

**AWARD NUMBER:** W81XWH-22-1-0591

**TITLE:** The RNA-Binding Protein Zfp106 as a Therapeutic Target for Suppression of Amyotrophic Lateral Sclerosis

**PRINCIPAL INVESTIGATOR:** Brian Black, PhD

**CONTRACTING ORGANIZATION:** University of California, San Francisco, CA

**REPORT DATE:** August 2023

**TYPE OF REPORT:** Annual

**PREPARED FOR:** U.S. Army Medical Research and Development Command  
Fort Detrick, Maryland 21702-5012

**DISTRIBUTION STATEMENT:** Approved for Public Release; Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

<b>REPORT DOCUMENTATION PAGE</b>			<i>Form Approved</i> OMB No. 0704-0188		
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
<b>1. REPORT DATE</b> August 2023		<b>2. REPORT TYPE</b> Annual		<b>3. DATES COVERED</b> 15Jul2022-14Jul2023	
<b>4. TITLE AND SUBTITLE</b>  The RNA-Binding Protein Zfp106 as a Therapeutic Target for Suppression of Amyotrophic Lateral Sclerosis			<b>5a. CONTRACT NUMBER</b> W81XWH-22-1-0591		
			<b>5b. GRANT NUMBER</b> AL210129		
			<b>5c. PROGRAM ELEMENT NUMBER</b>		
<b>6. AUTHOR(S)</b> Brian Black, PhD			<b>5d. PROJECT NUMBER</b>		
			<b>5e. TASK NUMBER</b>		
			<b>5f. WORK UNIT NUMBER</b>		
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  The Regents of the University of California, San Francisco 1855 Folsom St, Ste 425, San Francisco, CA 94143-4249			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  1		
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>  U.S. Army Medical Research and Development Command Fort Detrick, Maryland 21702-5012			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>		
			<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>		
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  Approved for Public Release; Distribution Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> Amyotrophic lateral sclerosis (ALS) comprises a heterogeneous group of neurodegenerative disorders, characterized by loss of motor neurons and concomitant denervation of skeletal muscles. All forms of ALS share loss of motor neurons as a common feature. Aberrant RNA metabolism and processing, often resulting in the accumulation of RNA and/or protein aggregates is a major underlying cause of ALS, including the most common familial form of ALS, which is caused by abnormal expansion of GGGGCC repeats in the first intron of the C9orf72 gene. Unfortunately, there are no effective therapies currently available for treating ALS caused by dysregulated RNA metabolism. The recently described zinc finger RNA binding protein Zfp106 suppresses neurotoxicity, cellular death, and neurodegeneration caused by expansion of GGGGCC repeats through direct binding to the repeats. Therefore, the overall objective of this proposal is to leverage the ability of Zfp06 to suppress neurodegeneration by defining mechanisms for increasing the expression of Zfp106 in vivo. In a first specific aim, we sought to identify determinants within Zfp106 responsible for RNA-binding and suppression of RNA aggregation, RAN translation, and neurotoxicity. Towards this aim, we identified a small, ~ 45kD C-terminal fragment of Zfp106 that is sufficient for RNA binding. We also determined that co-expression of Zfp106 in human cells suppressed the presence of toxic RNA aggregates and the repeat-associated non-ATG-dependent (RAN) translation of dipeptide repeat proteins. These seminal findings have been favorably reviewed at PNAS and are currently published on bioRxiv (Celona et al., Zfp106 binds to G-quadruplex RNAs and inhibits RAN translation and formation of RNA foci caused by G4C2 repeats. Proceedings of the National Academy of Sciences USA, in revision for invited resubmission; bioRxiv, 2023.03.05.531055; doi: 10.1101/2023.1103.1105.531055) In a second specific aim, we will define cis-regulatory elements (enhancers) responsible for motor neuron expression of human ZNF106 using in vivo transgenic techniques in mice. Towards this goal, we made significant progress in cloning elements from ZNF106, successfully cloning 4/6 elements to be tested. In the next project period, the remaining elements will be cloned, and all will be tested in vivo by mouse transgenesis.					
<b>15. SUBJECT TERMS</b> None listed.					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>USAMRDC</b>
Unclassified	Unclassified	Unclassified	Unclassified	43	<b>19b. TELEPHONE NUMBER (include area code)</b>

## TABLE OF CONTENTS

	<u>Page</u>
1. Introduction	4
2. Keywords	4
3. Accomplishments	4
4. Impact	9
5. Changes/Problems	10
6. Products	11
7. Participants & Other Collaborating Organizations	13
8. Special Reporting Requirements	16
9. Appendices	17

## 1. INTRODUCTION:

Zfp106 is an endogenous, natural suppressor of ALS-like neurodegeneration caused by dysregulated RNA metabolism. The ultimate goal of the work proposed here is to develop adeno-associated virus (AAV)-based therapeutics to suppress the neurodegenerative effects of ALS by increasing expression of Zfp106, an innate inhibitor of neurodegeneration. In order to develop this therapeutic idea and determine if Zfp106 is a viable target for pre-clinical studies using the basic therapeutic strategies highlighted above, we will take a two-pronged strategy. First, we will define the minimum fragment of Zfp106 necessary to suppress neurodegeneration (**Specific Aim 1**). In parallel, we will also identify motor neuron-specific human *ZNF106* enhancers that can be activated to increase expression in the context of ALS (**Specific Aim 2**). These studies will establish the proof-of-concept for Zfp106/ZNF106 as a viable therapeutic target by defining the minimal region of Zfp106 that can ameliorate ALS and that is small enough to be packaged into an AAV vector and by identifying and characterizing *ZNF106* motor neuron enhancers that can be targeted by AAV-delivered CRISPRa. We believe this will have broad application for ameliorating many forms of ALS caused by dysregulation of RNA processing and metabolism.

## 2. KEYWORDS:

amyotrophic lateral sclerosis, ALS, ZNF106, Zfp106, G-quadruplex, RNA-binding, *C9orf72*, neurodegeneration, zinc finger, suppression, RNA foci, neuroprotection

## 3. ACCOMPLISHMENTS:

**What were the major goals of the project?**

**The major goals of this project are:**

**(Major Task 1) Identify regions of the Zfp106 protein responsible for binding to GGGGCC-containing RNA repeats and other G-quadruplex sequences in electrophoretic mobility shift assays (EMSA) in vitro.** This major task was expected to take 12 months (7/15/23) and was completed on time (6/15/23).

**(Major Task 2) Determine regions of the Zfp106 protein responsible for disrupting the formation of RNA foci in human cells in culture using an established system.** This major task was expected to take 12-18 months (7/15/23-1/15/24) and was completed ahead of schedule (6/15/23).

**(Major Task 3) Determine regions of Zfp106 responsible for suppressing repeat-associated non-ATG (RAN) translation in cultured neuronal cells.** This major task was expected to take 12-18 months (7/15/23-1/15/24) and was completed ahead of schedule (6/15/23).

**(Major Task 4) Identify transcriptional enhancers from the human *ZNF106* locus with activity in transgenic mice.** This work was expected to take 16 months (11/15/23) and is approximately 50% complete. We expect to complete this work by 7/14/24.

**(Major Task 5) Characterize the detailed temporal and spatial pattern of human *ZNF106* motor neuron enhancers in transgenic mice.** This work was expected to take 24 months (7/14/24). This work is currently 0% complete as it is dependent on completion of Major Task 4. We are anticipating on time completion by 7/14/24.

**What was accomplished under these goals?**

During this reporting period, we completed an important paper (Celona *et al.*, **Zfp106 binds to G-quadruplex RNAs and inhibits RAN translation and formation of RNA foci caused by G4C2 repeats. *Proceedings of the National Academy of Sciences USA*, in revision for invited resubmission. *bioRxiv*, 2023.03.05.531055; doi: 10.1101/2023.1103.1105.531055) that reports several of the major activities accomplished during this reporting period and represents completion of several of the major tasks and goals of this grant. A copy of this preprint is provided in the appendix.**

During this reporting period, we conducted electrophoretic mobility shift assays (EMSA) and circular dichroism (CD) experiments that established that Zfp106 binds to GGGGCC repeats, which is the causative lesion in *C9orf72* ALS, and other G-quadruplex-forming sequences. Importantly, we established that the C-terminal zinc fingers and WD40 motif containing region is necessary and sufficient for the binding activity (see attached manuscript preprint and **Fig. 1**). Importantly, the cDNA that encodes this small region is sufficiently small to package into an adeno-associated virus vector.

We also performed transfection experiments in human U-2OS cells with Zfp106 expression constructs. The U2-OS cells stably express GGGGCC repeats or other control RNAs. In these experiments, we showed that Zfp106 suppresses the formation of toxic RNA foci that are formed when GGGGCC repeats are expressed, such as they are in the context of *C9orf72* ALS (see attached manuscript preprint and **Fig. 2**). This is a major finding as it shows that *Zfp106* expression is sufficient to suppress one of the major pathological features of *C9orf72* ALS.



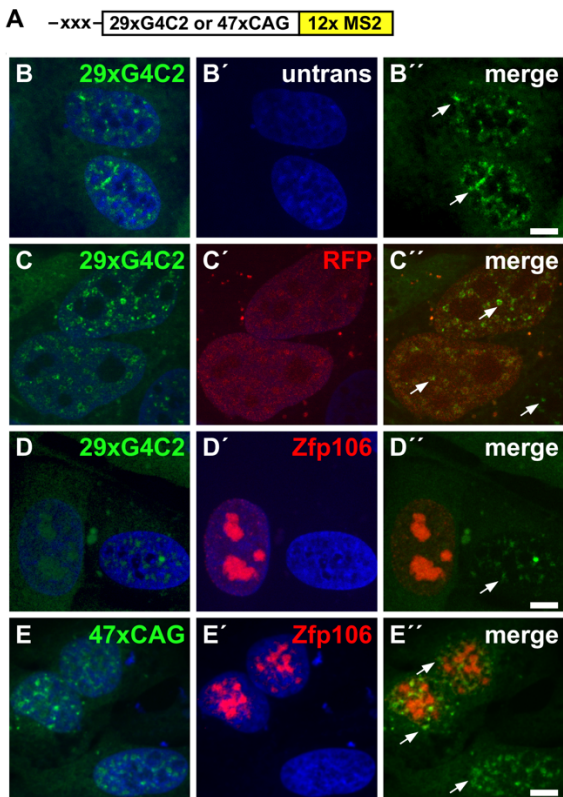


Fig. 2. Representative fluorescence images of U-2OS cells, stably expressing either 29× GGGGCC (B,C,D) or 47× CAG repeats (E) fused to 12× MS2 RNA stem-loops (schematic shown in A), that are untransfected (B), or have been transfected with nuclear RFP (C) or RFP-Zfp106 (D,E). Cell nuclei were counterstained 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.

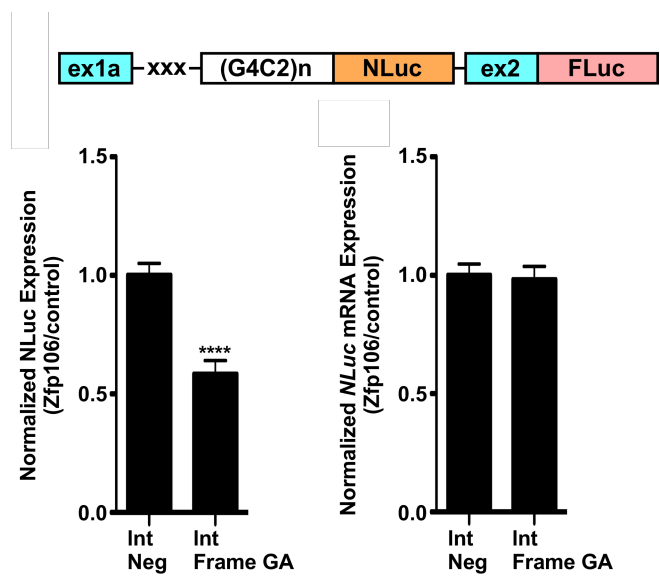


Fig. 3. NLuc luminescence (left) and mRNA levels (right) in cells co-transfected with bicistronic splicing dual-luciferase reporters with *C9orf72* exons and intron (Int Neg, no repeats; Int Frame GA, 70× GGGGCC repeats) and Zfp106 or control vector. NLuc luminescence or mRNA expression was normalized 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. Data were analyzed by two-way ANOVA and Bonferroni's Multiple Comparison Test; \*\*\*\* p<0.0001; xxx indicates stop codons.

**What opportunities for training and professional development has the project provided?**

Two senior level postdocs (Specialists) in the lab, Drs. Barbara Celona and Pauline Colombier, worked on this project during the reporting period. Each received active one-on-one mentorship and training from the Principal Investigator, Dr. Brian Black. Dr. Celona had the opportunity to present her findings at several round table discussions, and both Drs. Celona and Colombier presented their work at the UCSF Cardiovascular research Institute's annual scientific retreat. Dr. Celona wrote the first draft of the manuscript that describes many of the findings reported here (Celona *et al.*, Zfp106 binds to G-quadruplex RNAs and inhibits RAN translation and formation of RNA foci caused by G4C2 repeats. *Proceedings of the National Academy of Sciences USA*, in revision for invited resubmission. *bioRxiv*, 2023.03.05.531055; doi: 10.1101/2023.1103.1105.531055), offering her intense training in manuscript writing and preparation. Similarly, Dr. Colombier has begun work on two manuscripts, both currently in preparation, offering her similar training in manuscript writing and preparation. These training activities will continue into the next reporting period.

**How were the results disseminated to communities of interest?**

Nothing to Report.

**What do you plan to do during the next reporting period to accomplish the goals?**

During the next reporting period, we will continue with the stated goals as originally proposed. We will finalize publication of our work (Celona *et al.*, Zfp106 binds to G-quadruplex RNAs and inhibits RAN translation and formation of RNA foci caused by G4C2 repeats. *Proceedings of the National Academy of Sciences USA*, in revision for invited resubmission. *bioRxiv*, 2023.03.05.531055; doi: 10.1101/2023.1103.1105.531055), making any (unanticipated at this point) revisions requested by reviewers. We will continue our efforts to clone all human *ZNF106* putative enhancer elements from the genome in preparation for generation and analysis of transgenic mice in order to determine neuron-specific regulatory elements from human *ZNF106*. We will complete the approvals process from the UCSF IACUC and from DoD (ACURO) and then only after those approvals are in place,

we will generate and analyze *ZNF106-lacZ* transgenic mice. We anticipate publication of a second manuscript near the end of the final reporting period.

#### 4. IMPACT:

**What was the impact on the development of the principal discipline(s) of the project?**

Our work under this project to date will have a major impact on the neurodegeneration field. First, this work establishes *ZNF106* as an important regulator of RNA G-quadruplexes and establishes it as a putative therapeutic target for treatment of neurodegeneration, particularly for *C9orf72* ALS/FTD. Our work under this project to date also provides mechanistic information about how *Zfp106* functions in RNA processing and how that processing when defective is central to neurodegenerative diseases through formation of RNA aggregates and aberrant protein synthesis.

**What was the impact on other disciplines?**

Nothing to Report.

**What was the impact on technology transfer?**

Nothing to Report.

**What was the impact on society beyond science and technology?**

Nothing to Report.

**5. CHANGES/PROBLEMS:**

**Changes in approach and reasons for change**

Nothing to Report.

**Actual or anticipated problems or delays and actions or plans to resolve them**

We have experienced delays with IACUC protocol submission and approval, and as a result this has resulted in delayed ACURO submission. The delays were due to refining our proposed use of mice and development of a more extensive approval overall in order to encompass a wider range of studies than those simply proposed here. Notably, this has resulted in **no delay in the work overall**, as we've focused our efforts on the biochemical and cell culture-based studies and on our cloning efforts. This delay will be resolved imminently. The IACUC protocol has been submitted and conditionally approved and the ACURO protocol will be submitted imminently. No animal work will commence until ACURO approval, but we are not anticipating any delays in the overall progress of the project since we're not yet quite ready to begin animal studies as we are still focused on enhancer cloning.

**Changes that had a significant impact on expenditures**

Nothing to Report.

**Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

**Significant changes in use or care of human subjects**

Nothing to Report.

**Significant changes in use or care of vertebrate animals**

Nothing to Report. No animal studies have commenced as we are not ready for that work yet. Once ACURO approval has been granted and enhancer cloning is complete, animals studies will begin. We do not anticipate any changes to their proposed use.

**Significant changes in use of biohazards and/or select agents**

Nothing to Report.

**6. PRODUCTS:**

- **Publications, conference papers, and presentations**

**Journal publications.**

Celona B, Wu H, Dang B, Kratochvil HT, DeGrado WF, **Black BL. 2023.** Zfp106 binds to G-quadruplex RNAs and inhibits RAN translation and formation of RNA foci caused by G4C2 repeats. *Proceedings of the National Academy of Sciences USA*, in revision for invited resubmission. *bioRxiv*, 2023.03.05.531055; doi: 10.1101/2023.1103.1105.531055. Acknowledgement of Federal (DoD) support.

**Books or other non-periodical, one-time publications.**

Nothing to Report.

**Other publications, conference papers and presentations.**

Nothing to Report.

- **Website(s) or other Internet site(s)**

Nothing to Report.

- **Technologies or techniques**

Nothing to Report.

- **Inventions, patent applications, and/or licenses**

Nothing to Report.

- **Other Products**

Nothing to Report.

## **7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS**

**What individuals have worked on the project?**

Example:

<i>Name:</i>	<i>Brian Black</i>
<i>Project Role:</i>	<i>Principal Investigator</i>
<i>Researcher Identifier (e.g. ORCID ID):</i>	<i>0000-0002-6664-8913</i>
<i>Nearest person month worked:</i>	<i>1.8</i>
<i>Contribution to Project:</i>	<i>Dr. Black oversaw all aspects of the project and helped write the recently submitted paper describing these findings.</i>
<i>Name:</i>	<i>Barbara Celona</i>
<i>Project Role:</i>	<i>Specialist</i>
<i>Researcher Identifier (e.g. ORCID ID):</i>	<i>0000-0001-5391-8155</i>
<i>Nearest person month worked:</i>	<i>6</i>
<i>Contribution to Project:</i>	<i>Dr. Celona participated in all aspects of the project, including performing EMSA and cell culture experiments and writing the first draft of the paper that reported these findings.</i>
<i>Name:</i>	<i>Pauline Colombier</i>
<i>Project Role:</i>	<i>Specialist</i>
<i>Researcher Identifier (e.g. ORCID ID):</i>	<i>0000-0001-6424-2525</i>
<i>Nearest person month worked:</i>	<i>6</i>
<i>Contribution to Project:</i>	<i>Dr. Colombier participated in enhancer cloning and transfection experiments and she also played a major intellectual role in the project, consulting on the written manuscript.</i>
<i>Name:</i>	<i>Onyekachim Okpo</i>
<i>Project Role:</i>	<i>Research technician</i>
<i>Researcher Identifier (e.g. ORCID ID):</i>	<i>n/a</i>
<i>Nearest person month worked:</i>	<i>2.4</i>
<i>Contribution to Project:</i>	<i>Ms. Okpo provided technical support for experiments performed by Drs. Celona and Colombier and performed additional molecular biology experiments under the direction of Dr. Black.</i>

**Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Dr. Black's previous active support from NIDDK (R01DK119621) has ended. There was no overlap. Dr. Black's previous active support from the Chan Zuckerberg Initiative has ended. There was no overlap. Dr. Black's previous pending application for support from NHLBI (HL165231) was not funded (no overlap). Dr. Black has received new support (salary only, 5% effort; no overlap) on Dr. Zian Tseng's R01 HL157247, *Clonal Hematopoiesis of Indeterminate Potential and Risk of Autopsy-defined Sudden Cardiac Death*. Dr. Black has received new support (salary only, 5% effort; no overlap) on Dr. Edward Gerstenfeld's R01 HL159069, *Elucidating Mechanistic Relationships Between Atrial Ectopy, Atrial Remodeling and Atrial Fibrillation*. Dr. Black has pending support (salary only, 2.5% effort; no overlap) on Dr. Vasanth Vedantham's R01 submitted to NHLBI, *Transcriptional Regulation of Sinoatrial Node Development*.

**What other organizations were involved as partners?**

Nothing to Report.

## 8. SPECIAL REPORTING REQUIREMENTS

**COLLABORATIVE AWARDS:**

**QUAD CHARTS:**

## 9. APPENDICES:

See attached manuscript preprint:

Celona *et al.*, Zfp106 binds to G-quadruplex RNAs and inhibits RAN translation and formation of RNA foci caused by G4C2 repeats. *Proceedings of the National Academy of Sciences USA*, in revision for invited resubmission. *bioRxiv*, 2023.03.05.531055; doi: 10.1101/2023.1103.1105.531055

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

Barbara Celona<sup>1</sup>, Haifan Wu<sup>1,2,†</sup>, Bobo Dang<sup>1,2,†</sup>, Huong T. Kratochvil<sup>1,2</sup>, William F. DeGrado<sup>1,2</sup>, and Brian L. Black<sup>1,3,\*</sup>

<sup>1</sup>Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA, 94143

<sup>2</sup>Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, 94143

<sup>3</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, 94143

†Present addresses: HW, Wichita State University, Wichita, KS, USA; BD, School of Life Sciences, Westlake University, and Westlake Institute for Advanced Study, Hangzhou, China

\*Correspondence:

Brian L. Black  
555 Mission Bay Blvd. S, MC 3120  
San Francisco, CA 94143-3120

brian.black@ucsf.edu

Keywords: *C9orf72*; amyotrophic lateral sclerosis; RAN translation; G-quadruplex; RNA-binding protein

## ABSTRACT

Expansion of intronic GGGGCC repeats in the *C9orf72* gene causes amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). 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 via a previously unknown mechanism. Here, we show that Zfp106 inhibits formation of RNA foci and significantly reduces RAN translation caused by GGGGCC-repeats in mammalian 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. These data suggest that Zfp106 suppresses GGGGCC repeat-mediated cytotoxicity through alteration of the repeat's G-quadruplex structure.

## INTRODUCTION

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 4-6 weeks of age followed by rapidly progressing motor neuron degeneration and death by 16 weeks (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 recent studies suggest that

pathogenesis may be the result of multiple non-mutually 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 established that Zfp106 binds directly and specifically to GGGGCC RNA repeats, and that co-expression 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 and suppresses GGGGCC-induced RAN translation in mammalian 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 co-transfected a Zfp106 expression plasmid with a pCMV-(GGGGCC)<sub>30</sub>-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. 1A). 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. 1A',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. 1B). These data suggest that Zfp106 inhibited translation of the eGFP mRNA specifically containing 30× GGGGCC repeats.

An important distinction between the expression of the GGGGCC repeats from the pCMV-(GGGGCC)<sub>30</sub>-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)<sub>30</sub>-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 if Zfp106 co-expression also inhibits RAN translation, we co-transfected 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. 1C, Fig. S1A) (45). As an internal control, the reporter construct produces

firefly luciferase (FLuc) by canonical AUG-dependent initiation (45). Co-expression 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. 1C'). Inhibition occurred in a Zfp106 dose-dependent manner (Fig. S1A'). Steady-state NLuc mRNA levels were unaffected by co-expression of Zfp106 (Fig. 1C''). 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 (Fig. S1B). Importantly, these data demonstrate that Zfp106 suppresses translation, including RAN translation, of GGGGCC-repeat-containing RNAs. Moreover, 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 neuroprotective effect of Zfp106 *in vivo*.

### **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 (46). Therefore, we examined the influence of Zfp106 co-expression on the formation of RNA foci induced by overexpression of GGGGCC-repeat-containing RNAs (Fig. 2). Jain and Vale (2017) described a system in which 29× GGGGCC repeats tagged with 12× MS2-hairpin loops can be visualized via co-expression of YFP-tagged MS2-coat binding protein, which binds to the MS2 hairpin loops (Fig. 2A) (38). Consistent with previously published results, induction of 29× GGGGCC RNA repeats in this system resulted in formation of numerous nuclear foci (Fig. 2B) (38). Remarkably, the presence of foci was substantially reduced by co-expression of Zfp106 compared to nuclear-RFP control transfected cells (Fig. 2C,D,F). In contrast, Zfp106 had no effect on the formation of RNA foci induced by 47× CAG repeats (Fig. 2E,G). As an alternate method for detecting foci, we also examined the number of foci

per nucleus by using RNA FISH 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 (Fig. S2).

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 central to 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 suggests that it plays roles in cellular RNA processing and that its RNA-binding activity is unlikely to be restricted to only 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 Zfp106 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. 3A). Zfp106 bound to all sequences predicted to form G-quadruplexes, while it failed to bind significantly to all non-G-quadruplex-forming sequences (Fig. 3A). Indeed, Zfp106 bound efficiently to GGGGCC repeats under conditions that

promote G-quadruplex formation (Fig. 3A). 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. 3A). Likewise, Zfp106 showed little to no detectable binding to a GC stem-loop sequence (Fig. 3A'') or to CAG repeats (Fig. 3B), consistent with the observations that Zfp106 did not suppress translation or formation of foci from transcripts containing CAG repeats (Fig. 1B, Fig. 2E,G).

The full length isoform of Zfp106 used in these studies is ~150kD, 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 S3A) (3, 47). 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)<sub>4</sub> repeats (Fig. 3C). 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, showed no detectable binding to (GGGGCC)<sub>4</sub> repeats (Fig. S3A,B).

GGGGCC RNA repeats form parallel G-quadruplexes (12). In addition to binding to (GGGGCC)<sub>4</sub> and (GGGGCC)<sub>3</sub> repeats, Zfp106 also bound efficiently to other parallel G-quadruplex-forming RNA sequences, including sequences in TERRA and NEAT1 (Fig. 3A',A'') (48, 49). Mutation of TERRA repeats with G>U substitutions, which impede G-quadruplex formation, prevented Zfp106 binding (Fig. 3A'). Interestingly, the predicted anti-parallel G-quadruplex-forming sequence from ORN-N (50) was also bound by Zfp106 (Fig. 3A'). 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 (51, 52). 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 (53).

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)<sub>4</sub> EMSA increased Zfp106 binding, whereas addition of LiCl to Zfp106-(GGGGCC)<sub>4</sub> EMSA decreased Zfp106 binding (Fig. 3D,D'). 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 (54, 55). Similarly, addition of increasing concentration of TMPyP4, a cationic porphyrin derivative that binds and distorts the G-quadruplex structure of GGGGCC repeats (56, 57), decreased Zfp106 binding to (GGGGCC)<sub>4</sub> repeats (Fig. 3D'').

### **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, 58). Therefore, we reasoned that the disruption of RNA foci and suppression of neurotoxicity caused by GGGGCC repeats in cells and flies might be due to 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 circular dichroism (CD). CD spectroscopy confirmed that the (GGGGCC)<sub>4</sub> RNA molecules formed the expected parallel G-quadruplex structure, with a spectral minimum at 236 nm and a spectral maximum at 264 nm (Fig. 4A). The CD spectra for Zfp106[WD40+ZnF] alone showed negative peaks in the 210-220nm range

but gave minimal to no signal at 264 nm (Fig. 4C), allowing us to monitor the G-quadruplex structure of (GGGGCC)<sub>4</sub> in the presence and absence of Zfp106 during temperature unfolding (Fig. 4A,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)<sub>4</sub> 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 250-300 nm range (Fig. 4A',B'). Closer examination of CD spectra in the 250-300 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 G-quadruplex conformation (Fig. 4 A').

Interestingly, this shift in the peak from 264 nm to 274 nm occurred at lower temperature in the presence of Zfp106 (Fig. 4B'), suggesting that Zfp106 stabilized the G-quadruplex conformation normally favored at higher temperature. 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. 4D). We also conducted essentially identical experiments with full-length Zfp106 at 1:2 molar ratio of protein to RNA, and similar results were obtained (Fig. S4), confirming a change in the (GGGGCC)<sub>4</sub> 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)<sub>4</sub> might form both intermolecular and intramolecular G-quadruplexes in a concentration- and temperature-dependent manner (59). However, when we analyzed (GGGGCC)<sub>4</sub> repeats by analytical ultracentrifugation, we failed to detect intermolecular interactions between (GGGGCC)<sub>4</sub> molecules under the same buffer conditions and RNA concentration used in our CD analysis (data not shown).

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*. Regardless, the studies presented here support the idea that Zfp106's suppressive effect on GGGGCC-mediated cytotoxicity might be through alteration or stabilization of specific G-quadruplex structures formed by GGGGCC repeats. Future studies will dissect how different G-quadruplex conformations contribute to disease in C9orf72 ALS.

### **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, 60). 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 RBP RanGAP1 (61). 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 (62). Future studies will

address the functional importance of Zfp106 binding to G-quadruplex-forming cellular RNAs and if dysregulation of the “G4ome” (62) is a contributing mechanism for neurodegeneration in *C9orf72* ALS/FTD or other neurodegenerative diseases.

## MATERIALS AND METHODS

### Plasmids

GFP reporter constructs with either 3× or 30× GGGGCC repeats inserted in the 5'-UTR of an eGFP cDNA in plasmids pCMV-(GGGGCC)<sub>3</sub>-EGFP and pCMV-(GGGGCC)<sub>30</sub>-EGFP, respectively, have been described previously (44). Plasmid pCMV-(CAG)<sub>6</sub>-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)<sub>60</sub>-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 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 of the full-length protein, 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) 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].

### **Cell culture, transfections, and reporter assays**

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).

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 Multidetector 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 reverse transcriptase (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). 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 Supplemental Table S1. Data were normalized to housekeeping gene *Gapdh* by the  $2^{-\Delta\Delta C_t}$  method (63).

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. 24 h post-transfection, cells were fixed in 4% paraformaldehyde for 10 minutes 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 minutes 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 hybridization 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.

### **Electrophoretic mobility shift assay (EMSA)**

100 pmol of single-stranded RNA (Integrated DNA Technologies, Inc.) were 5'-end labeled with [ $\gamma$ - $^{32}$ 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 (64) 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. Equivalent amount of total protein was used for each EMSA. Parallel purification from HEK293T cells transfected with the 3 $\times$ FLAG-2 $\times$ 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% acrylamide-TBE gels, which were subsequently dried and subjected to autoradiography. For competition experiments, the purified proteins were preincubated for 10 minutes at room temperature with 30-fold molar excess of unlabeled single-stranded RNA competitor. The single-stranded RNA probe sequences are provided in Supplemental Table S2.

### **Circular dichroism (CD) spectroscopy**

CD spectra from 400 to 190 nm were collected at temperatures between 8°C and 98°C, with a 5°C/minute temperature gradient, using a Jasco J-810 equipped with a Jasco Peltier temperature control system. CD spectra were recorded in the presence of 4  $\mu$ M (GGGGCC) $_4$  RNA, either alone or with addition of 4  $\mu$ M Zfp106 C-terminal fragment (Zfp106[WD40+ZnF]) or 2  $\mu$ M 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.

## **ACKNOWLEDGEMENTS**

HTK was supported by NIH grant K99GM138753. BC received support from the UCSF 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 CDMRP to BLB with additional support from NIH grants R35 GM122603 to WFD and DK119621 and HL146366 to BLB. We are grateful to Peng Jin (Emory) and Shuying Sun (Johns Hopkins) for providing plasmids and Ron Vale (Janelia) and Ankur Jain (MIT) for sharing cells and for helpful advice. We thank Bill Seeley, Sarat Vatsavayai, Jen Yokoyama, and members of the Black and DeGrado labs for helpful discussions.

## **AUTHOR CONTRIBUTIONS**

Conceptualization, BC, WFD, and BLB; Methodology, Formal Analysis, and Investigation, BC, HW, BD, HTK; Writing – Original Draft, BC and BLB; Writing – Reviewing and Editing, all authors; Visualization, BC and BLB; Project Administration, BC and BLB; Supervision, BLB and WFD; Funding Acquisition, BLB and WFD. All authors approved the final manuscript. The authors declare no competing interests.

## REFERENCES

1. D. M. Anderson *et al.*, Severe muscle wasting and denervation in mice lacking the RNA-binding protein ZFP106. *Proceedings of the National Academy of Sciences of the United States of America* **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** (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. *Nature communications* **6**, 6674 (2015).
6. L. Tafforeau *et al.*, The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Molecular cell* **51**, 539-551 (2013).
7. P. I. Joyce *et al.*, Deficiency of the zinc finger protein ZFP106 causes motor and sensory neurodegeneration. *Human molecular genetics* **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 FTL/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 TFE3 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** (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)<sub>n</sub> repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *The Journal of biological chemistry* **288**, 9860-9866 (2013).
35. A. Kovanda, M. Zalar, P. Šket, J. Plavec, B. Rogelj, Anti-sense DNA d(GGCCCC)<sub>n</sub> 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** (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 molecular medicine* **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)<sub>n</sub> 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. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 7778-7783 (2013).
45. W. Cheng *et al.*, C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2 $\alpha$  phosphorylation. *Nature communications* **9**, 51 (2018).
46. 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).
47. M. Hatayama, J. Aruga, Characterization of the tandem CWCH2 sequence motif: a hallmark of inter-zinc finger interactions. *BMC Evol Biol* **10**, 53 (2010).
48. M. Lalonde, P. Chartrand, TERRA, a Multifaceted Regulator of Telomerase Activity at Telomeres. *J Mol Biol* **432**, 4232-4243 (2020).
49. T. Yamazaki *et al.*, Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Molecular cell* **70**, 1038-1053.e1037 (2018).

50. 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).
51. A. Bajc Česnik *et al.*, Nuclear RNA foci from C9ORF72 expansion mutation form paraspeckle-like bodies. *J Cell Sci* **132** (2019).
52. 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** (2021).
53. 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).
54. 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).
55. C. Bardin, J. L. Leroy, The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res* **36**, 477-488 (2008).
56. B. Zamiri, K. Reddy, R. B. Macgregor, Jr., C. E. Pearson, TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)<sub>n</sub> repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *The Journal of biological chemistry* **289**, 4653-4659 (2014).
57. 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).
58. J. U. Guo, D. P. Bartel, RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353** (2016).
59. 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).
60. S. Mizielinska, A. M. Isaacs, C9orf72 amyotrophic lateral sclerosis and frontotemporal dementia: gain or loss of function? *Curr Opin Neurol* **27**, 515-523 (2014).
61. K. Zhang *et al.*, The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56-61 (2015).
62. A. Cammas, S. Millevoi, RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res* **45**, 1584-1595 (2017).
63. 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).
64. 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).

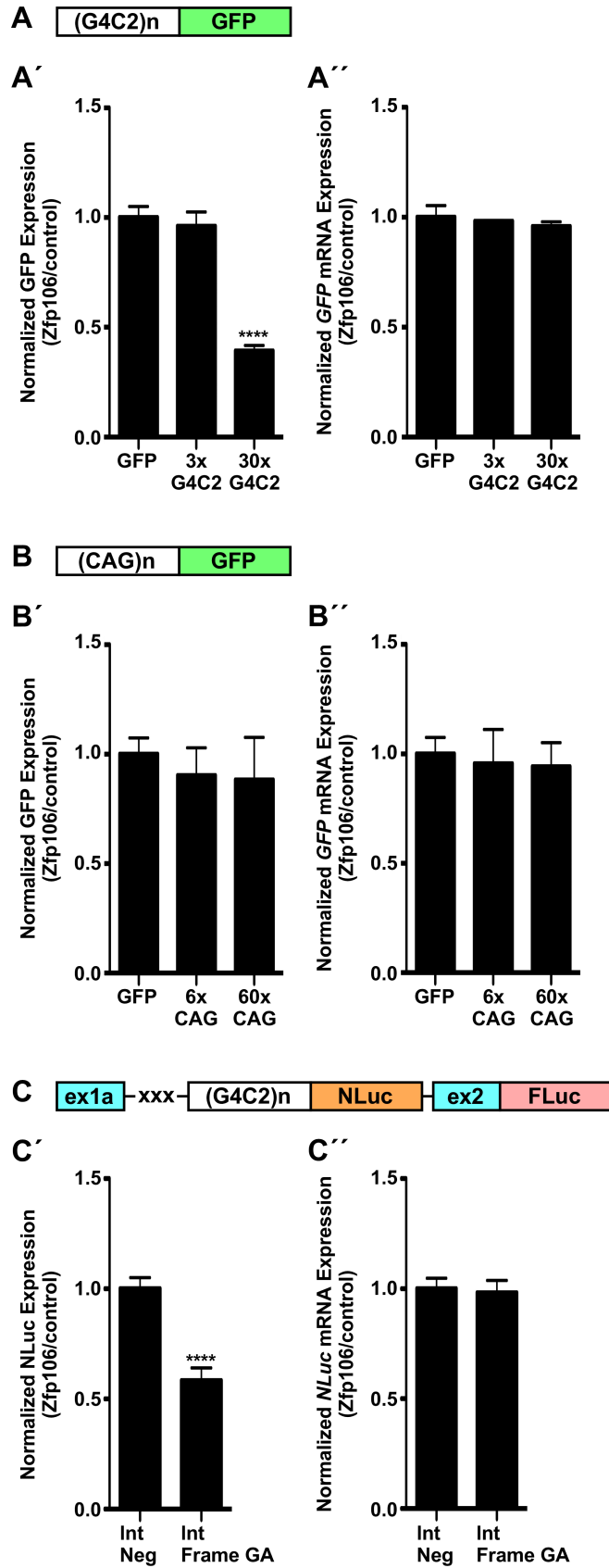
## FIGURE LEGENDS

**Figure 1. Zfp106 suppresses RAN translation.** eGFP fluorescence (A',B') or NLuc luminescence (C') and mRNA levels (A'',B'',C'') in Neuro-2a cells (A,B) or HEK293T cells (C) co-transfected with pCMV-(GGGGCC)<sub>n</sub>-EGFP constructs (no-repeats, 3×, or 30× GGGGCC repeats) and Zfp106 or control vector (A); co-transfected with pCMV-(CAG)<sub>n</sub>-EGFP-containing constructs (no-repeats, 6×, or 60× CAG repeats) and Zfp106 or control vector (B); or co-transfected 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,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. Data were analyzed by two-way ANOVA and Bonferroni's Multiple Comparison Test; \*\*\*\* p<0.0001. Schematic depictions of transfected constructs are shown in (A,B,C); xxx indicates stop codons.

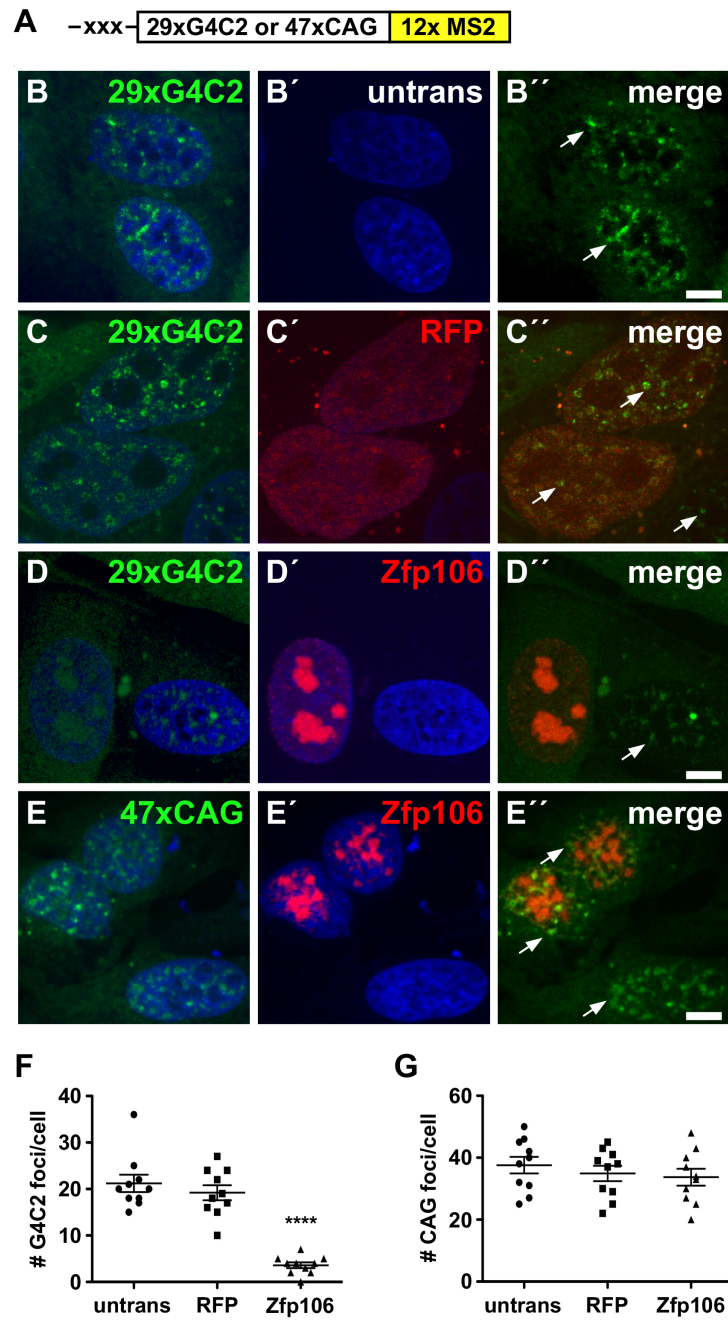
**Figure 2. Zfp106 suppresses formation of GGGGCC-containing RNA foci.** Representative fluorescence images of U-2OS cells, stably expressing either 29× GGGGCC (B,C,D) or 47× CAG repeats (E) fused to 12× MS2 RNA stem-loops (schematic shown in A), that are untransfected (B), or have been transfected with nuclear RFP (C) or RFP-Zfp106 (D,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. Zfp106 co-expression 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.

**Figure 3. Zfp106 binds G-quadruplex RNAs.** (A-A'') RNA EMSAs performed with purified Zfp106 protein demonstrate that Zfp106 directly binds G-quadruplex forming RNAs *in vitro*: GGGGCC [(G4C2)<sub>3</sub>, (G4C2)<sub>4</sub>] repeats (A), ORN and TERRA (A') and NEAT1 (A''). Zfp106 showed no binding to the following non-G-quadruplex-forming RNAs: antisense GGCCCC repeats (G2C4)<sub>4</sub>, (CUG)<sub>8</sub>, mutated GGGGCC [(MUT)<sub>4</sub>], (A4C2)<sub>3</sub>, mutated TERRA (mTERRA), and GC-rich stem-loop (GC-SL). (B) RNA EMSA using purified Zfp106 protein demonstrates that Zfp106 does not bind (CAG)<sub>8</sub> repeats under conditions in which the RNA binding protein MBNL1 efficiently binds (CAG)<sub>8</sub>. MBNL1 binding to (CAG)<sub>8</sub> is specific since binding is competed by 30× molar excess of self-competitor but not 30× molar excess of (GGGGCC)<sub>4</sub>. (C) RNA EMSA using purified Zfp106 C-terminus [WD40+ZnF] demonstrates that [WD40+ZnF] is sufficient to bind to (GGGGCC)<sub>4</sub> repeats *in vitro*. 30× molar excess of unlabeled self-competitor, but not 30× molar excess of unlabeled (AAAACC)<sub>4</sub> or (MUT)<sub>4</sub>, efficiently competed with [WD40+ZnF] binding to (GGGGCC)<sub>4</sub>, establishing specificity of the interaction. (D-D'') RNA EMSA using purified Zfp106 protein shows increased binding of Zfp106 to (GGGGCC)<sub>4</sub> with increasing concentration of KCl (3 – 30 – 100 mM) (C) and decreased binding with increasing concentrations of LiCl (3 – 30 – 100 mM) and TMPyP4 (10 – 25 – 50 μM).

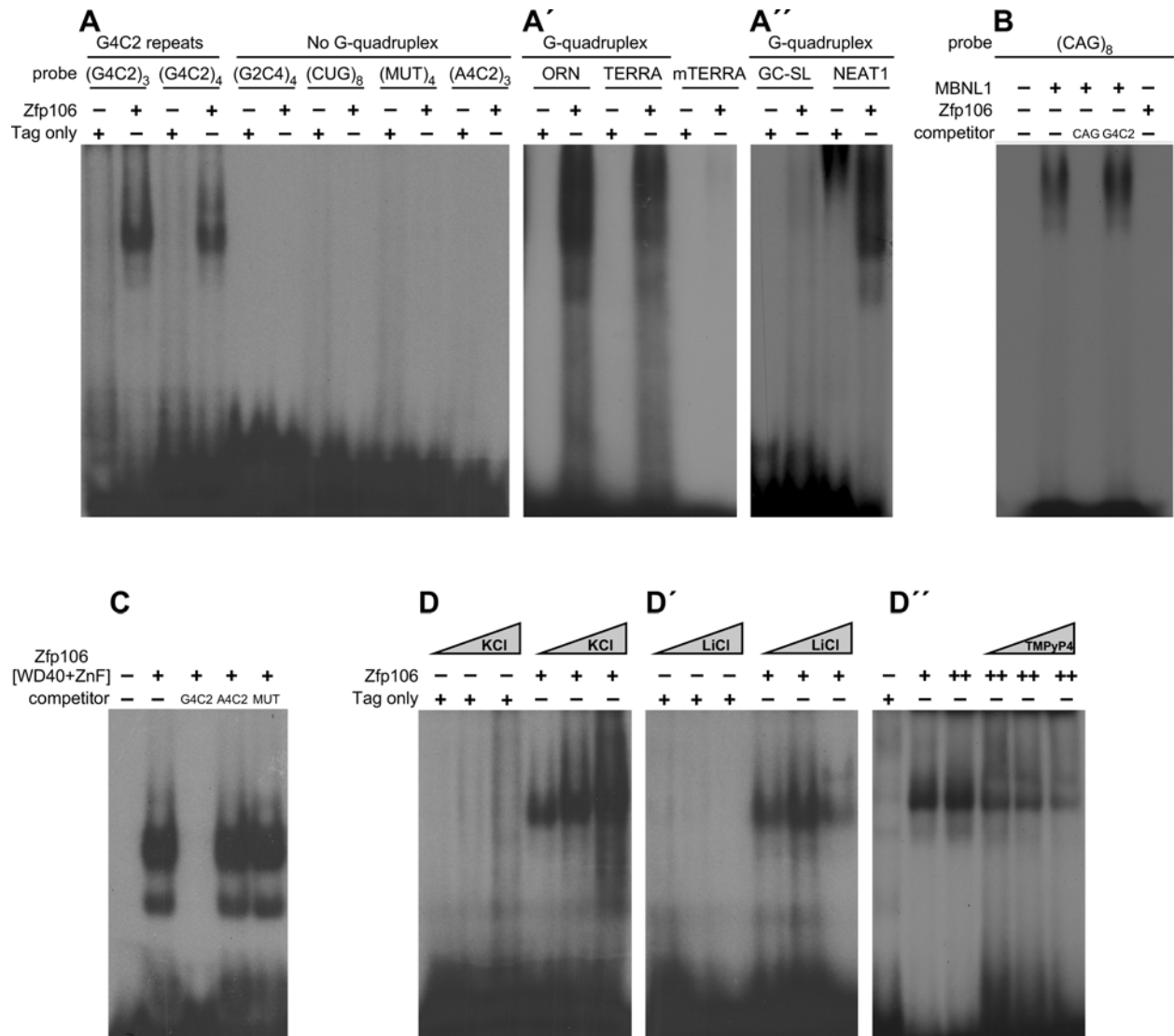
**Figure 4. Zfp106 [WD40+ZnF] alters the G-quadruplex structure of r(GGGGCC)<sub>4</sub>.** (A,A') CD spectra for (GGGGCC)<sub>4</sub> RNA showing a characteristic parallel G-quadruplex structure with a minimum at 236 nm and a maximum at 264 nm. (B,B') CD spectra for r(GGGGCC)<sub>4</sub> plus Zfp106 C-terminus [WD40+ZnF]. (C) CD spectra for Zfp106 C-terminus [WD40+ZnF] alone. Zfp106 C-terminus binding to (GGGGCC)<sub>4</sub> 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 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 temperature (red squares) than in control (black circles).



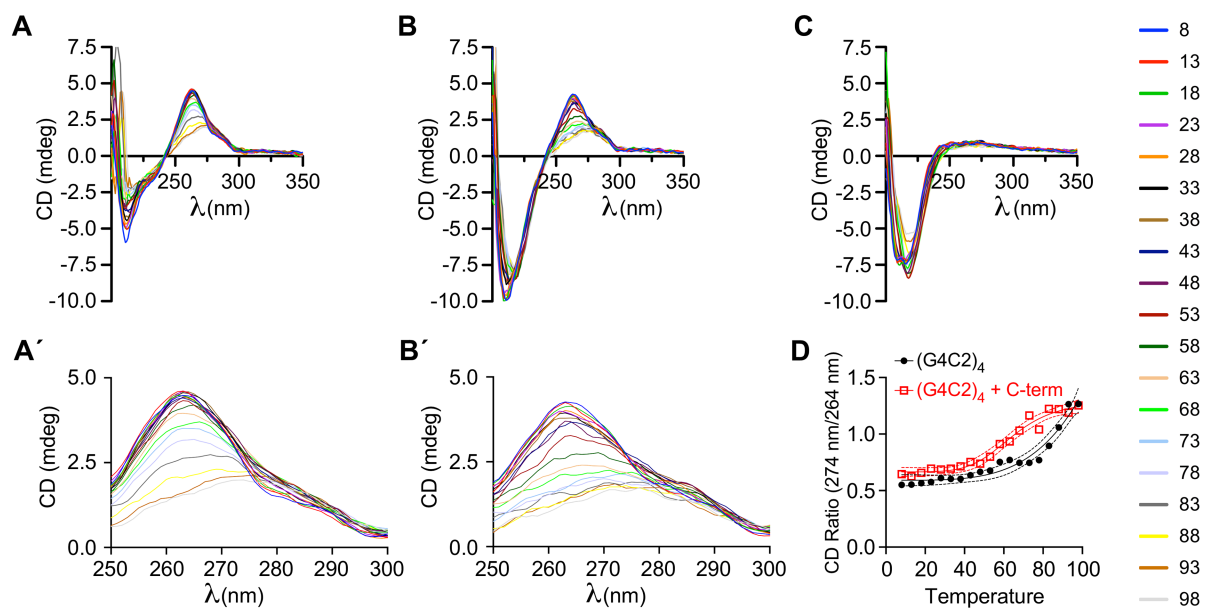
Celona et al., Figure 1



Celona et al., Figure 2



Celona et al., Figure 3



Celona et al., Figure 4