-2. Raw data were processed and plotted using Prism.

RNA-FISH and immunofluorescence microscopy

U2OS cells expressing the indicated RNA were fixed at 24 h following transfection and permeabilized by incubation for 10 min in methanol containing 10% (v/v) acetic acid. RNA was detected using a Cy3-labelled DNA probe (5′-/5Cy3/CTGCTGCTGCTGCTGCTGCTG-3′). Hybridization and washing buffers were obtained from Biosearch Technologies and used following the manufacturer's protocol. For immunofluorescence detection of proteins following RNA-FISH, methanol-fixed cells were stained using antibodies against TDP-43 (Proteintech, 10782-2-AP, 1:100) and G3BP1 (Proteintech, 66486-1-lg, 1:100), and Alexa Fluor 488 and 647-labelled secondary antibodies (Invitrogen, A-32731, 1:250, and Invitrogen, A-32728, 1:250, respectively). The samples were subsequently stained with DAPI and imaged using confocal microscopy as described above. The data were analysed using Zen 2 Blue software (version 2.3).

In vitro phase separation assays

In vitro TDP-43 droplet formation was induced in a buffer containing 20 mM HEPES (pH 7.4), 300 mM KCl, 6 mM $MgCl_2$, 0.02% NP-40, and 50% glycerol. Dextran (10%) was also added as a crowding agent to induce phase separation of TDP-43. The droplets on a coverslip were detected using a 100× oil immersion objective by confocal microscopy. After TDP-43 droplet formation following the addition of m^1A-RNA, partition coefficients were calculated for stably formed TDP-43-containing droplets based on the ratio of intensity of TDP-43 in phase-separated droplets over that located in the immediately adjacent region.

In vitro methyltransferase assay with TRMT6–TRMT61A

The expression plasmid, pET17b-6×HisTrm6-Trm61, was isolated from a strain provided by J. Finer-Moore[45] and was modified to contain a 3C

protease cleavage sequence C-terminal to the 6×His tag. Recombinant TRMT6–TRMT61A was purified following previously described procedures[45]. For the methyltransferase assay, synthetic $(CAG)_{16}$ RNA was annealed by first heating to 95 °C for 5 min and then cooling to 25 °C. The reactions were performed at room temperature overnight in 20 µl buffer containing 50 mM ammonium acetate, 3 mM $MgCl_2$, 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 1 mM *S*-adenosylmethionine, 2.5 µM RNA substrate and 12.5 µM TRMT6–TRMT61A complex. After the reaction, proteins were removed by chloroform extraction, and the RNA was subsequently digested by nuclease P1 and Antarctic phosphatase, and the levels of rA and m^1A were measured by LC–MS/MS, as described above.

C. elegans strains and maintenance

C. elegans strains were maintained at 20 °C on standard nematode growth medium agar plates seeded with *E. coli* OP50 bacteria unless otherwise stated[46]. A list of strains used in this study is provided in Supplementary Table 4.

To generate transgenic *C. elegans* strains (WG291 and WG300; Supplementary Tables 4) expressing human ALKBH3 (hALKBH3) driven by a pan-neuronal snb-1 promoter, we cloned 1.5 kb of DNA upstream of the *snb-1* gene, which is specifically expressed in the somatic nervous system. The cDNA of human *ALKBH3* gene was obtained from HEK293T cells and amplified by PCR. Two introns were inserted into *ALKBH3* to avoid mis-splicing. This *ALKBH3*-WT cDNA was used in site-directed mutagenesis experiments to generate the catalytically inactive mutant of ALKBH3. Blue fluorescence protein (BFP) was fused to the C-terminus of ALKBH3 to monitor the expression of ALKBH3 protein in worms. All fragments with *TBB2* 3′-UTR were ligated by employing PCR walking technique and inserted into pCR2.1-TOPO (Thermo Fisher). The primer sequences are listed in Supplementary Table 1. The plasmid was injected together with the *rol-6 (su1006)* dominant marker plasmid pRF4 into the germline cells of early-adult hermaphrodites, where the injection mixtures contained 50 µg ml^{-1} each of pRF4 and ALKBH3 plasmids[47]. The F_1 progenies of the transgenic strains generated in this manner carried heritable extrachromosomal multi-copy arrays of both the experimental and marker plasmids and were used for microscopy assessment. The gaps between neurons were quantified using ImageJ.

For RNAi experiments, *W02A11.1* RNAi constructs were obtained from the Vidal RNAi library. Hermaphrodite worms were fed *E. coli* OP50 containing an empty control vector (L4440) or expressing double-stranded RNA when they reached adulthood[48]. Plates for RNAi analysis were prepared by supplementing agar with 100 µg ml^{-1} carbenicillin and 1 mM IPTG after autoclave. Worms were synchronized and grown from hatching on RNAi-feeding plates unless otherwise stated. RNAi efficiency was verified by quantitative PCR with reverse transcription and western blot.

Statistical analysis

All statistical analyses were performed in GraphPad Prism (version 8.0.1) or Microsoft Excel 2016. The outcomes of all statistical tests including *P* values and number of samples are included in the figure panels or the corresponding figure legends. Significance was defined as any statistical outcome that resulted in a *P* value of less than 0.05, unless otherwise indicated[23,49].

Reporting summary

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

Data availability

Necessary data for evaluating this study are available in the main text and Supplementary Information. Uncropped gel images are provided in Supplementary Fig. 1. Source data are provided with this paper.

40. Menalled, L. B. et al. Early motor dysfunction and striosomal distribution of huntingtin microaggregates in Huntington's disease knock-in mice. *J. Neurosci.* **22**, 8266–8276 (2002).
41. Warrick, J. M. et al. Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in *Drosophila*. *Cell* **93**, 939–949 (1998).
42. van Eersel, J. et al. Cytoplasmic accumulation and aggregation of TDP-43 upon proteasome inhibition in cultured neurons. *PLoS ONE* **6**, e22850 (2011).
43. Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K. & Yuan, H. S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37**, 1799–1808 (2009).
44. Nilsen, T. W. 3′-end labeling of RNA with [5′-^{32}P]cytidine 3′,5′-bis(phosphate) and T4 RNA ligase 1. *Cold Spring Harb. Protoc.* **2014**, 444–446 (2014).
45. Finer-Moore, J., Czudnochowski, N., O'Connell, J. D. 3rd, Wang, A. L. & Stroud, R. M. Crystal structure of the human tRNA m^1A58 methyltransferase-$tRNA_3^{Lys}$ complex: refolding of substrate tRNA allows access to the methylation target. *J. Mol. Biol.* **427**, 3862–3876 (2015).
46. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
47. Conte, D. Jr, MacNeil, L. T., Walhout, A. J. M. & Mello, C. C. RNA interference in *Caenorhabditis elegans*. *Curr. Protoc. Mol. Biol.* **109**, 26.23.21–26.23.30 (2015).
48. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
49. Dominissini, D. et al. The dynamic N^1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).

Acknowledgements This work was supported by the National Institutes of Health (R35 NS097275 to N.M.B., R01 GM124349 to W.G., R35 ES031707 to Y.W. and R01 NS113612 to X.W.Y.). Huntington's disease research in the Yang laboratory is also supported by Hereditary Disease Foundation, CHDI Foundation, and donations from the families of patients with Huntington's disease. The authors thank R. Hai for access to his microscope; H. Yuan and J. Finer-Moore for providing the expression plasmids for recombinant TDP-43 and TRMT6–TRMT61A proteins; M. Maduro for assistance with RNAi in *C. elegans;* and the Bloomington *Drosophila* Stock Center for fly lines (NIH P40OD018537).

Author contributions Y.W., Y.S. and H.D. designed the experiments and wrote the paper. All authors reviewed and edited the manuscript. Y.S., J. Yin, Y.C. and G.G. performed m^1A and m^6A measurements. H.D. and W.G. carried out *C. elegans* experiments. H.D. conducted western blot and in vitro methyltransferase assay. X.D. analysed SILAC data. X.L. performed knockdown of *TRMT61A*. J. Yuan and F.T. analysed quantitative data. N.W. and X.W.Y. prepared mouse brain tissues. A.E.P. and N.M.B. performed *Drosophila* experiments.

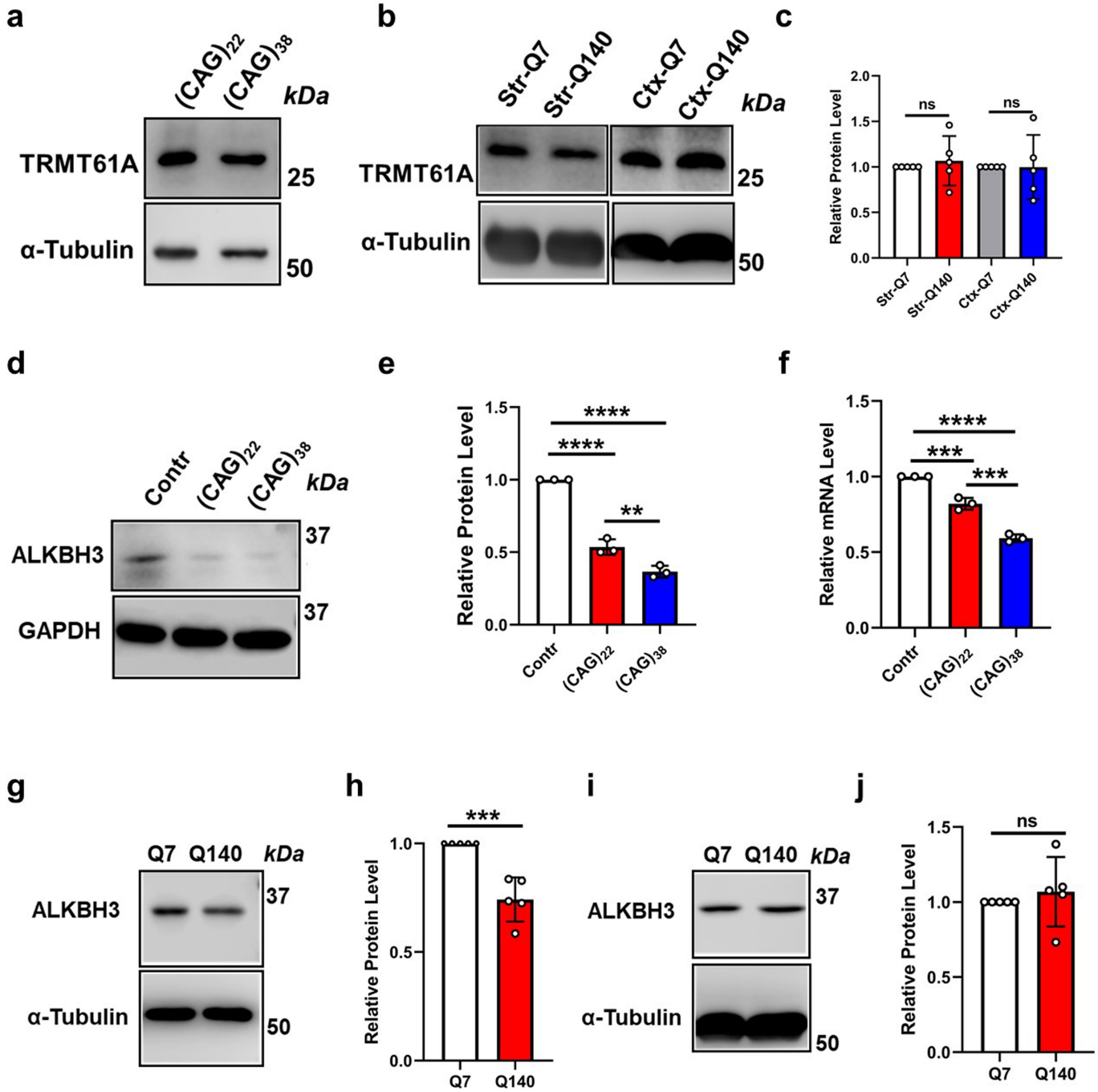
Competing interests The authors declare no competing interests.

Additional information

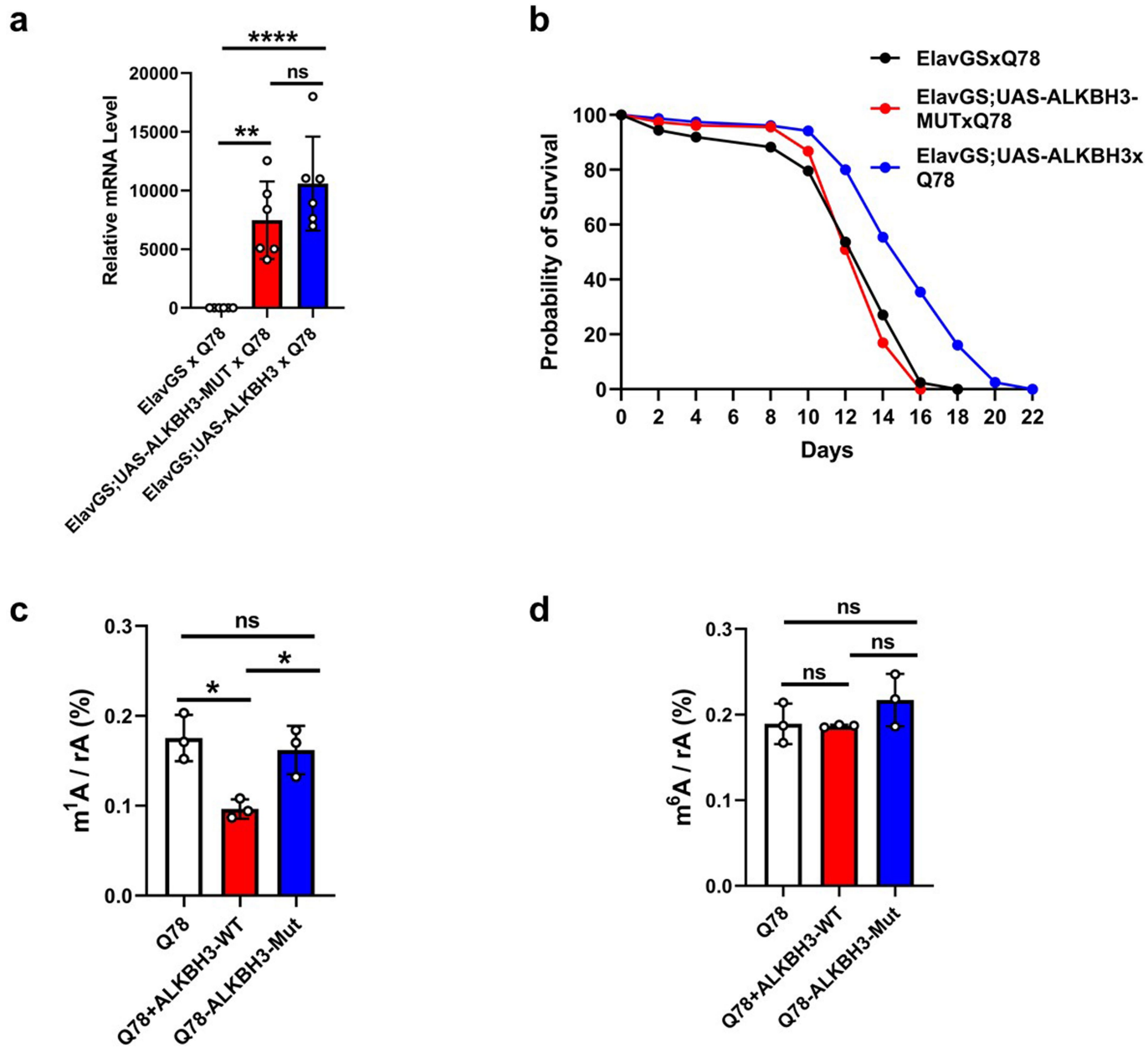
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-023-06701-5.

Correspondence and requests for materials should be addressed to Yinsheng Wang.

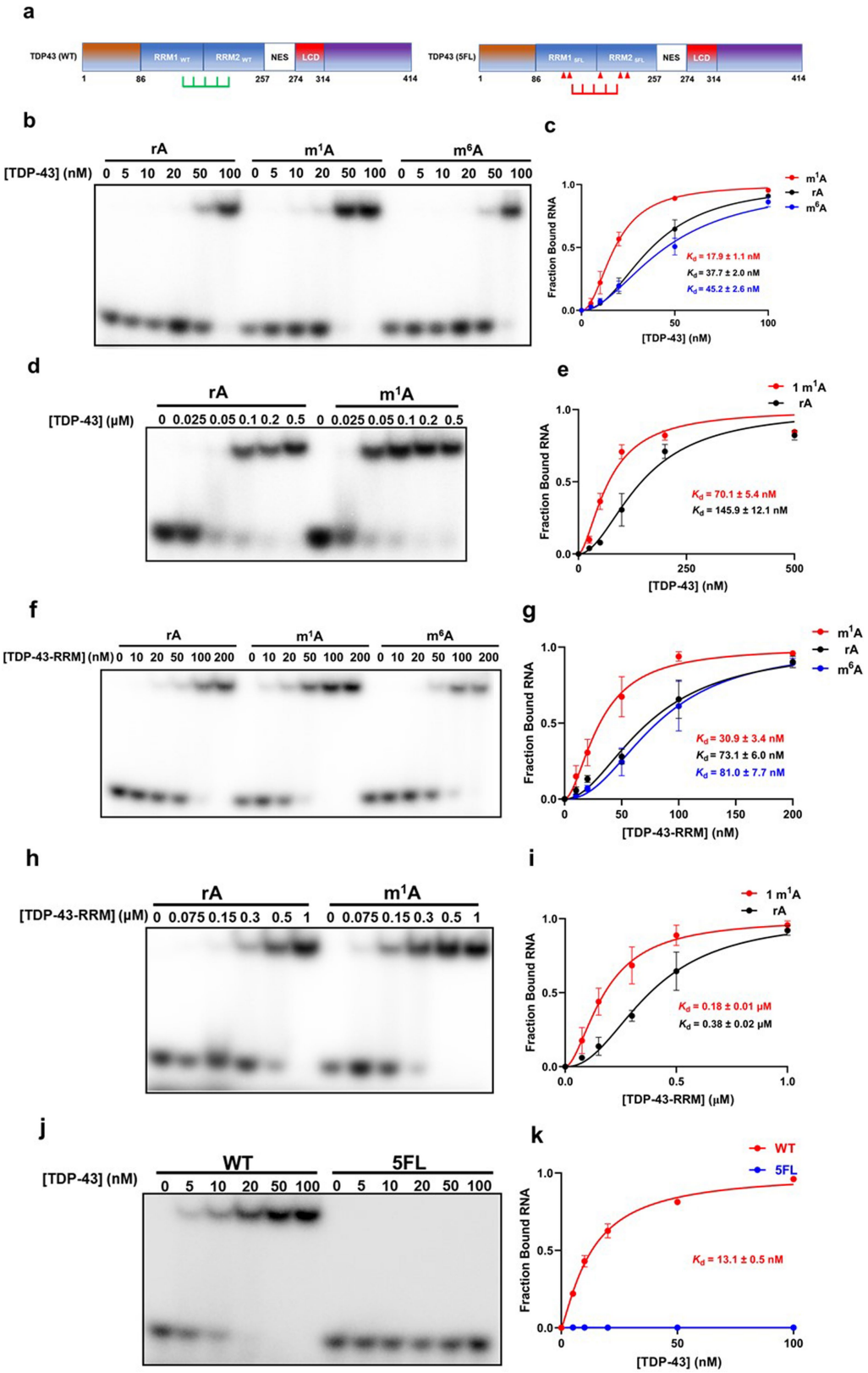
Peer review information *Nature* thanks Maria Carmo-Fonseca and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



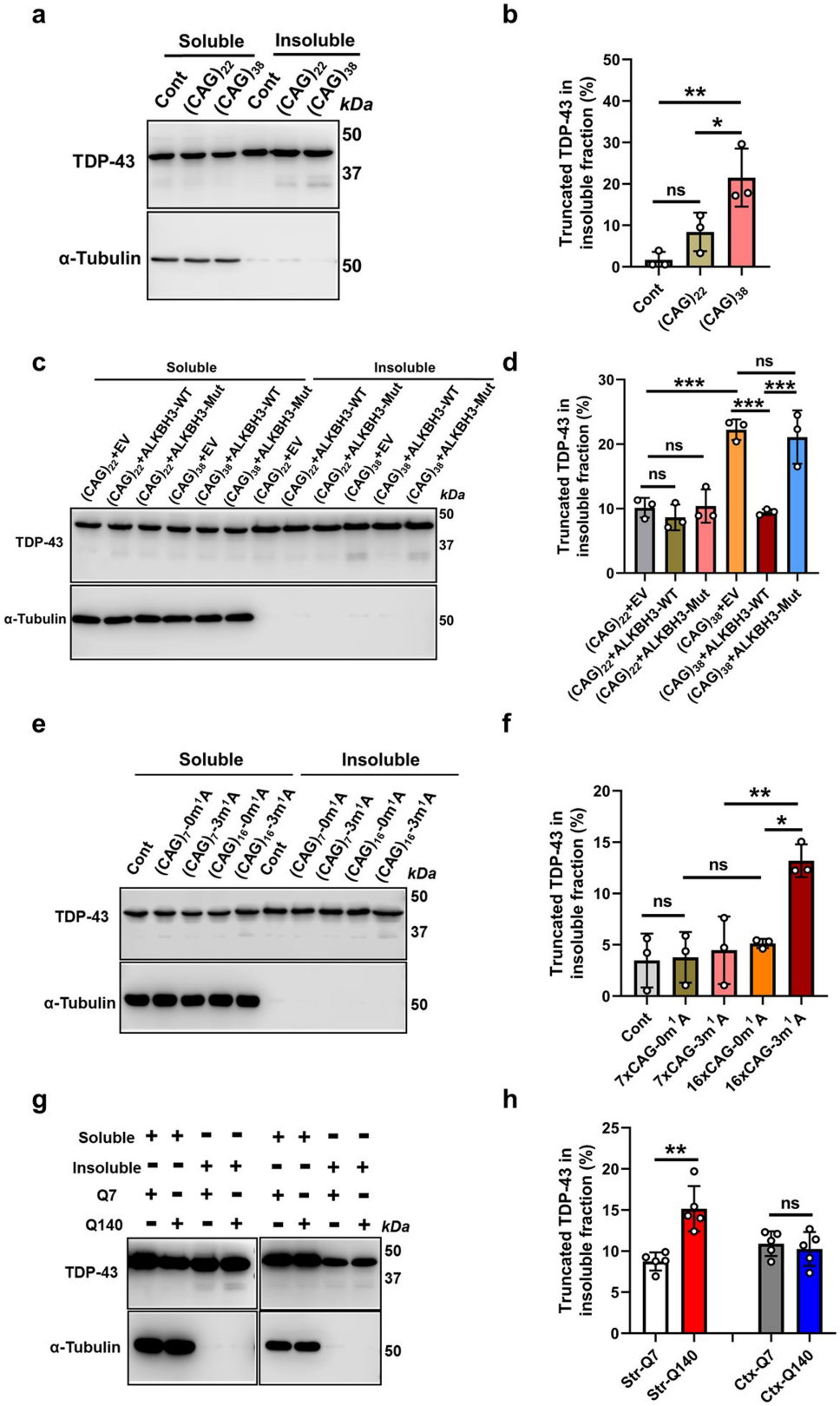
Extended Data Fig. 1 | Expression of CAG repeat RNA suppresses the expression of ALKBH3, but not TRMT61A in cells and striatum tissues of mice. a-**c**, Western blot images and quantification results showing the relative expression levels of TRMT61A in HEK293T cells with ectopic expression of $(CAG)_{22}$ and $(CAG)_{38}$ RNA (**a**) and in striatum (Str) and cortex (Ctx) tissues of Q7 and Q140 mice (**b**, **c**). n = 5 biologically independent samples. Data are mean ± s.d. *P* values were determined using Two-tailed Student's *t*-test. **d**-**e**, Western blot images and quantification results showing the relative expression levels of ALKBH3 in HEK293T cells with ectopic expression of $(CAG)_{22}$ and $(CAG)_{38}$ RNA. **f**, RT-qPCR results showing the relative expression levels of ALKBH3 mRNA in HEK293T cells with ectopic expression of $(CAG)_{22}$ and $(CAG)_{38}$ RNA, or with empty plasmid control (Contr). Data in **e**-**f** are mean ± s.d., and represent three biologically independent experiments. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. ns, $P > 0.05$; $^{**}P = 0.0041$ in **e**; $^{***}P = 0.0005$ between Contr and $(CAG)_{22}$, and $^{***}P = 0.0001$ between $(CAG)_{22}$ and $(CAG)_{38}$ in **f**; $^{****}P < 0.0001$. **g**-**j**, Western blot images and quantification results showing the relative expression levels of ALKBH3 protein in striatum (**g**-**h**) and cortex (**i**-**j**) tissues of Q7 and Q140 mice (n = 5 biologically independent samples of striatum or cortex). The loading control was detected in a separate gel in parallel for **d** and **g**. Data are mean ± s.d. *P* values of Str or Ctx were determined using Two-tailed Student's *t*-test. ns, $P > 0.05$; $^{***}P = 0.0041$ in **h**. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 2 | ALKBH3-WT, but not ALKBH3-Mut extends the lifespan of *Drosophila* expressing expanded Q78 CAG repeats, and reduces the ratios of m^1A/rA, but not m^6A/rA in CAG repeat RNA. a, The mRNA level of wild-type human ALKBH3 (ALKBH3-WT) or its catalytically inactive mutant (ALKBH3-Mut) in Q78 *Drosophila* was measured by RT-qPCR (n = 6 biologically independent samples, 15 heads per sample). Data are mean ± s.d. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. ns, $P > 0.05$; $^{**}P = 0.0016$; $^{****}P < 0.0001$; **b**, *Drosophila* lifespan was evaluated with co-expression of wild-type human ALKBH3 or its catalytically inactive mutant in Q78 *Drosophila* (n = 150 animals per survival curve). **c**-**d**, The levels of m^1A (**c**) and m^6A (**d**) in Q78 flies with expression of wild-type human or its catalytically inactive mutant. Data are mean ± s.d., and represent three biologically independent samples (30 heads per sample). Ten-day old male flies on RU486 food were used for the experiments. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. ns, $P > 0.05$; $^{*}P = 0.012$ between Q78 and Q78 + ALKBH3-WT, and $^{*}P = 0.027$ between Q78 + ALKBH3-WT and Q78 + ALKBH3-Mut in **c**.

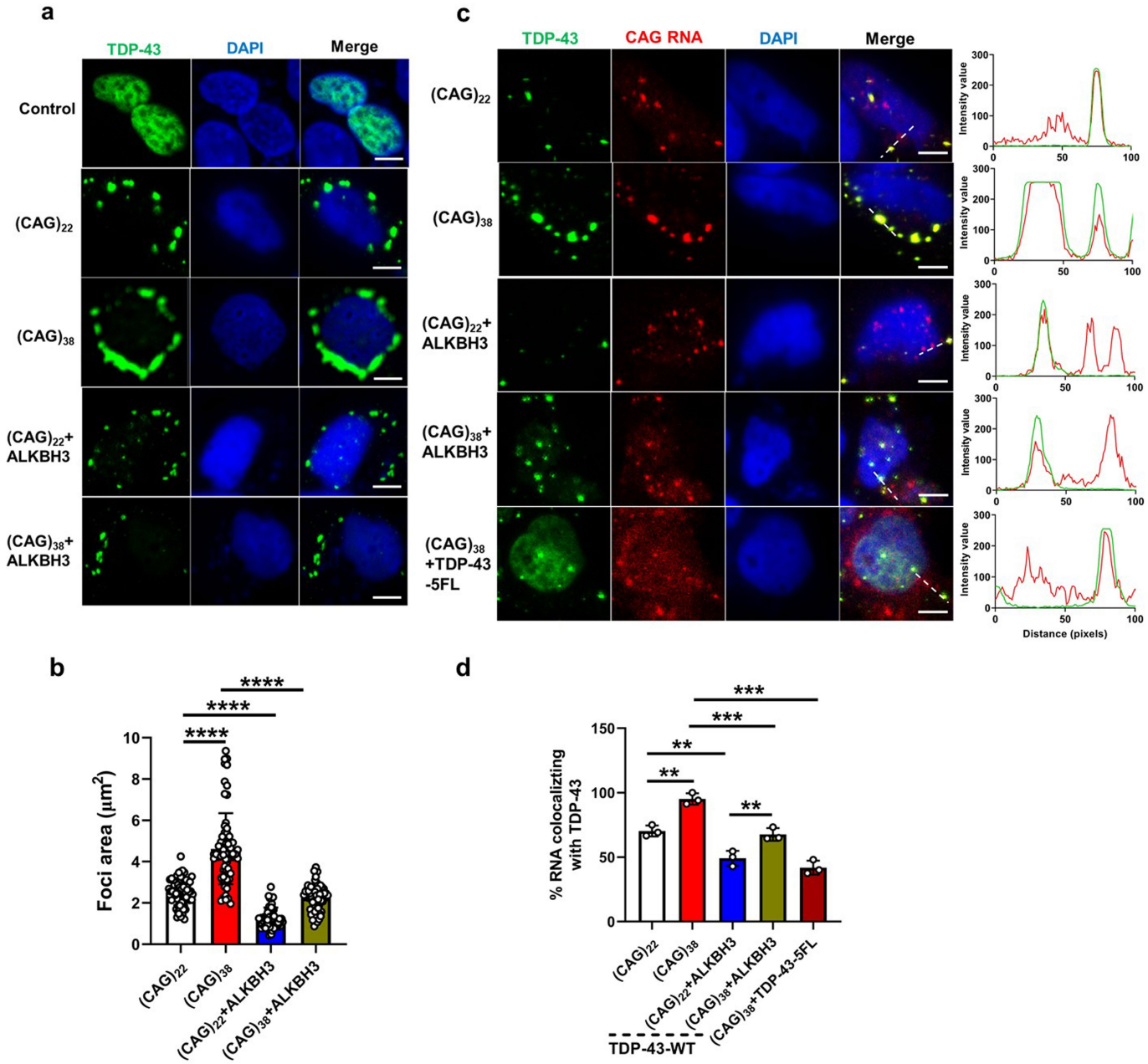


Extended Data Fig. 3 | Electrophoretic mobility shift assay (EMSA) for determining the binding affinities of TDP-43 with rA-, m^{1}A- and m^{6}A-carrying CAG repeat RNA. a, Construct designs for TDP-43 vectors containing functional (WT) or RNA-binding-defective (5FL) RRMs were used for EMSA. **b**-**i**, EMSA for measuring the binding affinities of full-length TDP-43 (**b**-**e**) and TDP-43-RRM (**f**-**i**) with rA-, m^{1}A- and m^{6}A-carrying RNA probes (5′- CCGUUC CGCCCXGGCCGCGCCCAGCUGGAAUGCA-3′, where 'X' is rA, m^{1}A or m^{6}A) (**b**-**c**, **f**-**g**) or with rA-, and m^{1}A-carrying CAG repeat RNA probes (5′-CAGCAGCAGC XGCAGCAGCAG-3′, where 'X' is rA or m^{1}A) (**d**-**e**, **h**-**i**). Protein concentrations ranged from 0 to 1000 nM. Data are mean ± s.e.m., and represent three biologically independent experiments. Shown in (**j**) and (**k**) are the gel image and quantification results for the K_d values for wild-type TDP-43 and its RRM mutant (TDP-43-5FL) in binding with $(CAG)_7$-1 m^{1}A. For gel source data, see Supplementary Fig. 1.

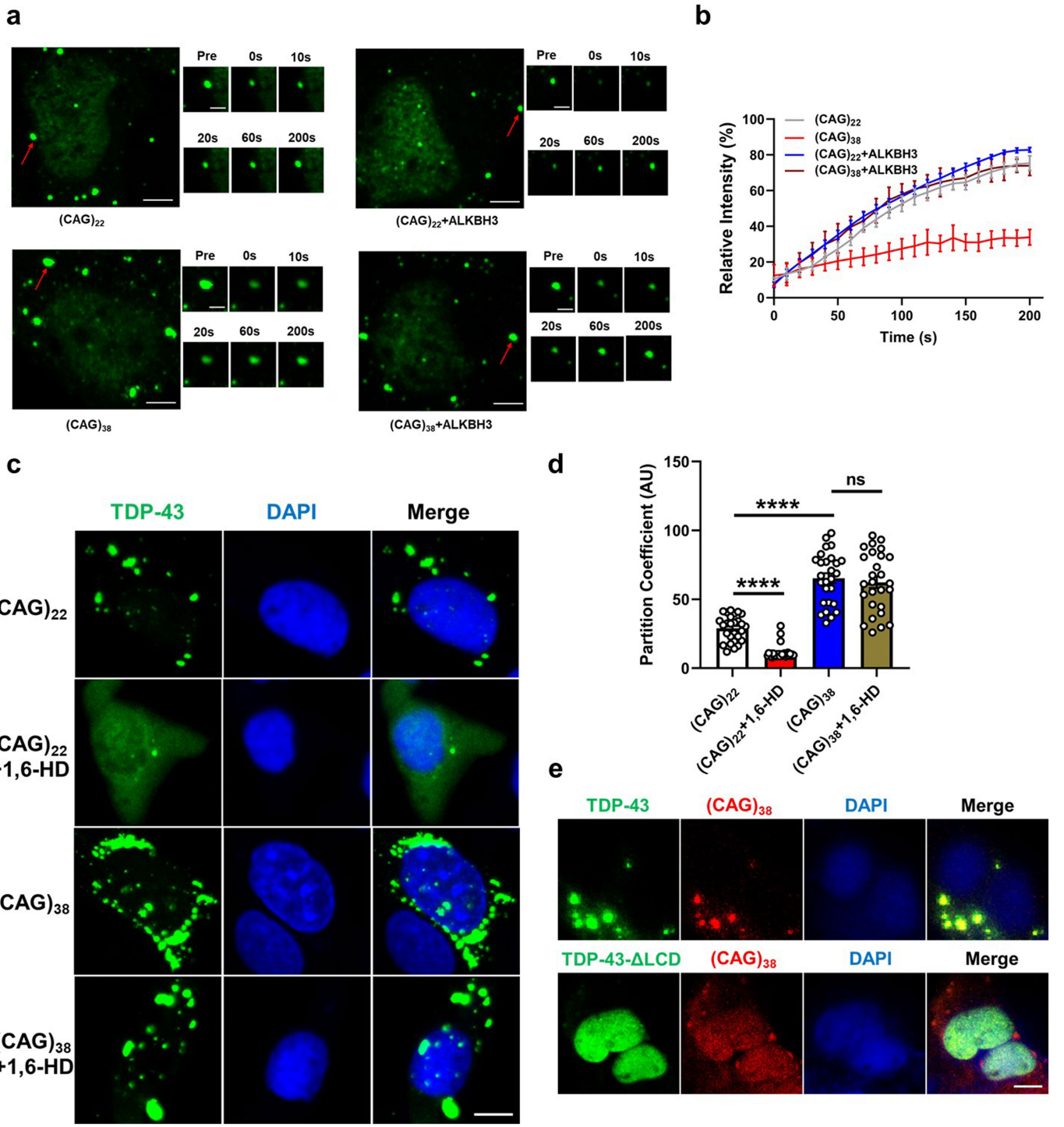


Extended Data Fig. 4 | See next page for caption.

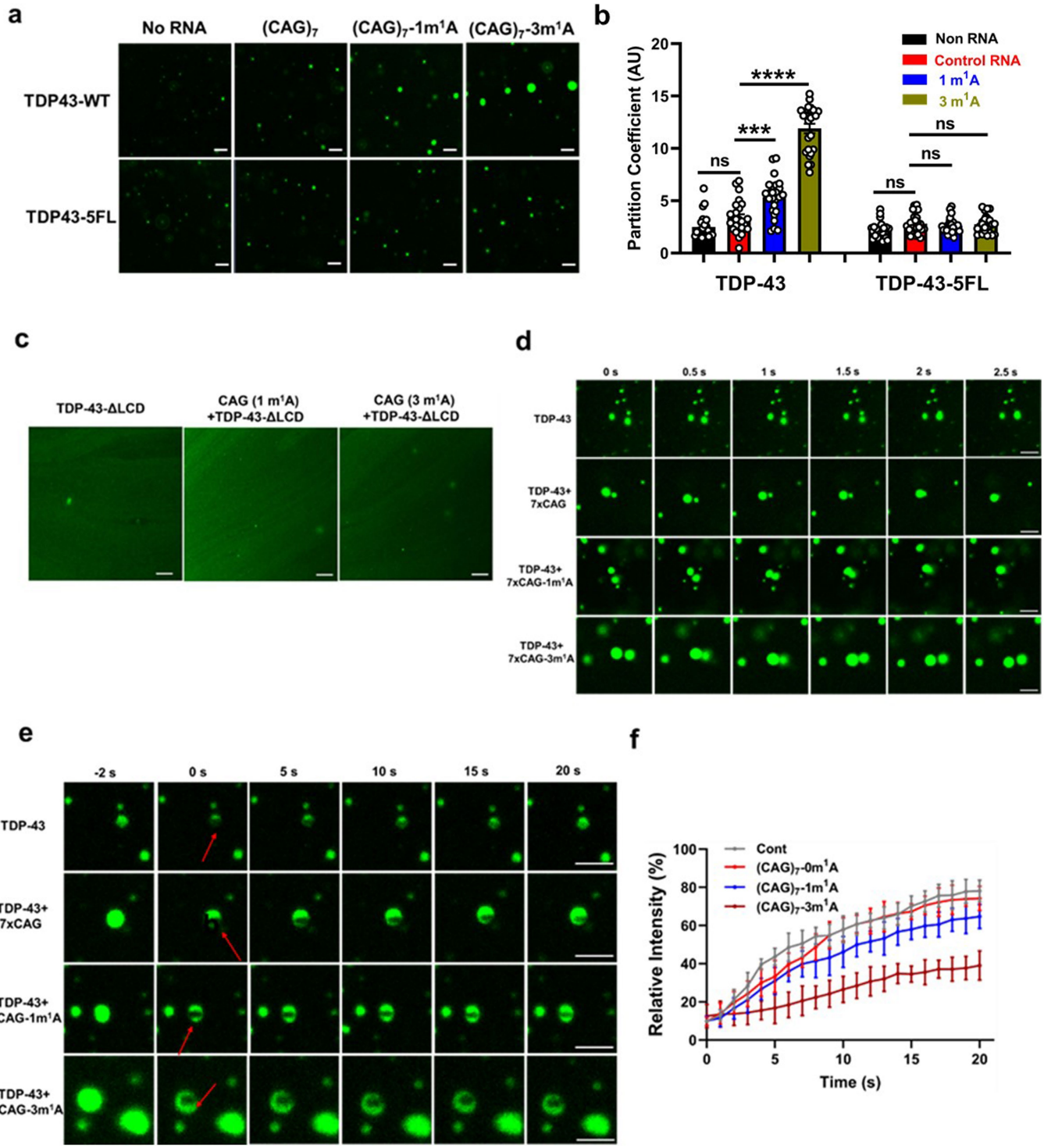
Extended Data Fig. 4 | m^1A in CAG repeat RNA induces truncation of TDP-43 protein. a-f, Detergent-solubility fractionation and Western blot for endogenous TDP-43 protein in cells with ectopic expression of CAG repeat RNA (**a**, **b**), with or without ectopic expression of ALKBH3-WT or ALKBH3-Mut (**c**, **d**), or with or without transfection of synthetic $(CAG)_7$ or $(CAG)_{16}$ RNAs containing zero or three m^1A (**e**, **f**). Data are mean ± s.d. and represent three biologically independent experiments. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. ns, $P > 0.05$; $^{**}P = 0.0066$ and $^{*}P = 0.041$ in **b**; $^{***}P = 0.0004$ between $(CAG)_{22}$ + EV and $(CAG)_{38}$ + EV, $^{***}P = 0.0003$ between $(CAG)_{38}$ + EV and $(CAG)_{38}$ + ALKBH3-WT, and $^{***}P = 0.0006$ between $(CAG)_{38}$ + ALKBH3-WT and $(CAG)_{38}$ + ALKBH3-Mut in **d**; $^{*}P = 0.011$, and $^{**}P = 0.0065$ in **f**; **g**-**h**, Quantification results showing the fractions of truncated TDP-43 protein in striatum (n = 5 biologically independent samples) and cortex (n = 5 biologically independent samples) tissues of mice. Data are mean ± s.d. *P* values were determined using Two-tailed Student's *t*-test. ns, $P > 0.05$; $^{**}P = 0.0013$. The loading control was detected in a separate gel in parallel for **a**, **c**, **e** and **g**. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 5 | m^{1}A triggers cytoplasmic mis-localization and aggregation of ectopically expressed EGFP-TDP-43. a, Representative images of U2OS cells co-expressing EGFP-TDP-43 and CAG repeat RNAs with or without ALKBH3 overexpression. **b**, Quantification results for the areas of EGFP-TDP-43 foci. $(CAG)_{22}$: n = 69; $(CAG)_{38}$: n = 69; $(CAG)_{22}$ + ALKBH3: n = 68; $(CAG)_{38}$ + ALKBH3: n = 66. Data are mean ± s.d., and represent three biologically independent experiments. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. $^{****}P < 0.0001$. **c**, RNA FISH for assessing the co-localization of EGFP-TDP-43 with CAG repeat RNAs. **d**, Percentages of CAG repeat RNAs co-localized with EGFP-TDP-43. Data are mean ± s.d. and represent three biologically independent experiments. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. $^{**}P = 0.0012$ between $(CAG)_{22}$ and $(CAG)_{38}$, $^{**}P = 0.0033$ between $(CAG)_{22}$ and $(CAG)_{22}$ + ALKBH3, $^{**}P = 0.0074$ between $(CAG)_{22}$ + ALKBH3 and $(CAG)_{38}$ + ALKBH3; $^{***}P = 0.0006$ between $(CAG)_{38}$ and $(CAG)_{38}$ + ALKBH3, and $^{***}P = 0.0002$ between $(CAG)_{38}$ and $(CAG)_{38}$ + TDP-43-5FL. Scale bar, 10 μm.



Extended Data Fig. 6 | $(CAG)_{38}$ RNA triggers phase transition of TDP-43 protein in cells. a, FRAP of EGFP-TDP-43 droplets in U2OS cells at 24 h following transfection. Scale bar, 5 µm. Higher magnification is shown in the right panel at different time points following photobleaching at 0 s (scale bar, 2 µm). **b**, Mean fluorescence intensity of EGFP-TDP-43 in the fully bleached area over time. Data are mean ± s.e.m. from the recovery curves obtained from three biologically independent experiments. **c**, Representative images of U2OS cells expressing $(CAG)_{22}$ RNA with or without treatment with 6% 1,6-hexanediol (1,6-HD) for 1 min. Scale bar, 10 µm. **d**, Partition coefficients of TDP-43 protein in $(CAG)_{22}$- or $(CAG)_{38}$-expressing cells with or without treatment with 6% 1,6-HD. The partition coefficient was calculated based on the fraction of EGFP-TDP-43 signal in the droplets over EGFP-TDP-43 signal in immediately adjacent region. n = 27 foci of EGFP-TDP-43 for each group with three biologically independent experiments. Data are mean ± s.e.m. *P* values were determined using Two-tailed Student's *t*-test. ns, $P > 0.05$; ****$P < 0.0001$. **e**, Phase separation of EGFP-TDP-43 and EGFP-TDP-43-ΔLCD proteins in U2OS cells with or without ectopic expression of $(CAG)_{38}$ mRNA. The experiments were repeated independently three times with similar results. Scale bar, 10 µm.



Extended Data Fig. 7 | m^{1}A modulates liquid-liquid phase separation of TDP-43 protein. a, The droplets of EGFP-TDP-43 and EGFP-TDP-43-5FL were detected in the presence of synthetic $(CAG)_7$ RNAs (50 ng) containing 0, 1 and 3 m^{1}A residues, or without RNA. **b**, Partition coefficients of wild-type and mutant TDP-43 proteins in the presence or absence of m^{1}A-containing CAG repeat RNA. 24 or 25 TDP-43-EGFP droplets observed in three biologically independent experiments for each group were quantified. n = 24 for TDP-43-WT + No RNA and TDP-43-WT + 3m^{1}A; n = 25 for TDP-43-WT + Control RNA, TDP-43-WT + 1m^{1}A,TDP-43-5FL + No RNA, TDP-43-5FL + Control RNA, TDP-43-5FL + 1m^{1}A, and TDP-43-5FL + 3m^{1}A. Data are mean ± s.e.m. *P* values were determined using one-way ANOVA with Tukey's multiple comparisons test. ns, $P > 0.05$; ***$P = 0.0009$; ****$P < 0.0001$. **c**, The droplets of TDP-43-ΔLCD were detected with CAG repeats RNAs (50 ng) or without RNA. Scale bar, 5 μm. **d**, TDP-43-EGFP (50 μM) was imaged using confocal microscopy. The fusions of TDP-43-EGFP droplets are also shown in Supplementary Videos 1–4. Scale bar, 5 μm. The experiments shown in **c** and **d** were repeated at least three times with similar results. **e**, Representative examples of FRAP analyses of in vitro phase-separated TDP-43 droplets in the presence of synthetic $(CAG)_7$ RNAs containing 0, 1 and 3 m^{1}A residues, or without RNA. FRAP are also shown in Supplementary Videos 5–8. **f**, FRAP curves displaying the relative TDP-43 intensity of phase-separated droplets in the presence of $(CAG)_7$ RNAs containing 0, 1 and 3 m^{1}A residues, or without RNA. The data represented mean ± s.e.m. from the recovery curves obtained from three biologically independent experiments. Scale bar, 2 μm.

nature portfolio

Corresponding author(s): Yinsheng Wang

Last updated by author(s): Sep 16, 2023

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 software was used in this study. For Western blot, membranes were scanned using Odyssey Fc Imager (LI-COR). The immunofluoresence data were imaged with an Zeiss 880 Inverted microscope (Zen 2 Blue, Version 2.3).
Data analysis	For statistical analysis, GraphPad Prism version 8.0.1 was used. For image analysis, Image J Fiji was used. For data classification and calculation, Microsoft Excel 2016 was used.

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

Necessary data for evaluating this study are available in the main text and Supplementary Information. Uncropped gel images are provided in Supplementary Fig. 1. Source data are provided with this paper.

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	No participants in this study
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

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 size was chosen to ensure reproducibility of the results. Sample sizes are noted for all experiments. At least 3 independent experiments were carried out for all of the assays and noted in the relevant figure legends.
Data exclusions	No data were excluded.
Replication	All experiments were performed in at least 3 biological replicates, which is indicated in the Figure Legends.
Randomization	Random images of cells were evaluated for immunofluorescence (IF). Randomization for other experiments were not subjective.
Blinding	Blinding is not relevant to the study as the output parameters are not subjective and therefore not subjected to this form of bias.

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	TDP-43 (Proteintech, 10782-2-AP, immunoblot, 1:1000; IF, 1:100), G3BP1 (Proteintech, 66486-1-lg, immunoblot, 1:1000; IF, 1:100), TRMT61A (Thermo Fisher, Cat #A305-858A-T, immunoblot, 1:1000), anti-Flag antibody (Cell Signaling Technology, 14793S, immunoblot, 1:2000), ALKBH3 (Cell Signaling Technology, #87620, immunoblot, 1:1000), anti-GAPDH antibody (Santa Cruz

Biotechnology, sc-32233, immunoblot, 1:5000), anti-α-Tubulin antibody (Santa Cruz Biotechnology, sc-32293, immunoblot, 1:5000), Anti-Rabbit IgG (whole molecule)-Peroxidase antibody produced in goat (Sigma, A0545, immunoblot, 1:2500), m-IgGκ BP-HRP (Santa Cruz Biotechnology, sc-516102, immunoblot, 1:2500), Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488 (Invitrogen, A-32731, IF, 1:250), Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 647 (Invitrogen, A-32728, IF, 1:250)

Validation

TDP-43 antibody is a rabbit polyclonal antibody recognizing N-terminal domain of TDP-43. Tested Reactivity Human, Mouse, Rat, Zebrafish. Host/Isotype: Rabbit/IgG. Positive WB detected in SH-SY5Y cells, HeLa cells, C2C12 cells, Neuro-2a cells. This antibody has been used in other publications (e.g., Nature, 2021, 594:117-123; Nature, 2022, 603:124-130) according to the Proteintech website.

G3BP1 antibody targets G3BP1 in WB, RIP, IP, IHC, IF, FC, CoIP, ELISA applications and shows reactivity with Human, mouse, rat, pig samples. Host/Isotype: Mouse/IgG.Positive WB detected in LNCaP cells, RAW 264.7 cells, pig brain tissue, HeLa cells, mouse brain tissue, HEK-293 cells, HepG2 cells, Jurkat cells, K-562 cells, HSC-T6 cells, PC-12 cells, NIH/3T3 cells, 4T1 cells. This antibody has been used in other publications (e.g., Nature, 2019, 573:590-594; Nature, 2020, 588:688-692) according to the Proteintech website.

TRMT61A antibody targets TRMT61A in WB and IHC applications and shows reactivity with Human, mouse, rat. Host/Isotype: Rabbit/IgG. Positive WB detected in Jurkat, BT-474, U-251MG, mouse brain, mouse pancreas, mouse spleen, mouse liver, rat testis and rat brain. The specificity of this antibody was verified by shRNA knockdown experiments conducted in this study.

Flag (DYKDDDDK) Tag antibody (D6W5B) Rabbit mAb detects exogenously expressed DYKDDDDK proteins in cells. The antibody recognizes the DYKDDDDK peptide, which is the same epitope recognized by Sigma-Aldrich Anti-FLAG M2 antibody, fused to either the amino-terminus or carboxy-terminus of the target protein. This antibody was applied in WB, IP, IHC, ChIP, eCLIP, IF.This antibody has been used in other publications (e.g., Nature, 2023, 619:819-827; Nat. Commun., 2023, 14:4217) according to the Cell Signaling Technology website.

ALKBH3 antibody targets ALKBH3 in WB, IP, ChIP, IF and IHC applications and shows reactivity with Human, mouse, rat. Host/Isotype: Rabbit/IgG. Positive WB detected in A549 and PC-3 cells. This antibody has been used in other publication (e.g., Cell Death Discov., 2023, 9:159) according to the Cell Signaling Technology website.

GAPDH antibody is recommended for detection of GAPDH of mouse, rat, human, rabbit and Xenopus origin by WB, IP and IF. Host: Mouse. This antibody has been used in other publications (e.g., Nat Commun., 2023, 14:159.; Nucleic Acids Res., 2022, 50:6990-7001) according to the Santa Cruz website.

Anti-alpha Tubulin Antibody (DM1A) is recommended for detection of α Tubulin of mouse, rat, human and avian origin by WB, IP, IF and IHC. Host: Mouse. This antibody has been used in other publications (e.g., Cell, 2023, 186:112-130.; Cell Death Dis., 2023, 14:220)

Goat anti-rabbit IgG-peroxidase conjugate associates with all rabbit Igs. It has been used in western blotting , immunofluorescence staining , immunochemistry and immunoprecipitations. This antibody has been used in other publication (e.g., Journal of Biological Chemistry, 279:47:P49384-49394)

Mouse IgGκ light chain binding protein (m-IgGκ BP) conjugated to horseradish peroxidase (HRP) is a strongly recommended alternative to conventional goat/rabbit anti-mouse IgG secondary antibodies for WB and IHC signal enhancement. This antibody has been used in other publication (e.g., Nat Commun., 2023, 14:1713)

Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody Alexa Fluor® 488 conjugate was performed to IF analysis. This antibody has been used in other publication (e.g., Development, 2023, 150: dev201376)

Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 647 conjugate was performed to IF analysis. This antibody has been used in other publication (e.g., Nat Commun., 2023, 14:5521)

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK293T (ATCC, Catalog # CRL-3216) and U2OS (ATCC, Catalog # HTB-96)
Authentication	HEK293T cells and U2OS cells were authenticated by ATCC. Cells were authenticated using using Short Tandem Repeat (STR) analysis as described in 2012 in ANSI Standard (ASN-0002) Authentication of Human Cell Lines.
Mycoplasma contamination	Not contaminated. Tested with LookOut® Mycoplasma PCR Detection Kit (Sigma).
Commonly misidentified lines (See ICLAC register)	None was used in the present study.

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	Mice were maintained and bred under standard conditions consistent with National Institutes of Health guidelines. The cages were

Laboratory animals	maintained on a 12:12 light/dark cycle, with food and water ad libitum. Brain tissues from wild-type and heterozygous HD knockin Q140 mice in C57BL/6 background at 6-month age were harvested and immediately frozen on dry ice.
Wild animals	No wild animals were used in this study.
Reporting on sex	Brain tissues from both male and female mice were used in this study.
Field-collected samples	No field collected samples were used in the study.
Ethics oversight	The mouse experiments were approved by the University of California, Los Angeles Institutional Animal Care and Use Committee.

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