PNAS RESEARCH ARTICLE | BIOCHEMISTRY

Zfp106 binds to G-quadruplex RNAs and inhibits RAN translation and formation of RNA foci caused by G4C2 repeats

Barbara Celona[a], Sally E. Salomonsson[b,c], Haifan Wu[a,d,1], Bobo Dang[a,d,2,3], Huong T. Kratochvil[a,d,4], Claire D. Clelland[b,c], William F. DeGrado[a,d,5], and Brian L. Black[a,e,5]

Affiliations are included on p. 8.

Contributed by William F. DeGrado; received December 16, 2022; accepted June 14, 2024; reviewed by Todd J. Cohen, Vincenzo A. Gennarino, and Oyoshi Takanori

Expansion of intronic GGGGCC repeats in the *C9orf72* gene causes amyotrophic lateral sclerosis (ALS) and frontotemporal dementia. Transcription of the expanded repeats results in the formation of RNA-containing nuclear foci and altered RNA metabolism. In addition, repeat-associated non-AUG (RAN) translation of the expanded GGGGCC-repeat sequence results in the production of highly toxic dipeptide-repeat (DPR) proteins. GGGGCC repeat–containing transcripts form G-quadruplexes, which are associated with formation of RNA foci and RAN translation. Zfp106, an RNA-binding protein essential for motor neuron survival in mice, suppresses neurotoxicity in a *Drosophila* model of *C9orf72* ALS. Here, we show that Zfp106 inhibits formation of RNA foci and significantly reduces RAN translation caused by GGGGCC repeats in cultured mammalian cells, and we demonstrate that Zfp106 coexpression reduces the levels of DPRs in *C9orf72* patient-derived cells. Further, we show that Zfp106 binds to RNA G-quadruplexes and causes a conformational change in the G-quadruplex structure formed by GGGGCC repeats. Together, these data demonstrate that Zfp106 suppresses the formation of RNA foci and DPRs caused by GGGGCC repeats and suggest that the G-quadruplex RNA-binding function of Zfp106 contributes to its suppression of GGGGCC repeat-mediated cytotoxicity.

C9orf72 | amyotrophic lateral sclerosis | RAN translation | G-quadruplex | RNA-binding protein

Significance

The most common inherited form of ALS and FTD is caused by expansion of a six-nucleotide GGGGCC repeat in the first intron of the *C9orf72* gene. GGGGCC repeat–containing RNAs form focal aggregates and are translated via repeat-associated non-AUG (RAN) translation into highly toxic dipeptide-repeat (DPR) proteins. We show that the zinc finger–containing RNA-binding protein Zfp106 inhibits formation of RNA foci, significantly reduces RAN translation caused by GGGGCC repeats, and suppresses the formation of DPRs in *C9orf72* patient-derived cells. We also show that Zfp106 binds to RNA G-quadruplexes and causes a conformational change in the G-quadruplex secondary structure formed by GGGGCC repeats. Taken together, these observations suggest that the G-quadruplex RNA-binding function of Zfp106 contributes to its suppression of GGGGCC repeat-mediated cytotoxicity.

Zinc finger protein 106 (Zfp106) is a Cys2-His2 (C2H2) zinc finger protein with four predicted zinc fingers and seven WD40 domains (1–3). Zfp106 primarily localizes to the nucleolus and to nuclear speckles adjacent to spliceosomes (1–3), and previous studies have demonstrated that it functions as an RNA-binding protein with suggested roles in multiple aspects of RNA metabolism, including pre-rRNA processing, splicing, and polyA mRNA binding (1, 2, 4–6). Knockout of the *Zfp106* gene in mice results in profound neuromuscular disease, resembling amyotrophic lateral sclerosis (ALS), with onset at approximately 4 to 6 wk of age followed by rapidly progressing motor neuron degeneration and death by 16 wk (1, 2, 7). Restoration of Zfp106 expression specifically to motor neurons significantly suppresses the phenotype in knockout mice, suggesting that the requirement of Zfp106 for neuromuscular function and motor neuron viability is autonomous to motor neurons (2).

The most common cause of familial ALS and frontotemporal dementia (FTD), a related neurodegenerative disorder of the brain, results from abnormal expansion of the hexanucleotide sequence GGGGCC in the first intron of the *C9orf72* gene (8, 9). Healthy individuals usually have fewer than 30 copies of the hexanucleotide repeat sequence, whereas affected patients may have hundreds to thousands of repeats in their *C9orf72* gene (10). The expanded GGGGCC repeats have been proposed to cause neurodegeneration through two alternative gain-of-function mechanisms. One proposed mechanism involves RNA toxicity through formation of pathogenic repeat RNA foci and sequestration of RNA-binding proteins (11–15). The other proposed mechanism for neurotoxicity of the repeats is via generation of dipeptide repeat proteins (DPRs) produced by aberrant repeat-associated non-ATG (RAN) translation (16–18). Loss-of-function of the *C9orf72* gene, which encodes a guanine nucleotide exchange factor implicated in autophagy and immune responses, exacerbates gain-of-function toxicity (19–24). Work in model organisms has strongly supported the notion that DPRs play a dominant role in the pathogenesis of *C9orf72* disease, but other studies suggest that pathogenesis may be the result of multiple nonmutually exclusive mechanisms (25–27). Indeed, both gain-of-function mechanisms, RNA toxicity and DPR production, may independently cause disruption of pathways such as stress granule formation and nucleocytoplasmic transport (27–32). Interestingly, previous work has

The authors declare no competing interest.

[1]Present address: Department of Chemistry, Wichita State University, Wichita, KS 67260.

[2]Present address: School of Life Sciences, Westlake University, Hangzhou 310030, China.

[3]Present address: Westlake Institute for Advanced Study, Hangzhou 310030, China.

[4]Present address: Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599.

[5]To whom correspondence may be addressed. Email: william.degrado@ucsf.edu or brian.black@ucsf.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2220020121/-/DCSupplemental.

Published July 23, 2024.

established that Zfp106 binds directly and specifically to GGGGCC RNA repeats, and that coexpression of Zfp106 suppresses the pathology and neurotoxicity induced by expression of 30 copies of GGGGCC in a Drosophila model of *C9orf72* ALS (2).

GGGGCC sense strand RNA repeats are capable of forming G-quadruplex or RNA hairpin structures (12, 33, 34), while antisense GGCCCC repeats form i-motifs and protonated hairpins (35). Stem-loops or hairpins are the most common RNA secondary structures and play important roles in transcription, RNA processing, mRNA export, mRNA stability, and translation (36). G-quadruplexes are secondary structures that form specifically in guanine (G)-rich nucleic acids, where a tetrad of G-rich strands is coordinated by a cation (37). These secondary structures in RNA have been proposed to play multiple diverse and important roles in RNA biology, including roles in splicing, polyadenylation, transcriptional termination, translational enhancement or repression, and mRNA trafficking (37). In the context of *C9orf72* ALS/FTD, GGGGCC repeat-induced RNA foci form by phase separation, which has been suggested to be dependent on the G-quadruplex structure (30, 38, 39). Interestingly, small molecules that bind to hairpin or G-quadruplex structures formed by GGGGCC repeats reduce both DPR translation and formation of foci, demonstrating that targeting secondary structures formed by GGGGCC repeats may be therapeutically relevant (33, 40, 41).

Here, we show that Zfp106 binds to G-quadruplex-forming RNA sequences, including to GGGGCC repeats when folded in the G-quadruplex conformation. Moreover, we find that Zfp106 binding alters the G-quadruplex structure of GGGGCC repeats, and we show that Zfp106 specifically inhibits formation of GGGGCC-mediated RNA foci. We also show that Zfp106 inhibits GGGGCC-induced RAN translation in mammalian cells and reduces the levels of DPRs in *C9orf72* patient-derived cells. These results support a role for Zfp106 in suppressing both of the gain-of-function mechanisms associated with *C9orf72* ALS/FTD and suggest that Zfp106 recognition and binding to G-quadruplexes may play a role in the pathobiology of neurodegeneration, including *C9orf72* ALS/FTD.

Results and Discussion

Zfp106 Suppresses GGGGCC-Induced RAN Translation. To gain insight into how Zfp106 suppresses the cytopathic effects induced by expression of GGGGCC repeats, we cotransfected a Zfp106 expression plasmid with a pCMV-$(GGGGCC)_{30}$-EGFP expression plasmid or a control pCMV plasmid with only 3 copies of the GGGGCC repeat sequence into Neuro-2A cells and measured eGFP RNA and protein expression in the presence or absence of Zfp106 expression (Fig. 1*A*). Zfp106 significantly inhibited protein expression from the eGFP mRNA containing 30 copies of the GGGGCC repeats without affecting the steady-state level of the mRNA (Fig. 1 *A′* and *A″*). Zfp106 did not affect eGFP protein expression from plasmids containing 6× or 60× Huntington's disease-associated CAG repeats (42, 43) in the 5′-UTR of the eGFP cDNA (Fig. 1*B*). These data suggest that Zfp106 inhibited translation of the eGFP mRNA specifically containing 30× GGGGCC repeats. Interestingly, expression of Zfp106 increased the nuclear retention of the eGFP mRNA that contained 30× GGGGCC repeats (*SI Appendix*, Fig. S1*A*), suggesting that the translational inhibition mediated by coexpression of Zfp106 with 30× GGGGCC repeat-containing mRNAs (Fig. 1*A′*) might be due to alterations in nucleocytoplasmic transport of the repeat-containing RNA.

An important distinction between the expression of the GGGGCC repeats from the pCMV-$(GGGGCC)_{30}$-EGFP reporter compared to the expression of the repeats in the context of *C9orf72* ALS/FTD is that the repeats expressed from the pCMV-$(GGGGCC)_{30}$-EGFP reporter are present in the 5′-UTR, and GFP protein expression occurs by standard AUG-dependent translation (44), whereas translation of the repeats in the disease context in humans occurs via RAN translation (16–18). Therefore, to determine whether Zfp106 coexpression also inhibits RAN translation, we cotransfected the Zfp106 expression plasmid with a bicistronic reporter plasmid that produces NanoLuc luciferase (NLuc) by RAN translation induced by 70× GGGGCC repeats from an intronic sequence derived from the human *C9orf72* gene [Fig. 1*C* and *SI Appendix*, Fig. S1*C*; (45)]. As an internal control, the reporter construct produces firefly luciferase (FLuc) by canonical AUG-dependent initiation (45). Coexpression of Zfp106 significantly and specifically reduced RAN translation of NLuc from the reporter construct without affecting AUG-dependent translation of FLuc from the same plasmid (Fig. 1*C′*). Inhibition occurred in a Zfp106-dose-dependent manner (*SI Appendix*, Fig. S1*C″*). Importantly, steady-state NLuc mRNA levels were unaffected by coexpression of Zfp106 (Fig. 1*C″*). However, nuclear retention of the NLuc mRNA was increased slightly but significantly (*SI Appendix*, Fig. S1*B*), further supporting the idea that Zfp106 expression results in retention of GGGGCC-containing RNAs in the nucleus. Zfp106 also inhibited RAN translation from a second bicistronic reporter plasmid from which NLuc is produced by RAN translation induced by 70× GGGGCC repeats from a mature mRNA, rather than from intronic sequences, also without affecting steady-state RNA level *SI Appendix*, Fig. S1*D*.

To determine whether Zfp106 inhibits RAN translation in a disease-relevant context, we transduced induced pluripotent stem cells (iPSCs) derived from a *C9orf72* patient with lentiviral vectors expressing either GFP-Zfp106 or GFP alone. One week after transduction, GFP-positive cells were sorted, and poly-GA, one of the DPRs produced by RAN translation of sense GGGGCC repeats, was quantified by meso scale discovery (MSD) immunoassay (Fig. 1*D*), using an antibody previously validated through both positive and negative selection (46). Importantly, overexpression of GFP-Zfp106 but not the GFP only control significantly reduced poly-GA expression in *C9orf72* patient iPSCs (Fig. 1*D′*). Since DPRs produced via GGGGCC repeat–induced RAN translation are major contributors to *C9orf72* pathology in animal models and are strongly associated with disease in humans (16–18), these results suggest that Zfp106 inhibition of RAN translation may contribute to the previously reported neuroprotective effect of Zfp106 in vivo (2).

Zfp106 Inhibits the Formation of GGGGCC Repeat–Induced RNA Foci. GGGGCC repeat–containing RNAs also form foci in affected motor neurons, and these foci are strongly associated with pathology and disease (11–15). Moreover, the presence of RNA foci and DPRs are not mutually exclusive, and the two pathology-associated phenomena often co-occur (47). Therefore, we examined the influence of Zfp106 coexpression on the formation of RNA foci induced by overexpression of GGGGCC repeat–containing RNAs (Fig. 2). Jain and Vale described a system in which 29× GGGGCC repeats tagged with 12× MS2-hairpin loops can be visualized via coexpression of YFP-tagged MS2-coat binding protein, which binds to the MS2 hairpin loops (Fig. 2*A*) (38). Consistent with previously published results, induction of 29× GGGGCC RNA repeats in this system resulted in formation of numerous nuclear foci (Fig. 2*B*) (38). Remarkably, the presence of foci was substantially reduced by coexpression of Zfp106 compared to nuclear-RFP control transfected cells (Fig. 2 *C*, *D*, and *F*). In contrast, Zfp106 had no effect on the formation of RNA foci induced by 47× CAG repeats (Fig. 2 *E* and *G*). As an alternate method for detecting foci, we also examined the number of foci per nucleus by using RNA FISH

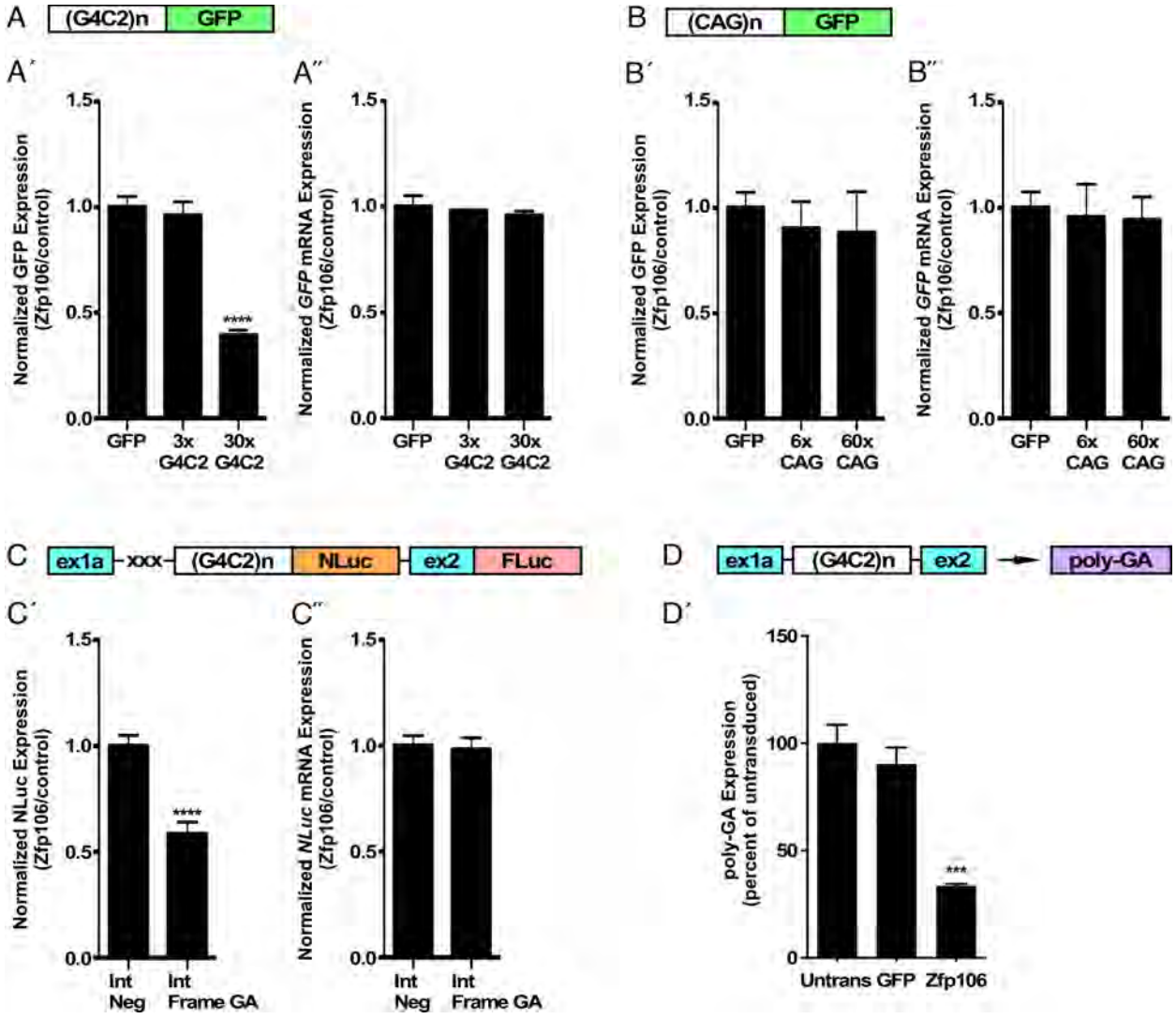

Fig. 1. Zfp106 suppresses RAN translation. eGFP fluorescence (*A′* and *B′*) or NLuc luminescence (*C′*) and mRNA levels (*A″*–C″) in Neuro-2a cells (*A* and *B*) or HEK293T cells (*C*) cotransfected with pCMV-$(GGGGCC)_n$-EGFP constructs (no-repeats, 3×, or 30× GGGGCC repeats) and Zfp106 or control vector (*A*); cotransfected with pCMV-(CAG)n-EGFP–containing constructs (no-repeats, 6×, or 60× CAG repeats) and Zfp106 or control vector (*B*); or cotransfected with bicistronic splicing dual-luciferase reporters with *C9orf72* exons and intron (Int Neg, no repeats; Int Frame GA, 70× GGGGCC repeats) and Zfp106 or control vector (*C*). (*A* and *B*) GFP fluorescence was normalized to total protein content; *eGFP* mRNA was normalized to *Gapdh*. (*C*) NLuc luminescence (*C′*) or mRNA expression (*C″*) was normalized to FLuc in each sample. Data are expressed as ratios of Zfp106-transfected cells over control. The mean ratio from the no-repeat construct was set to a normalized value of 1. (*D′*) Poly-GA levels quantified by MSD immunoassay in *C9orf72* patient iPSCs transduced with GFP or GFP-Zfp106 lentivirus (from plasmids pLV-EF1a-GFP and pLV-EF1a-GFP-Zfp106). Data are expressed as the percentage of the untransduced control that was set to 100%. Schematic depictions of transfected constructs are shown in (*A*–*C*); xxx indicates stop codons. Schematic depiction of the endogenous *C9orf72* repeat expansion is shown in *D*. N = 3 biological replicates. Data were analyzed by one-way ANOVA and Bonferroni's multiple comparison test; ***$P < 0.001$; ****$P < 0.0001$.

probes directed against MS2 hairpin loops, which also showed that Zfp106 expression substantially abrogated formation of foci when compared to the number of foci in cells lacking Zfp106 expression (*SI Appendix*, Fig. S2*A*).

Taken together, the results in Figs. 1 and 2 demonstrate that Zfp106 inhibits RAN translation and the formation of RNA foci caused by GGGGCC repeats. These observations have important implications for understanding the role of Zfp106 in neuroprotection and neurodegeneration in gain- and loss-of-function models of ALS.

Zfp106 Binds to G-Quadruplex-Forming RNA Sequences. The observations that Zfp106 suppressed RAN translation and formation of RNA foci caused by GGGGCC repeats, combined with our previously published work showing that Zfp106 binds directly and specifically to GGGGCC repeats (2), suggested that RNA-binding by Zfp106 might be involved in its repressive effect on repeat-induced pathology. Our previous studies also showed that Zfp106 interacts directly with a wide array of other RNA-binding proteins, which further suggests that it plays roles in cellular RNA processing and that its RNA-binding activity is unlikely to be restricted solely to GGGGCC repeats (2). Because GGGGCC repeats are capable of forming G-quadruplex structures and because RNA G-quadruplexes are associated with phase separation and formation of RNA foci (34, 38), we reasoned that Zfp106 might bind more generally to G-quadruplex forming RNA sequences and that this ability might confer its repressive effects on GGGGCC-repeat toxicity.

We use electrophoretic mobility shift assay (EMSA) to test the ability of full-length Zfp106 (Fig. 3*A*) to bind to numerous RNA repeat sequences, including those predicted to form G-quadruplexes and those predicted to form hairpins or other non-G-quadruplex secondary structures (Fig. 3*B*). Zfp106 bound to all sequences predicted to form G-quadruplexes, while it failed to bind

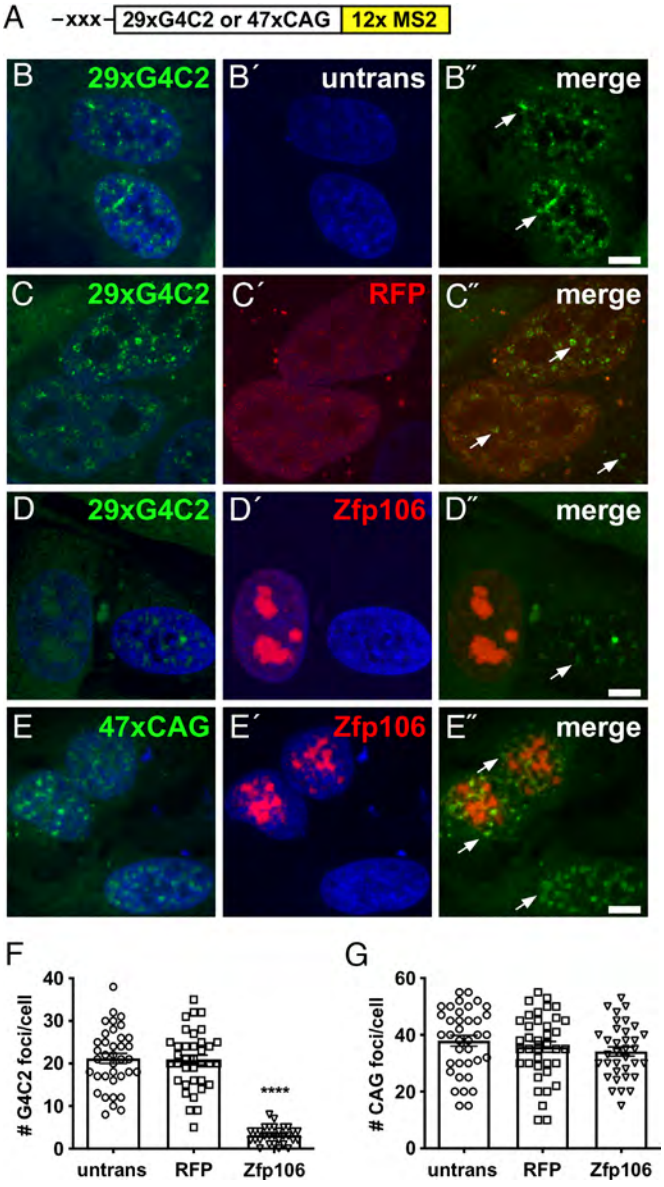

Fig. 2. Zfp106 suppresses formation of GGGGCC-containing RNA foci. Representative fluorescence images of U-2OS cells, stably expressing either 29× GGGGCC (*B–D*) or 47× CAG repeats (*E*) fused to 12× MS2 RNA stem-loops (schematic shown in *A*), that are untransfected (*B*), or have been transfected with nuclear RFP (*C*) or RFP-Zfp106 (*D* and *E*). Cell nuclei were counterstained with DAPI (blue). (*B–E*) green and blue channel merge; (*B′–E′*) red and blue channel merge; (*B″–E″*) green and red channel merge. Arrows mark cells with >10 GGGGCC nuclear foci. (Scale bars, 10 μm.) Quantification of GGGGCC RNA repeat foci (*F*) and CAG RNA repeat foci (*G*) per cell analyzed by one-way ANOVA and Bonferroni's multiple comparison test. For each experimental condition, a minimum of 30 cells from at least 10 fields were imaged across three independent transfection experiments and analyses. Zfp106 coexpression significantly reduced the number of GGGGCC RNA foci (*F*) but had no significant effect on the formation of CAG RNA foci (*G*); ****$P < 0.0001$.

significantly to all non-G-quadruplex-forming sequences (Fig. 3*B*). Indeed, Zfp106 bound efficiently to GGGGCC repeats under conditions that promote G-quadruplex formation (Fig. 3*B*). In contrast, under those same conditions, Zfp106 showed no binding to antisense GGCCCC repeats (non-G-quadruplex-forming), CUG repeats, GGGGCC repeats with G>A substitutions that impede G-quadruplex formation (MUT), or AAAACC repeats (Fig. 3*B*). Likewise, Zfp106 showed little to no detectable binding to a GC stem-loop (GC-SL) sequence (Fig. 3*B″*) or to CAG repeats (Fig. 3*C*), consistent with the observations that Zfp106 did not suppress translation or formation of foci from transcripts containing CAG repeats (Figs. 1*B* and 2 *E* and *G*).

The full-length isoform of Zfp106 used in these studies is ~150 kD, comprises 1,099 amino acids, and is composed primarily of an N-terminal zinc finger, predicted low complexity domains, seven WD-40 domains and two C-terminal zinc fingers (Fig. 3*A*) (3, 48). To gain additional insight into RNA binding by Zfp106, we next sought to identify determinants in Zfp106 responsible for binding to GGGGCC repeats. We found that a C-terminal 43kD-fragment of Zfp106, comprising amino acid residues 712–1099 and encompassing the WD40 domains and the two C-terminal zinc fingers was sufficient for robust, specific binding to $(GGGGCC)_4$ repeats (Fig. 3*D*). Other regions of Zfp106, including a fragment encoding the N terminus and a fragment including N-terminal and C-terminal zinc fingers, but lacking intervening low complexity regions and WD40 repeats, showed no detectable binding to $(GGGGCC)_4$ repeats (*SI Appendix*, Fig. S3).

GGGGCC RNA repeats form parallel G-quadruplexes (12). In addition to binding to $(GGGGCC)_4$ and $(GGGGCC)_3$ repeats, Zfp106 also bound efficiently to other parallel G-quadruplex-forming RNA sequences, including sequences in TERRA and NEAT1 (Fig. 3 *B′* and *B″*) (49, 50). Mutation of TERRA repeats with G>U substitutions, which impede G-quadruplex formation, prevented Zfp106 binding (Fig. 3*B′*). Interestingly, the predicted antiparallel G-quadruplex-forming sequence from ORN-N (51) was also bound by Zfp106 (Fig. 3*B′*). Zfp106 binding to NEAT1 RNA is particularly interesting since GGGGCC RNA repeats have been suggested to replace NEAT1 RNA as a scaffold in paraspeckle-like structures and to form foci with characteristics of paraspeckles (52, 53). Moreover, the DPR protein poly-PR binds to and up-regulates NEAT1, resulting in increased paraspeckle formation, which might contribute to the neurotoxic effect of poly-PR (54).

As further evidence for the specificity of Zfp106 binding to G-quadruplex-forming structures, rather than to primary sequences in RNA, we found that addition of KCl to Zfp106-$(GGGGCC)_4$ EMSA increased Zfp106 binding, whereas addition of LiCl to Zfp106-$(GGGGCC)_4$ EMSA decreased Zfp106 binding (Fig. 3 *E* and *E′*). This is significant because it is well understood that G-quadruplexes are differentially stabilized or destabilized by monovalent cations with potassium having a stabilizing effect and lithium exerting a destabilizing effect (55, 56). Similarly, addition of increasing concentration of TMPyP4, a cationic porphyrin derivative that binds and distorts the G-quadruplex structure of GGGGCC repeats (57, 58), decreased Zfp106 binding to $(GGGGCC)_4$ repeats (Fig. 3*E″*).

Zfp106 Alters the G-Quadruplex Structure of GGGGCC Repeats upon RNA Binding. The threshold number of ~30 GGGGCC repeats for disease is similar to the repeat length in which cells in culture exhibit foci, but higher than the number of repeats required for RNA gelation in vitro (38). For instance, as few as 5× GGGGCC RNA repeats form spherical, gel-like clusters in vitro, but this number of repeats is clearly not sufficient for formation of foci or phase separation in vivo (38). The discrepancy between the number of repeats required for phase transition and gelation in vitro versus in vivo may be due to cellular proteins that alter the G-quadruplex structure in a way that prevents phase separation in vivo (38, 59). Therefore, we reasoned that the disruption of RNA foci and suppression of neurotoxicity caused by GGGGCC repeats in cells and flies might involve the ability of Zfp106 to bind to and alter the RNA G-quadruplex secondary structure. To test this idea, we analyzed the interaction of the C-terminal 43kD-fragment of Zfp106 (Zfp106[WD40+ZnF]) with GGGGCC repeats by

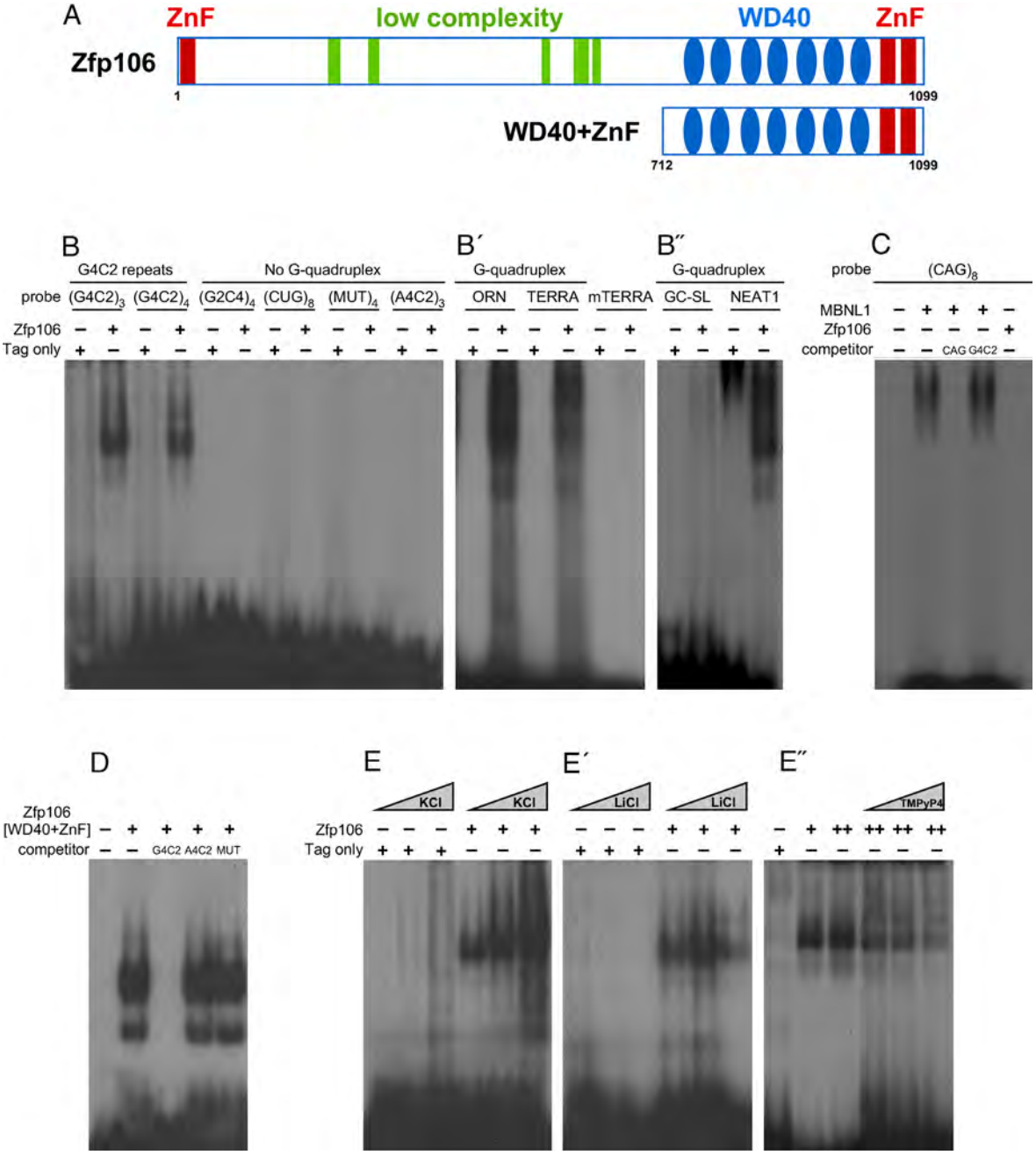

Fig. 3. Zfp106 binds G-quadruplex RNAs. Schematic (*A*) of Zfp106 full-length and C-terminal deletion construct (WD40+ZnF) tested for binding to $(GGGGCC)_4$ and other G-quadruplex and non-G-quadruplex sequences via RNA EMSA. Numbers denote the position of the amino acid residues relative to the full-length Zfp106 isoform XM_006499044.2. (*B*–*B*″) RNA EMSAs performed with purified Zfp106 protein demonstrate that Zfp106 directly binds G-quadruplex forming RNAs in vitro: GGGGCC [$(G4C2)_3$, $(G4C2)_4$] repeats (*B*), ORN and TERRA (*B*′) and NEAT1 (*B*″). Zfp106 showed no binding to the following non-G-quadruplex-forming RNAs: antisense GGCCCC repeats $(G2C4)_4$, $(CUG)_8$, mutated GGGGCC [$(MUT)_4$], $(A4C2)_3$, mutated TERRA (mTERRA), and GC-SL. (*C*) RNA EMSA using purified Zfp106 protein demonstrates that Zfp106 does not bind $(CAG)_8$ repeats under conditions in which the RNA-binding protein MBNL1 efficiently binds $(CAG)_8$. MBNL1 binding to $(CAG)_8$ is specific since binding is competed by 30× molar excess of self-competitor but not 30× molar excess of $(GGGGCC)_4$. (*D*) RNA EMSA using purified Zfp106 C terminus [WD40+ZnF] demonstrates that [WD40+ZnF] is sufficient to bind to $(GGGGCC)_4$ repeats in vitro. 30× molar excess of unlabeled self-competitor, but not 30× molar excess of unlabeled $(AAAACC)_4$ or $(MUT)_4$, efficiently competed with [WD40+ZnF] binding to $(GGGGCC)_4$, establishing specificity of the interaction. (*E*–*E*″) RNA EMSA using purified Zfp106 protein shows increased binding of Zfp106 to $(GGGGCC)_4$ with increasing concentration of KCl (3–30–100 mM) and decreased binding with increasing concentrations of LiCl (3–30–100 mM) and TMPyP4 (10–25–50 μM).

circular dichroism (CD). CD spectroscopy confirmed that the $(GGGGCC)_4$ RNA molecules formed the expected parallel G-quadruplex structure, with a spectral minimum at 236 nm and a spectral maximum at 264 nm (Fig. 4*A*). The CD spectra for Zfp106[WD40+ZnF] alone showed negative peaks in the 210 to 220 nm range but gave minimal to no signal at 264 nm (Fig. 4*C*), allowing us to monitor the G-quadruplex structure of $(GGGGCC)_4$ in the presence and absence of Zfp106 during temperature unfolding (Fig. 4 *A* and *B*).

Addition of the C-terminal 43kD-fragment of Zfp106 (Zfp106[WD40+ZnF]) at an equimolar ratio of protein to RNA resulted in a conformational change in the $(GGGGCC)_4$ G-quadruplex structure, reflected by the lowering and broadening of the peak at 264 nm and by a shift and change in shape of the CD spectra in the 250 to 300 nm range (Fig. 4 *A*′ and *B*′). Closer examination of CD spectra in the 250 to 300 nm range showed a shift in the peak from 264 nm to 274 nm for G4C2 repeats alone as the temperature increased, suggesting a change in the

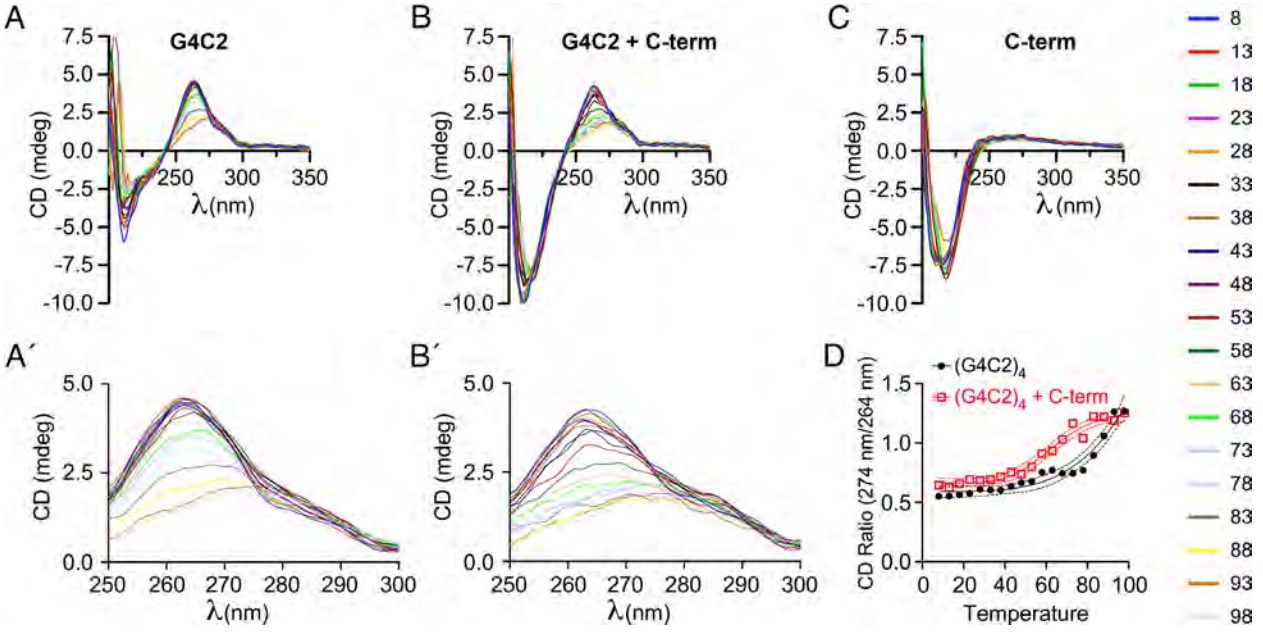

Fig. 4. Zfp106 [WD40+ZnF] alters the G-quadruplex structure of r(GGGGCC)$_4$. (*A* and *A′*) CD spectra for (GGGGCC)$_4$ RNA showing a characteristic parallel G-quadruplex structure with a minimum at 236 nm and a maximum at 264 nm. (*B* and *B′*) CD spectra for r(GGGGCC)$_4$ plus Zfp106 C terminus [WD40+ZnF]. (*C*) CD spectra for Zfp106 C terminus [WD40+ZnF] alone. Zfp106 C terminus binding to (GGGGCC)$_4$ RNA repeats caused a change in shape of the CD spectra in the 250 to 300 nm region with a shift of the peak at 264 nm to 274 nm (*A′* and *B′*). CD spectra were measured during a thermal unfolding with increasing temperature from 8 to 98 °C, depicted by the color scale on the *Right* of the figure. (*D*) Ratio of CD absorbance at 274 nm/264 nm from 8 to 98 °C shows that binding of Zfp106 C terminus [WD40+ZnF] caused a conformational shift in the G-quadruplex structure with a shift from absorbance at 264 nm to 274 nm occurring at lower temperatures (red squares) than in control (black circles). Representative data from one out of three independent experiments performed with three separate protein preparations are shown.

G-quadruplex conformation (Fig. 4*A′*). Interestingly, this shift in the peak from 264 nm to 274 nm occurred at lower temperatures in the presence of Zfp106 (Fig. 4*B′*), suggesting that the presence of Zfp106 promotes the G-quadruplex conformation normally favored at higher temperature in the absence of Zfp106. This observation was more clearly indicated by comparing the ratio of the CD absorption at 274 nm to 264 nm in the presence and absence of Zfp106 C terminus (Fig. 4*D*). We also conducted essentially identical experiments with full-length Zfp106 at 1:2 molar ratio of protein to RNA, and similar results were obtained (*SI Appendix*, Fig. S4), confirming a change in the (GGGGCC)$_4$ G-quadruplex structure upon full-length Zfp106 binding.

The ability of Zfp106 to induce a conformational change in the G-quadruplexes formed by GGGGCC repeats is particularly interesting given prior studies that have suggested that phase transition and formation of RNA foci is dependent on the G-quadruplex structure (30, 38, 39). Indeed, as discussed earlier, small molecules targeting the G-quadruplex structure of GGGGCC repeat RNAs can reduce RNA foci formation, DPR levels, and GGGGCC-induced toxicity (33, 41). It has been suggested that (GGGGCC)$_4$ might form both intermolecular and intramolecular G-quadruplexes in a concentration- and temperature-dependent manner (60). However, when we analyzed (GGGGCC)$_4$ repeats by analytical ultracentrifugation, we failed to detect intermolecular interactions between (GGGGCC)$_4$ molecules under the same buffer conditions and RNA concentration used in our CD analysis. Thus, our data suggest that Zfp106 binding to GGGGCC repeats induces a shift in the intramolecular G-quadruplex structure, although we cannot rule out a possible effect on intermolecular forms at higher concentrations of GGGGCC repeats such as might occur in vivo.

Given the observation that the C-terminal WD40+ZnF region of Zfp106 was sufficient for G-quadruplex binding and conformational shift, we tested the effect of that fragment on RAN translation and RNA foci formation caused by GGGGCC repeats. Interestingly, neither the C-terminal WD40+ZnF nor the N-terminal N+LCR significantly affected translation of GGGGCC repeat–containing RNAs via conventional translation (*SI Appendix*, Fig. S1*E*) or via RAN translation (*SI Appendix*, Fig. S1*F*). Similarly, neither the C-terminal WD40+ZnF nor the N-terminal N+LCR significantly affected the formation of RNA foci caused by expression of 29× GGGGCC repeats (*SI Appendix*, Fig. S2 *B–D*). Taken together, these observations indicate that other regions of the Zfp106 protein in addition to the C-terminal DNA-binding region, including the N-terminal region, the low complexity domains, or both (Fig. 4*A*), are essential for its function in suppressing the formation of RNA foci and RAN translation in vivo.

Overall, the studies presented here demonstrate that the Zfp106 C-terminal RNA-binding domain is necessary, but not sufficient, for its inhibitory effects on repeat-induced foci formation and RAN translation and support the notion that Zfp106's suppressive effect on GGGGCC-mediated cytotoxicity is likely through its RNA-binding activity and subsequent alteration of specific G-quadruplex structures formed by GGGGCC repeats.

GGGGCC Repeats May Exert a "Sponge"-Like Effect on Zfp106 or Other G-Quadruplex RNA-Binding Proteins. Previous studies have demonstrated that GGGGCC RNA repeats sequester various G-quadruplex-binding proteins involved in diverse cellular functions, and this sequestration may contribute to the cytotoxic effects of the repeats (11, 61). For example, the G-quadruplex binding protein nucleolin is sequestered by GGGGCC RNA repeats, resulting in nucleolin mislocalization and nucleolar stress in *C9orf72* ALS patients (12). Similarly, sequestration of the G-quadruplex binding protein hnRNPH/F by GGGGCC RNA repeats results in defective RNA splicing and polyadenylation (11, 13, 14). GGGGCC repeats have also been shown to compromise nucleocytoplasmic transport by sequestering the RNA-binding

protein RanGAP1 (62). Thus, it is interesting to speculate that part of the toxic effects of GGGGCC repeats in the context of *C9orf72* ALS/FTD may be due to sequestration of Zfp106. Indeed, loss of *Zfp106* function in mice causes severe ALS-like neuromuscular disease (1, 2, 7), supporting the idea that Zfp106 sequestration could contribute to neurodegenerative disease. However, precisely how Zfp106 loss-of-function or sequestration leads to neurodegeneration is unclear since the specific function of Zfp106 in RNA processing and metabolism remains to be determined. The observations that Zfp106 binds to G-quadruplex RNAs and induces conformational changes in the G-quadruplex structure of GGGGCC repeats suggests that Zfp106 might play a role in G-quadruplex RNA homeostasis and metabolism, as has been suggested for other G-quadruplex binding proteins (63). Future studies will address the functional importance of Zfp106 binding to G-quadruplex-forming cellular RNAs and whether dysregulation of the "G4ome" (63) is a contributing mechanism for neurodegeneration in *C9orf72* ALS/FTD or other neurodegenerative diseases.

Materials and Methods

Plasmids and Lentivirus. GFP reporter constructs with either 3× or 30× GGGGCC repeats inserted in the 5′-UTR of an eGFP cDNA in plasmids pCMV-$(GGGGCC)_3$-EGFP and pCMV-$(GGGGCC)_{30}$-EGFP, respectively, have been described previously (44). Plasmid pCMV-$(CAG)_6$-EGFP was generated by cloning a synthetic oligonucleotide for 6× CAG repeats into the EcoRI and KpnI sites in between the transcription site and the translation site of pEGFP-N3, as previously described for GGGGCC repeats (44). Similarly, plasmid pCMV-$(CAG)_{60}$-EGFP was generated by cloning 30× CAG repeats into the EcoRI and KpnI sites of pEGFP-N3, followed by cloning of an additional 30× CAG repeats into the XhoI and KpnI sites of pEGFP-N3, also as previously described for GGGGCC repeats (44). Bicistronic reporters pcDNA5 FRT TO FLuc2-noC9R-NLuc (Neg), pcDNA5 FRT TO FLuc2-C9R70-NLuc Frame GA (Frame GA) and bicistronic splicing reporters pcDNA5 FRT TO noC9R intron-NLuc-exFLuc 2 (Int-Neg) and pcDNA5 FRT TO C9R70 intron-NLuc-exFLuc 2 Frame GA (Int-Frame GA) have been described previously (45).

Plasmid pCMV-GFP-Zfp106 was created by cloning the Zfp106 cDNA into the XhoI-EcoRI sites in pEGFP-C2 (Clontech). Plasmid pCMV-RFP-Zfp106 was created by replacing EGFP in plasmid pCMV-GFP-Zfp106 with NLS-TagRFP. 3×FLAG-2×STREP-Zfp106 has been described previously (2). Truncation mutants and deletion mutants of Zfp106 were generated by PCR and confirmed by sequencing; Zfp106 cDNA fragments, encoding amino acids 712–1099 of the C-terminal region, encoding amino acids 1–711 of the N-terminal region, and a cDNA with an internal deletion such that the resultant cDNA encodes amino acids 1–178 and 1025–1099 were digested and subcloned into the NotI and XhoI sites of pcDNA4/TO (Invitrogen). In these constructs, pcDNA4/TO was modified to include N-terminal 3×FLAG-2× STREP tags to create plasmids 3×FLAG-2×STREP-Zfp106[Cterm], 3×FLAG-2×STREP-Zfp106[Nterm], and 3×FLAG-2×STREP-Zfp106[N+Cterms], respectively. 3×FLAG-2×STREP-Zfp106[Cterm] encodes Zfp106[WD40+ZnF].

For lentiviral transduction, GFP and GFP-ZFP106 were subcloned into the XbaI and EcoRI sites of a pLV-EF1a vector (generous gift from Michael McManus), which was modified by removal of the Staygold fluorescent reporter to allow for N-terminal GFP tagging. Lentiviral packaging and high-titer ($>10^8$) production was performed by Genewiz (Azenta Life Sciences).

Cell Culture, Transfections, Reporter Assays, and DPR Detection. Mouse Neuro-2a (ATCC, CCL-131; RRID:CVCL_0470) and human HEK293T cells (ATCC, CRL-3216; RRID:CVCL_0063) were obtained directly from ATCC and maintained in Dulbecco's modified Eagle Medium supplemented with 10% fetal bovine serum (FBS, Gibco, Waltham, MA) and penicillin-streptomycin (Gibco). Monoclonal U-2OS cell lines stably expressing Tet-On 3G transactivator protein (Clontech), a tandem-dimeric MS2 hairpin binding protein tagged with eYFP (MS2CP-YFP), and 29× GGGGCC or 47× CAG repeat under control of a doxycycline-inducible promoter have been described previously (38). All cell lines were routinely tested for Mycoplasma contamination with the MycoAlert PLUS Mycoplasma Detection Kit (Lonza, Switzerland, LT07-705) and found to be negative.

Neuro-2a and U-2OS cells were transfected using GenJet In Vitro DNA Transfection Reagent for Neuro-2A Cells or U-2OS respectively (Signagen); HEK293T cells were transfected using PolyJet DNA In Vitro Transfection Reagent (Signagen), according to the manufacturers' recommendations.

Protein purification from HEK293T cells was performed as described previously (2).

The iPSC line from a patient harboring the *C9orf72* mutation (~200 GGGGCC repeats) was generated previously, as described (64). iPSCs were maintained in mTesR+ medium (Stem Cell Technologies) and plated on Matrigel (Corning 356231). iPSCs were passaged at 60 to 80% confluency. The human *C9orf72* iPSC line had a normal karyotype and negative quarterly Mycoplasma testing. iPSCs were lysed in RIPA buffer containing 2% SDS and 3× Halt Protease and Phosphatase Inhibitor Cocktail (78441, Thermo Scientific). The lysates were then subjected to Poly-GA dipeptide repeat protein quantification by MSD immunoassay with antibody MABN889 (Sigma Aldrich), as described recently (46).

For reporter assays, three biological replicates were analyzed for each condition for all GFP and dual-luciferase experiments. For bicistronic reporters, gene expression was induced with 2 μg/mL doxycycline for 24 h prior to sample collection. Nano-luciferase (NLuc) and firefly luciferase (FLuc) activities were measured by using the Nano-Glo Dual Luciferase Assay on a Glomax Multidetection System (Promega). GFP fluorescence was measured using an Infinite M1000 microplate reader (Tecan Group Ltd.); samples were normalized by total protein concentration using the Bradford assay with protein assay dye reagent (Bio-Rad).

qPCR, Imaging of RNA Foci, and Fluorescence In Situ Hybridization (FISH). For RT-qPCR, total RNA was isolated from cells using Trizol (Invitrogen) and was treated for 1 h at 37 °C with DNaseI, followed by cDNA synthesis using random primers and the Omniscript RT kit (Qiagen). Nuclear and cytoplasmic RNA fractions were isolated with the PARIS Kit (Ambion), according to the manufacturer's instructions. qPCR was performed using the MAXIMA SYBR Green kit (Thermo Scientific) and a 7900HT Fast Real Time PCR System or a Quant Studio 5 Real Time PCR System (Applied Biosystems) with the primers listed in *SI Appendix*, Table S1. Data were normalized to housekeeping gene *Gapdh* by the $2-\Delta\Delta Ct$ method (65).

Repeat RNA expression in U-2OS cells stably transfected with 29× GGGGCC or 47× CAG repeats was induced by adding 1 μg/mL doxycycline at the time of transfection with the RFP-Zfp106 plasmid described above. Twenty-four hours posttransfection, cells were fixed in 4% paraformaldehyde for 10 min and imaged in 0.3 μm Z-stacks using a Nikon Ti-Eclipse microscope equipped with Yokogawa confocal CSU-X spinning disk module. The number of foci per cell was calculated using the FIJI 3D Objects Counter plugin as previously described (38).

For RNA FISH, repeat-expressing U-2OS cells were induced for 24 h and were then harvested and fixed in 2% paraformaldehyde for 10 min at room temperature. Following fixation, cells were permeabilized by overnight incubation in 70% ethanol at 4 °C. RNA was detected using Cy5-labeled DNA oligonucleotides designed against the MS2-hairpin sequence; probe sequences were previously described (38). Stellaris RNA FISH and wash buffers were purchased from Biosearch Technologies and used according to the manufacturer's recommendations. After labeling, samples were mounted in Prolong Gold antifade medium (Thermo Scientific, Inc.) and imaged using confocal microscopy as described above.

EMSA. One hundred picomoles of single-stranded RNA (Integrated DNA Technologies, Inc.) was 5′-end labeled with [γ-^{32}P]ATP, unincorporated radioactivity was removed, and RNA EMSA was performed as described previously (2). Full-length Zfp106 and indicated truncation and deletion mutants or MBNL1 protein were purified from total cellular lysates of transfected HEK293T using the Strep-Tag system (66) with Strep-Tactin Sepharose (IBA, catalog #2-1201-010), according to the manufacturer's protocol; 0.1% Triton X-100 was added to all buffers for cell lysis and chromatography. An equivalent amount of total protein was used for each EMSA. Parallel purification from HEK293T cells transfected with the 3×FLAG-2×STREP parental vector was performed, and the corresponding fractions in which Zfp106 and MBNL1 proteins were eluted were used in RNA EMSA as a negative control. RNA–protein mixtures were electrophoresed in 6% polyacrylamide-TBE gels, which were subsequently dried and subjected to autoradiography. For competition experiments, the purified proteins were preincubated for 10 min at room temperature with 30-fold molar excess of unlabeled

single-stranded RNA competitor. The single-stranded RNA probe sequences are provided in *SI Appendix*, Table S2.

CD Spectroscopy. CD spectra from 400 to 190 nm were collected at temperatures between 8 °C and 98 °C, with a 5 °C/min temperature gradient, using a Jasco J-810 equipped with a Jasco Peltier temperature control system. CD spectra were recorded in the presence of 4 μM $(GGGGCC)_4$ RNA, either alone or with addition of 4 μM Zfp106 C-terminal fragment (Zfp106[WD40+ZnF]) or 2 μM full-length Zfp106 in 10 mM sodium cacodylate (pH 7.35). Prior to use in CD, Zfp106 and Zfp106[WD40+ZnF] proteins were purified as described above and then buffer exchanged to 10 mM sodium cacodylate (pH 7.35) and concentrated using the Amicon Pro Purification System (Millipore, Sigma). Three scans were recorded and averaged for each sample. A CD spectrum of the buffer was recorded and subtracted from all spectra.

Statistical Analyses. Statistical analyses were performed using GraphPad Prism 5.0 (GraphPad Software; RRID:SCR_002798). Data were analyzed by one-way or two-way ANOVA followed by Bonferroni's multiple comparison test or Dunnett's multiple comparison test.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. HTK was supported by NIH Grant K99GM138753. B.C. received support from the University of California, San Francisco Program for Breakthrough Biomedical Research, which is partially funded by the Sandler Foundation. This work was supported primarily by Grant AL210129 from the Department of Defense Congressionally-Directed Medical Research Program and Grant NS126499 from the NIH to B.L.B. with additional support from NIH Grants R35 GM122603 to W.F.D. and DK119621 and HL146366 to B.L.B. S.E.S. and C.D.C. are supported by NIH Grants U19NS132303 and U01NS134062. We are grateful to Peng Jin (Emory), Shuying Sun (Johns Hopkins), and Michael McManus (UCSF) for providing plasmids and Ron Vale (Janelia) and Ankur Jain (Massachusetts Institute of Technology) for sharing cells and for helpful advice. We thank Bill Seeley, Sarat Vatsavayai, Jen Yokoyama, Michael McManus, and members of the Black and DeGrado labs for helpful discussions.

Author affiliations: [a]Cardiovascular Research Institute, University of California, San Francisco, CA 94143; [b]Weill Institute for Neurosciences, University of California, San Francisco, CA 94143; [c]Memory & Aging Center, Department of Neurology, University of California, San Francisco, CA 94143; [d]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143; and [e]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143

Author contributions: B.C., S.E.S., C.D.C., W.F.D., and B.L.B. designed research; B.C., S.E.S., H.W., B.D., and H.T.K. performed research; C.D.C. contributed new reagents/analytic tools; B.C., S.E.S., H.W., B.D., C.D.C., W.F.D., and B.L.B. analyzed data; and B.C. and B.L.B. wrote the paper.

Reviewers: T.J.C., The University of North Carolina at Chapel Hill; V.A.G., Columbia University Irving Medical Center; and O.T., Graduate School of Science and Technology, Shizuoka University.

1. D. M. Anderson *et al.*, Severe muscle wasting and denervation in mice lacking the RNA-binding protein ZFP106. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4494–4503 (2016).
2. B. Celona *et al.*, Suppression of C9orf72 RNA repeat-induced neurotoxicity by the ALS-associated RNA-binding protein Zfp106. *eLife* **6**, e19032 (2017).
3. H. Grasberger, G. I. Bell, Subcellular recruitment by TSG118 and TSPYL implicates a role for zinc finger protein 106 in a novel developmental pathway. *Int. J. Biochem. Cell Biol.* **37**, 1421–1437 (2005).
4. A. Castello *et al.*, Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
5. S. Ide, J. Dejardin, End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. *Nat. Commun.* **6**, 6674 (2015).
6. L. Tafforeau *et al.*, The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Mol. Cell* **51**, 539–551 (2013).
7. P. I. Joyce *et al.*, Deficiency of the zinc finger protein ZFP106 causes motor and sensory neurodegeneration. *Hum. Mol. Genet.* **25**, 291–307 (2016).
8. M. DeJesus-Hernandez *et al.*, Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
9. A. E. Renton *et al.*, A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
10. J. D. Rohrer *et al.*, C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. *Lancet Neurol.* **14**, 291–301 (2015).
11. J. Cooper-Knock *et al.*, Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain* **137**, 2040–2051 (2014).
12. A. R. Haeusler *et al.*, C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195–200 (2014).
13. Y. B. Lee *et al.*, Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep.* **5**, 1178–1186 (2013).
14. M. Prudencio *et al.*, Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat. Neurosci.* **18**, 1175–1182 (2015).
15. D. Sareen *et al.*, Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci. Transl. Med.* **5**, 208ra149 (2013).
16. I. Kwon *et al.*, Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139–1145 (2014).
17. S. Mizielinska *et al.*, C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345**, 1192–1194 (2014).
18. K. Mori *et al.*, The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335–1338 (2013).
19. A. Burberry *et al.*, Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. *Sci. Transl. Med.* **8**, 347ra393 (2016).
20. Q. Zhu *et al.*, Reduced C9ORF72 function exacerbates gain of toxicity from ALS/FTD-causing repeat expansion in C9orf72. *Nat. Neurosci.* **23**, 615–624 (2020).
21. Q. Shao *et al.*, C9orf72 deficiency promotes motor deficits of a C9ALS/FTD mouse model in a dose-dependent manner. *Acta Neuropathol. Commun.* **7**, 32 (2019).
22. Y. Shi *et al.*, Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat. Med.* **24**, 313–325 (2018).
23. C. Sellier *et al.*, Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. *EMBO J.* **35**, 1276–1297 (2016).
24. J. Ugolino *et al.*, Loss of C9orf72 enhances autophagic activity via deregulated mTOR and TFEB signaling. *PLoS Genet.* **12**, e1006443 (2016).
25. L. Nguyen, J. D. Cleary, L. P. W. Ranum, Repeat-associated non-ATG translation: Molecular mechanisms and contribution to neurological disease. *Annu. Rev. Neurosci.* **42**, 227–247 (2019).
26. A. Schmitz, J. Pinheiro Marques, I. Oertig, N. Maharjan, S. Saxena, Emerging perspectives on dipeptide repeat proteins in C9ORF72 ALS/FTD. *Front. Cell Neurosci.* **15**, 637548 (2021).
27. F. Frottin, M. Pérez-Berlanga, F. U. Hartl, M. S. Hipp, Multiple pathways of toxicity induced by C9orf72 dipeptide repeat aggregates and G(4)C(2) RNA in a cellular model. *eLife* **10**, e62718 (2021).
28. S. Boeynaems *et al.*, Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Sci. Rep.* **6**, 20877 (2016).
29. J. Chew *et al.*, Aberrant deposition of stress granule-resident proteins linked to C9orf72-associated TDP-43 proteinopathy. *Mol. Neurodegener.* **14**, 9 (2019).
30. M. M. Fay, P. J. Anderson, P. Ivanov, ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. *Cell Rep.* **21**, 3573–3584 (2017).
31. B. D. Freibaum *et al.*, GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129–133 (2015).
32. S. Hutten, D. Dormann, Nucleocytoplasmic transport defects in neurodegeneration–Cause or consequence? *Semin. Cell Dev. Biol.* **99**, 151–162 (2020).
33. Z. Su *et al.*, Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* **83**, 1043–1050 (2014).
34. K. Reddy, B. Zamiri, S. Y. Stanley, R. B. Macgregor Jr., C. E. Pearson, The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.* **288**, 9860–9866 (2013).
35. A. Kovanda, M. Zalar, P. Šket, J. Plavec, B. Rogelj, Anti-sense DNA d(GGCCCC)n expansions in C9ORF72 form i-motifs and protonated hairpins. *Sci. Rep.* **5**, 17944 (2015).
36. P. Svoboda, A. Di Cara, Hairpin RNA: A secondary structure of primary importance. *Cell Mol. Life Sci.* **63**, 901–908 (2006).
37. S. Millevoi, H. Moine, S. Vagner, G-quadruplexes in RNA biology. *Wiley Interdiscip. Rev. RNA* **3**, 495–507 (2012).
38. A. Jain, R. D. Vale, RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
39. E. G. Conlon *et al.*, The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *eLife* **5**, e17820 (2016).
40. Z. F. Wang *et al.*, The hairpin form of r(G(4)C(2))(exp) in c9ALS/FTD Is repeat-associated non-ATG translated and a target for bioactive small molecules. *Cell Chem. Biol.* **26**, 179–190.e112 (2019).
41. R. Simone *et al.*, G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Mol. Med.* **10**, 22–31 (2018).
42. The Huntington's Disease Collaborative Research Group, A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).
43. G. Raca, E. Y. Siyanova, C. T. McMurray, S. M. Mirkin, Expansion of the (CTG)(n) repeat in the 5′-UTR of a reporter gene impedes translation. *Nucleic Acids Res.* **28**, 3943–3949 (2000).
44. Z. Xu *et al.*, Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7778–7783 (2013).
45. W. Cheng *et al.*, C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. *Nat. Commun.* **9**, 51 (2018).
46. S. E. Salomonsson *et al.*, Validated assays for the quantification of C9orf72 human pathology. *Sci. Rep.* **14**, 828 (2024).
47. Y. Yuva-Aydemir, S. Almeida, F. B. Gao, Insights into C9ORF72-related ALS/FTD from Drosophila and iPSC models. *Trends Neurosci.* **41**, 457–469 (2018).
48. M. Hatayama, J. Aruga, Characterization of the tandem CWCH2 sequence motif: A hallmark of inter-zinc finger interactions. *BMC Evol. Biol.* **10**, 53 (2010).
49. M. Lalonde, P. Chartrand, TERRA, a multifaceted regulator of telomerase activity at telomeres. *J. Mol. Biol.* **432**, 4232–4243 (2020).

50. T. Yamazaki *et al.*, Functional domains of NEAT1 architectural lncRNA induce paraspeckle assembly through phase separation. *Mol. Cell* **70**, 1038–1053.e1037 (2018).
51. C. D. Xiao, T. Shibata, Y. Yamamoto, Y. Xu, An intramolecular antiparallel G-quadruplex formed by human telomere RNA. *Chem. Commun. (Camb)* **54**, 3944–3946 (2018).
52. A. Bajc Česnik *et al.*, Nuclear RNA foci from C9ORF72 expansion mutation form paraspeckle-like bodies. *J. Cell Sci.* **132**, jcs224303 (2019).
53. M. Malnar, B. Rogelj, SFPQ regulates the accumulation of RNA foci and dipeptide repeat proteins from the expanded repeat mutation in C9orf72. *J. Cell Sci.* **134**, jcs256602 (2021).
54. H. Suzuki, Y. Shibagaki, S. Hattori, M. Matsuoka, C9-ALS/FTD-linked proline-arginine dipeptide repeat protein associates with paraspeckle components and increases paraspeckle formation. *Cell Death Dis.* **10**, 746 (2019).
55. C. C. Hardin, T. Watson, M. Corregan, C. Bailey, Cation-dependent transition between the quadruplex and Watson–Crick hairpin forms of d(CGCG3GCG). *Biochemistry* **31**, 833–841 (1992).
56. C. Bardin, J. L. Leroy, The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res.* **36**, 477–488 (2008).
57. B. Zamiri, K. Reddy, R. B. Macgregor Jr., C. E. Pearson, TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *J. Biol. Chem.* **289**, 4653–4659 (2014).
58. H. Alniss, B. Zamiri, M. Khalaj, C. E. Pearson, R. B. Macgregor Jr., Thermodynamic and spectroscopic investigations of TMPyP4 association with guanine- and cytosine-rich DNA and RNA repeats of C9orf72. *Biochem. Biophys. Res. Commun.* **495**, 2410–2417 (2018).
59. J. U. Guo, D. P. Bartel, RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, aaf5371 (2016).
60. M. J. Wortman *et al.*, A synthetic Pur-based peptide binds and alters G-quadruplex secondary structure present in the expanded RNA repeat of C9orf72 ALS/FTD. *Biochim. Biophys. Acta Mol. Cell Res.* **1867**, 118674 (2020).
61. S. Mizielinska, A. M. Isaacs, C9orf72 amyotrophic lateral sclerosis and frontotemporal dementia: Gain or loss of function? *Curr. Opin. Neurol.* **27**, 515–523 (2014).
62. K. Zhang *et al.*, The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56–61 (2015).
63. A. Cammas, S. Millevoi, RNA G-quadruplexes, emerging mechanisms in disease. *Nucleic Acids Res.* **45**, 1584–1595 (2017).
64. M. Pribadi *et al.*, CRISPR-Cas9 targeted deletion of the *C9orf72* repeat expansion mutation corrects cellular phenotypes in patient-derived iPS cells. bioRxiv [Preprint] (2016). (Accessed 8 May 2024).
65. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402–408 (2001).
66. T. G. Schmidt, A. Skerra, The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* **2**, 1528–1535 (2007).

PNAS

Supporting Information for
Zfp106 binds to G-quadruplex RNAs and inhibits RAN translation and formation of RNA foci caused by G4C2 repeats

Barbara Celona, Sally E. Salomonsson, Haifan Wu, Bobo Dang, Huong T. Kratochvil, Claire D. Clelland, William F. DeGrado, and Brian L. Black

William F. DeGrado
Email: william.degrado@ucsf.edu

Brian L. Black
Email: brian.black@ucsf.edu

This PDF file includes:
Figures S1 to S4
Tables S1 to S2

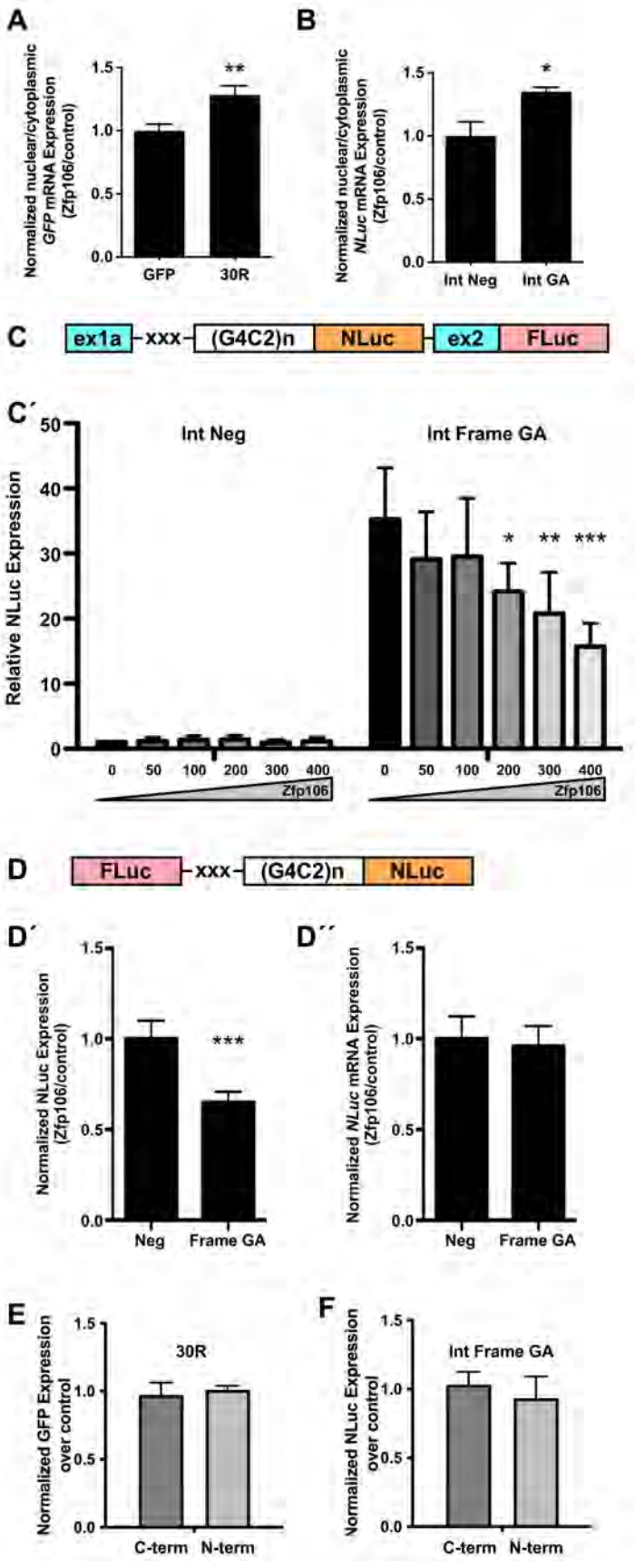
A
Normalized nuclear/cytoplasmic GFP mRNA Expression (Zfp106/control)
1.5
1.0
0.5
0.0
**
GFP
30R
B
Normalized nuclear/cytoplasmic NLuc mRNA Expression (Zfp106/control)
1.5
1.0
0.5
0.0
*
Int Neg
Int GA
C
ex1a
xxx
(G4C2)n
NLuc
ex2
FLuc
C´
Relative NLuc Expression
50
40
30
20
10
0
Int Neg
Int Frame GA
* ** ***
0 50 100 200 300 400
Zfp106
0 50 100 200 300 400
Zfp106
D
FLuc
xxx
(G4C2)n
NLuc
D´
Normalized NLuc Expression (Zfp106/control)
1.5
1.0
0.5
0.0
***
Neg
Frame GA
D´´
Normalized NLuc mRNA Expression (Zfp106/control)
1.5
1.0
0.5
0.0
Neg
Frame GA
E
Normalized GFP Expression over control
1.5
1.0
0.5
0.0
30R
C-term
N-term
F
Normalized NLuc Expression over control
1.5
1.0
0.5
0.0
Int Frame GA
C-term
N-term

Fig. S1. Zfp106 suppresses RAN translation. (A) Ratio of nuclear to cytoplasmic RNA levels indicates nuclear retention of 30R-GFP mRNA but not GFP-only (no repeats) mRNA in Zfp106-transfected cells compared to control. (B) Ratio of nuclear to cytoplasmic RNA levels indicates nuclear retention of NLuc-mRNA from Int-GA (70× GGGGCC repeats) but not Int-neg (no-repeats) in Zfp106-transfected cells compared to control (C). NLuc fluorescence in HEK293T cells co-transfected with bicistronic splicing dual-luciferase reporters with *C9orf72* exons and intron (Int Neg, no repeats; Int Frame GA, 70× GGGGCC repeats) and increasing amounts of Zfp106 (shown at the bottom of the graph in ng/cm^2). Parental control vector DNA was added to each transfection to ensure an equivalent amount of plasmid DNA up to 400 ng/cm^2. No-repeat control transfected with parental control was set to a normalized value of 1. Zfp106 co-expression significantly suppressed RAN translation from the Int Frame GA construct containing 70× repeats in a dosage-dependent fashion. (D) NLuc luminescence (D′) as a readout of RAN translation and mRNA levels (B′′) in HEK293T cells co-transfected with bicistronic (no splicing) dual-luciferase reporters (Neg, no repeats; Frame GA, 70× GGGGCC repeats) and Zfp106 or control vector. NLuc luminescence was normalized to FLuc in each sample. Data are expressed as ratios of Zfp106-transfected cells over control-transfected cells. The mean ratio from the no-repeat construct (Neg) was set to a normalized value of 1. (E,F) eGFP fluorescence (E) or NLuc luminescence (F) in Neuro-2a cells (E) or HEK293T cells (F) co-transfected with pCMV-(GGGGCC)$_{30}$-EGFP construct (30× GGGGCC repeats) and C-terminal (WD40+ZnF) or N-terminal (N+LCR) Zfp106 deletion constructs (E) or co-transfected with bicistronic splicing dual-luciferase reporter with *C9orf72* exons Int Frame GA (70× GGGGCC repeats) and C-terminal or N-terminal Zfp106 deletion constructs. GFP fluorescence was normalized to total protein content; NLuc luminescence was normalized to FLuc in each sample. Data are expressed as ratios of C-term-transfected or N-term-transfected cells over control. All data were analyzed by one-way ANOVA and Bonferroni's Multiple Comparison Test; * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Schematic depictions of transfected constructs are shown in (C,D); xxx indicates stop codons.

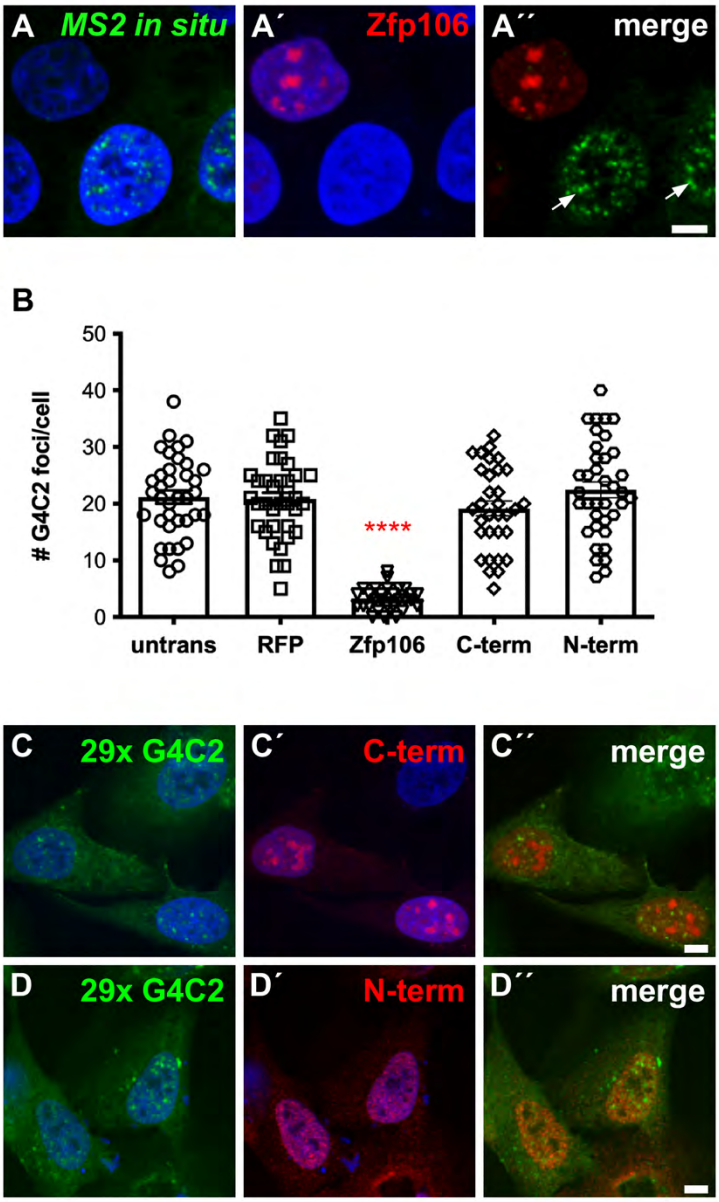

Fig. S2. Zfp106 suppresses formation of GGGGCC-containing RNA foci. RNA fluorescence *in situ* hybridization (FISH) using probes directed against MS2 hairpin loops (green) confirms that full length Zfp106 (A) co-expression (red), but not co-expression of the Zfp106 C-terminus (C) nor N-terminus (D), significantly reduced the number of GGGGCC RNA foci in the 29× GGGGCC U-2OS cell line. Cell nuclei were counterstained with DAPI (blue). The green and blue channels are merged in (A,C,D). The red and blue channel s are merged in (A′,C′,D′). The green and red channels are merged in A′′,C′′,D′′). Arrows in (A′′) mark cells with >10 GGGGCC nuclear foci. Scale bars, 10 μm. (B) Quantification of GGGGCC RNA repeat foci per cell analyzed by one-way ANOVA and Bonferroni's Multiple Comparison Test. For each experimental condition, a minimum of 30 cells from at least 10 fields were imaged across three independent experiments. Zfp106 co-expression significantly reduced the number of GGGGCC RNA foci (B) while neither the C-terminus of Zfp106 (C-term) nor the N-terminus of Zfp106 (N-term) had a significant effect on the formation of GGGGCC RNA foci. Note that data in (B) for untransfected, RFP, and Zfp106 are identical to data in Fig 2F; **** $p<0.0001$.

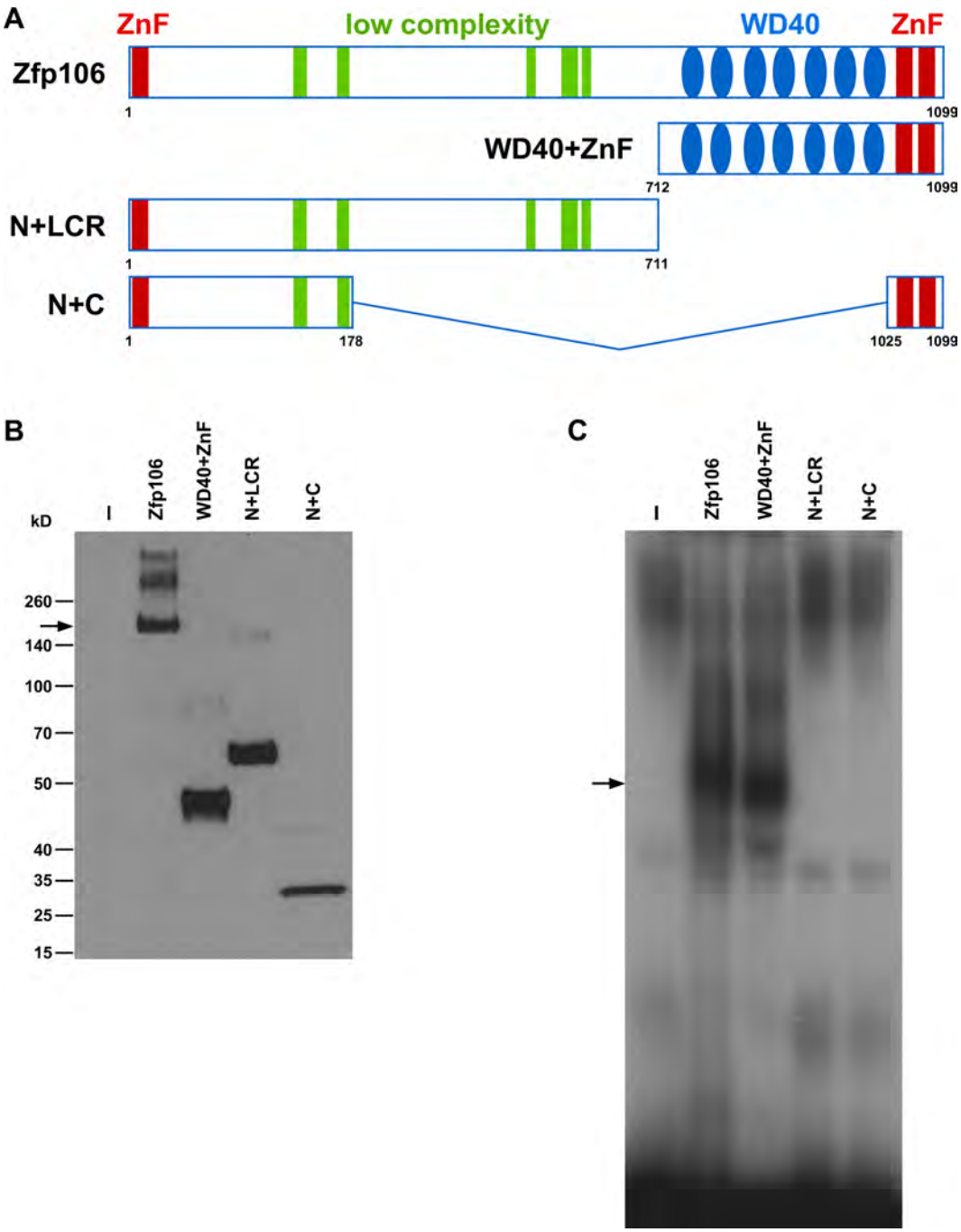

Fig. S3. The Zfp106 C-terminus [WD40+ZnF] is sufficient for bind to r(GGGGCC)$_4$. Schematic (A) and western blot (B) of Zfp106 deletion constructs tested for binding to (GGGGCC)$_4$ via RNA EMSA. Note that the schematic depictions of Zfp106 and for [WD40+ZnF] are identical to the schematics shown in Fig. 3A. Numbers denote the position of the amino acid residues relative to the full-length Zfp106 isoform XM_006499044.2. The arrow in (B) indicates full-length Zfp106. The vertical dash (–) indicates parental vector control. (C) Full-length Zfp106 and [WD40+ZnF] comparably bound to r(GGGGCC)$_4$, while [the N-terminal and low complexity regions (LCR) [N+LCR] and [N+C] showed no detectable binding to r(GGGGCC)$_4$ by EMSA. The arrow in (C) indicates specific gel shift. The vertical dash (–) indicates parental vector control.

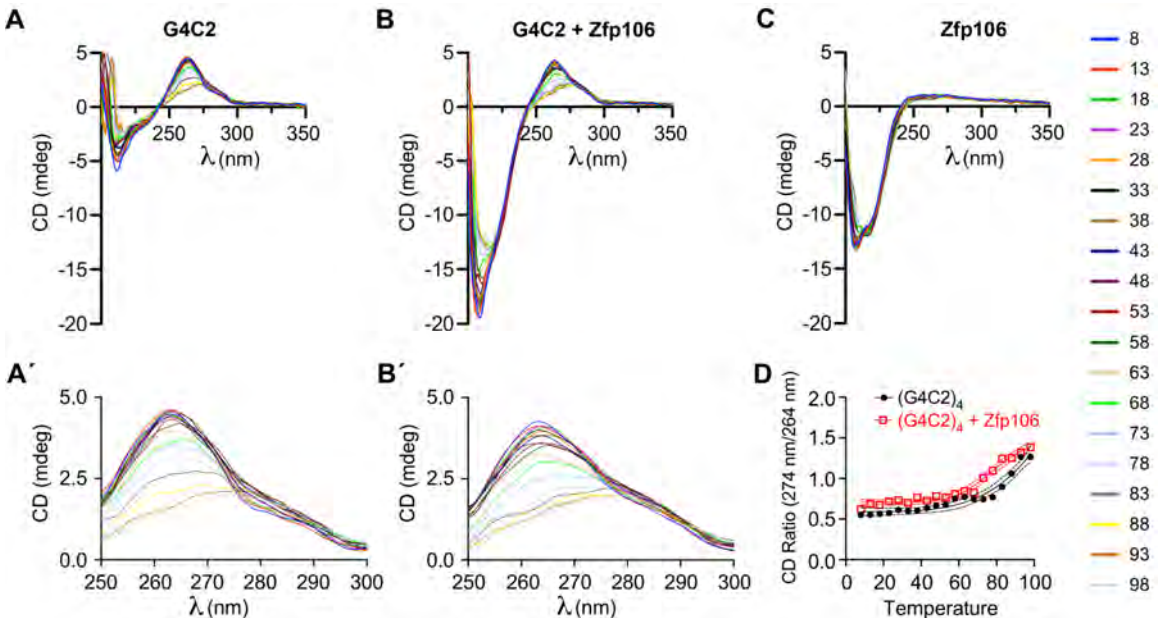

Fig. S4. Zfp106 alters the G-quadruplex structure of r(GGGGCC)$_4$. (A,A′) CD spectra for (GGGGCC)$_4$ RNA showing a characteristic parallel G-quadruplex structure with a minimum at 236 nm and a maximum at 264 nm. (B,B′) CD spectra for r(GGGGCC)$_4$ plus full length Zfp106. (C) CD spectra for Zfp106 alone. Zfp106 binding to GGGGCC)$_4$ RNA repeats caused a change in shape of the CD spectra in the 250-300 nm region with a shift of the peak at 264 nm to 274 nm (A′,B′). CD spectra were measured during a thermal unfolding with increasing temperature from 8 to 98°C, depicted by the color scale on the right of the figure. D) Ratio of CD absorbance at 274 nm/264 nm from 8 to 98°C shows that binding of Zfp106 caused a conformational shift in the G-quadruplex structure with a shift from absorbance at 264 nm to 274 nm occurring at lower temperature (red squares) than in control (black circles) even when present in a 1:2 molar ratio. Representative data from one out of 3 independent experiments performed with three separate protein preparations are shown. Note: panels A,A′ are identical to panels A,A′ in Fig. 4.

Table S1. Primers used for RT-qPCR (qPCR)

gene	5′ primer	3′ primer
egfp	5′-gcacgacttcttcaagtccgccatgcc-3′	5′-gcggatcttgaagttcaccttgatgcc-3′
GAPDH (human)	5′-gagtcaacggatttggtcgt-3′	5′-ttgattttggagggatctcg-3′
gapdh (mouse)	5′-accacagtccatgccatcac-3′	5′-tccaccaccctgttgctgta-3′
NLuc	5'-gtccgtaactccgatccaaag-3′	5′-tgccatagtgcaggatcacct-3′
FLuc	5′-gtgacttcccatttgccacc-3′	5′-tgatctggttgccgaagatg-3′
FLuc (spliced mRNA)	5′-gaggtgcgtcaaacagcgac-3′	5′-tttggcatcttccctcgagg-3′

Table S2. RNA probes used in RNA EMSA.

probe	RNA oligonucleotide sequence
(G4C2)3	5′-GGGGCCGGGGCCGGGGCC
(G4C2)4	5′-GGGGCCGGGGCCGGGGCCGGGGCC-3′
(G2C4)4	5′-GGCCCCGGCCCCGGCCCCGGCCCC-3′
(CUG)8	5′-CUGCUGCUGCUGCUGCUGCUGCUG-3′
MUT	5′-GAGGCCGGGACCGAGACCGAGGCC-3′
(A4C2)3	5′-AAAACCAAAACCAAAACC-3′
ORN	5′-AGGGUUAGGGUUAGGGUUAGGG-3′
TERRA	5′-UUAGGGUUAGGGUUAGGGUUAGGG-3′
mTERRA	5′-UUAGGGUUAGUGUUAGUGUUAGGG-3′
GC stem-loop	5′-GCGCGCGCGCGCGCGCGCGAGAGCGCGCGCGCGCGCGC-3′
NEAT1	5′-GGGAGGGAGGGAGGG-3′