Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins and are neurotoxic

Youn-Bok Lee[1], Han-Jou Chen[1], João N. Peres[2], Jorge Gomez-Deza[1], Jan Attig[5,6], Maja Štalekar[3], Claire Troakes[1], Agnes L. Nishimura[1], Emma L Scotter[1], Caroline Vance[1], Yoshitsugu Adachi[4], Valentina Sardone[1,7], Jack W Miller[1], Bradley N Smith[1], Jean-Marc Gallo[1], Jernej Ule[6], Frank Hirth[4], Boris Rogelj[3], Corinne Houart[2], Christopher E Shaw[1]

Supplemental Experimental Procedures

G4C2 cloning

In the first step, eight G4C2 DNA repeat oligonucleotides were annealed to generate short, double stranded, sticky end oligomers (5' PHO-GGG GCC GGG GCC GGG GCC GGG GCC GGG GCC GGG GCC GGG GCC 3', 5' PHO-CCG GCC CCG GCC CCG GCC CCG GCC CCG GCC CCG GCC CCG GCC 3'). These were incubated overnight (95 ºC to 25 ºC). G4C2 DNA extension was performed using DNA ligase for 2 hours; in this process G4C2 DNA double strands were stretched. To insert the G4C2 DNA product into the shuttle plasmid (pcDNA-Gateway directional TOPO expression vector, Invitrogen, kit K2440-20), we added a 5' adaptor (CAC CTC TAG A, PHO-CCT CTA GAG GTG) for 30 min in which we included both a CACC motif for Topo cloning and an XbaI restriction site (Figure S1A). We also designed a 3' end adaptor (TCT AGA, PHO-GGT CTA GA) with an XbaI restriction site. In the final step of ligation, the 3' adaptor was added and the mixture incubated for a further 30 min. These mixes were separated on a 1.2% agarose gel. Figure S1B shows band smears of increasing molecular weight, which corresponded to increasing numbers of G4C2 repeats. Bands were cut from the gel at the locations indicated (Figure S1B) and the DNA was extracted (QIAquick Gel extraction kit) and subsequently processed for Topo cloning following the manufacturer's instructions (Invitrogen: K2440-20). Colonies were picked and the size of the insert determined by digestion with XbaI. Integrity of G4C2 repeats was confirmed by DNA sequencing (Source-bioscience DNA sequencing, Cambridge), which revealed uninterrupted G4C2 repeats for 8x and 38x, whereas 72x shows a C to A reversion at position 13.

Cell cult*ure*

Human cell line SH-SY5Y cells were cultured in Dulbecco's Modified Eagle Medium (DMEM), F-12 medium supplemented with 10% fetal bovine serum (FBS), 2 mM L-

glutamine, penicillin-streptomycin 50 U/ml and 50 μg/ml respectively. Plasmid DNA transfections were performed in SH-SY5Y cells (1 x 10^6 cells, 24 well plate) by using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions.

RNA pull-down assay

G4C2 RNAs were transcribed *in vitro* and treated with DNase I and precipitated with lithium chloride (Ambion: Am1330). The RNA pull-down assay was performed as described previously (Tsai et al., 2010) with the following modifications. Three micrograms of biotinylated RNA was heated to 90 °C for 2 min in RNA structure buffer (10 mM Tris pH 7, 0.1 M KCl, 10 mM $MgCl_2$), and then incubated at room temperature for 20 minutes. Cells ($1x10^7$) were resuspended in 2 ml PBS, 2 ml nuclear isolation buffer (1.28 M sucrose, 40 mM Tris-HCl pH 7.5; 20 mM $MgCl_2$; 4% Triton X-100), and 6 ml water on ice for 20 min. Nuclei were pelleted by centrifugation at 2,500 x g for 15 min. The pellet was resuspended in 1 ml RNA immunoprecipitation (RIP) buffer (150 mM KCl, 25 mM Tris pH 7.4, 0.5 mM DTT, 0.5 % NP40, 1 mM PMSF and protease inhibitors (Roche complete protease inhibitor) and centrifuged at 13,000 rpm for 10 min. 100 ng of biotinylated RNA was then mixed with 50 μg of nuclear extract in RIP buffer and incubated at RT for 30 min. Then 50 μL of pre-cleared lysate was added to each binding reaction and further incubated at RT for a further 30 min. Beads were washed five times and then boiled in SDS buffer.

Fluoroscence-activated cell sorting (FACS)

SH-SY5Y cells were trypsinized, washed with ice-cold PBS, and stained using APC-Annexin V (Cat 550475, BD Biosciences) according to the manufacturer's protocol. The cells were then analysed with a FACS flow cytometer (Canto II, BD Biosciences).

Cell counting

The counting was carried out in the same manner as described previously (Chen et al., 2010; Mitchell et al., 2010) for EGFP vs G4C2 foci counting, the coverslips were systematically analyzed and all the foci positive cells were counted until a total of 250 EGFP positive cells was achieved. For active caspase-3 counting, all the active caspase-3/foci positive cells were counted until we reach a total of 250 foci positive cells. The background of active caspase-3 was estimated by counting the frequency of active caspase-3 positive but foci negative cells on the same coverslip.

Nothernblot

A digoxigenin (DIG) labeled probe was generated by PCR from 100 pg of EGFP plasmid using forward (5'-GTGCAGTGCTTCAGCCGCTA-3') and reverse primers (5'-CTGCTTGTCGGCCATGATAT-3') (Roche). 5 μg of RNA extracted from G4C2 plasmid transfected SH-SY5Y cells was separated on a 0.8% formaldehyde gel in MOPS buffer (200 mM MOPS, 50 mM sodium Acetate, 10 mM EDTA). RNA was transferred onto a nylon membrane by capillary blotting overnight (Hybond N+, GE Healthcare). The membrane was cross-linked with UV for 1 min and prehybridized in 20 ml of DIG EasyHyb solution (Roche) at 65 °C for 1 hour. Hybridization was performed at 48 °C overnight in a hybridization chamber. The membrane was then washed in 50 ml of first wash buffer (2xSSC, 0.1% SDS) at room temperature for 5 min and then washed two times in 50 ml of second wash buffer (0.1xSSC, 0.1% SDS) at 68 °C for 15 min each. Membrane bound DIG probes were detected using anti-DIG antibody labeled with alkaline phosphatase and CSPD chemiluminescent substrate (Roche) and bands visualized on x-ray film after 1-2 hours.

Zebrafish DNA injection and analysis

Wild-type embryos were injected with 30 pg of EGFP, EGFP-G4C2 x8, x38 or x72 DNA vectors at the one-cell stage. Injected embryos were kept at 28 °C and fixed at prim-5 stage in 4% PFA for 4 hours at room temperature and then washed with PBS and kept in methanol at -20 °C. Apoptotic cells were detected by terminal transferase dUTP nick-end labeling (TUNEL) in whole-mount embryos using the Apoptag Red *In situ* Apoptosis Detection Kit (Millipore). The previously published protocol was followed (Williams and Holder, 2000) with the exception that the embryos were permeabilised with proteinase K for 10 minutes and the digoxigenin-tagged DNA was detected using a 1:3000 dilution of anti-DIG-Rhodamine antibody (Roche). EGFP expressing cells were detected by immunohistochemistry as previously described (Shanmugalingam et al., 2000). A rabbit anti-EGFP antibody was used (AMS Biotechnology), followed by the anti-rabbit Alexa-488 secondary antibody (Invitrogen), both at 1:1000 dilution. Embryos were incubated with Hoechst staining solution to visualize the nucleus (1:10,000 dilution of 10 mg/ml, Invitrogen). Embryos were imaged with a Nikon eclipse 80i confocal microscope, and the number of TUNEL and active caspase-3 positive cells was determined in a 3D reconstruction of the z-stacks, using the spot detection function in the IMARIS X64 6.02 software. Three independent experiments were carried out on three separate days and for each condition, 5 embryos were analyzed.

Rat tissue nuclear extracts

Rat brain cortex was homogenised in 5 ml of ice cold buffer A [10 mM HEPES (pH 7.4), 10 mM NaCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, EDTA-free protease inhibitor cocktail (Roche)] on ice. Homogenate was incubated on ice for 75 minutes and then centrifuged for 10 min at 3000 g, 4 °C. Supernatant was collected as cytosol fraction. Pellet was washed 3 times with 1.5 ml of buffer A and supernatants were added to cytosol fraction. Washed pellet as resuspended in 2 ml of ice cold lysis buffer [10 mM HEPES (pH 7.4), 400 mM KCl, 10 mM $MgCl_2$, 1% IGEPAL CA-630 (Sigma-Aldrich), EDTA-free protease inhibitor cocktail], incubated on ice, sonicated and then centrifuged for 10 min at 9223 g, 4 °C. Supernatant containing nuclear extract was transferred to a fresh falcon tube, diluted with dilution buffer [10 mM HEPES (pH 7.4), 10 mM $MgCl_2$, 0.33% IGEPAL CA-630, EDTA-free protease inhibitor cocktail] and recentrifuged at 16100 g, 4 °C for 5 min.

Aptamer linked G4C2 Cloning

Streptavidin-binding S1 aptamer (Srisawat and Engelke, 2001), 48 GGGGCC repeats and DsRed ORF 1-369, both containing S1 aptamer on 3 end, were cloned between HindIII (or BamHI for S1 alone) and XbaI in pcDNA3 vector.

RNA Aptamer pull-down assay

S1, 48xG4C2-S1 and RFP-S1 plasmids were linearised with restriction digestion with XbaI behind S1 and *in vitro* transcribed using T7 promoter on pcDNA3 vector with TranscriptAid T7 High Yield Transcription Kit (Fermentas). RNA pull-down was performed according to (Butter et al., 2009) with some modifications. RNAs were incubated with streptavidin magnetic beads (Promega) in RNA-binding buffer [10 mM HEPES (pH 7.4), 100 mM KCl, 10 mM $MgCl_2$, 0.5% IGEPAL CA-630] for 40 min at room temperature. Beads with bound RNA were washed 3 times in RNA-binding buffer and then incubated with 3 mg of rat brain cortex nuclear extract with RiboLock RNase inhibitor (Fermentas) and 50 ug of yeast tRNA for 4 h on 4 °C. Beads were washed 5 times with RNA-binding buffer prior to elution.

Western blot

Rat brain cortex nuclear extract (1.5 ug) and 1/7 of RNA pull-down eluates were run in 12% SDS-PAGE and transferred to nitrocellulose membrane. The membrane was blocked and probed with hnRNP H (Abcam) and subsequently with a HRP-conjugated anti-rabbit secondary antibody (Dianova). Chemiluminiscent detection was performed with Luminol reagent (Santa Cruz). The membrane was exposed to film for 1.5 min and film was developed.

Splice Analysis of candidate alternative exons

RNA was isolated from cells transduced with lentivirus expressing a shRNA targeted against hnRNP H , or shRNA control (sigma), using Qiagen RNA mini extraction kit. cDNA was generated using the RevertAid First strand cDNA synthesis kit (Thermo Scientific) according to manufacturer protocol, using random primers. Splicing effects were analyzed after PCR with gene-specific primers using the Qiaxcel capillary electrophoresis device (Qiagen).

Primer sequences

TARBP2	F	CACAGTGACCCAGGAGTCTG
TARBP2	R	TCGTAGAGAATCCCAGGTGC
pA	F	GTGCAGTGCTTCAGCCGCTA
pA	R	CTGCTTGTCGGCCATGATAT
pB	F	ATCCGCCACAACATCGAGGA
pB	R	GGCCCCTCTAGAGGTGAAGG
pC	F	ATCCGCCACAACATCGAGGA
pC	R	ACGCGTAGAATCGAGACCGA

Supplemental References

Butter, F., Scheibe, M., Mörl, M., and Mann, M. (2009). Unbiased RNA-protein interaction screen by quantitative proteomics. Proceedings of the National Academy of Sciences of the United States of America *106*, 10626–10631.

Chen, H.J., Anagnostou, G., Chai, A., Withers, J., Morris, A., Adhikaree, J., Pennetta, G., and de Belleroche, J.S. (2010). Characterization of the properties of a novel mutation in VAPB in familial amyotrophic lateral sclerosis. J. Biol. Chem. *285*, 40266–40281.

Mitchell, J., Paul, P., Chen, H.J., Morris, A., Payling, M., Falchi, M., Habgood, J., Panoutsou, S., Winkler, S., Tisato, V., et al. (2010). Familial amyotrophic lateral sclerosis is associated with a mutation in D-amino acid oxidase. Proceedings of the National Academy of Sciences of the United States of America *107*, 7556–7561.

Shanmugalingam, S., Houart, C., Picker, A., Reifers, F., Macdonald, R., Barth, A., Griffin, K., Brand, M., and Wilson, S.W. (2000). Ace/Fgf8 is required for forebrain commissure formation and patterning of the telencephalon. Development *127*, 2549–2561.

Srisawat, C., and Engelke, D.R. (2001). Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. Rna *7*, 632–641.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689–693.

Williams, J.A., and Holder, N. (2000). Cell turnover in neuromasts of zebrafish larvae. Hearing Research *143*, 171–181.

Figure S1

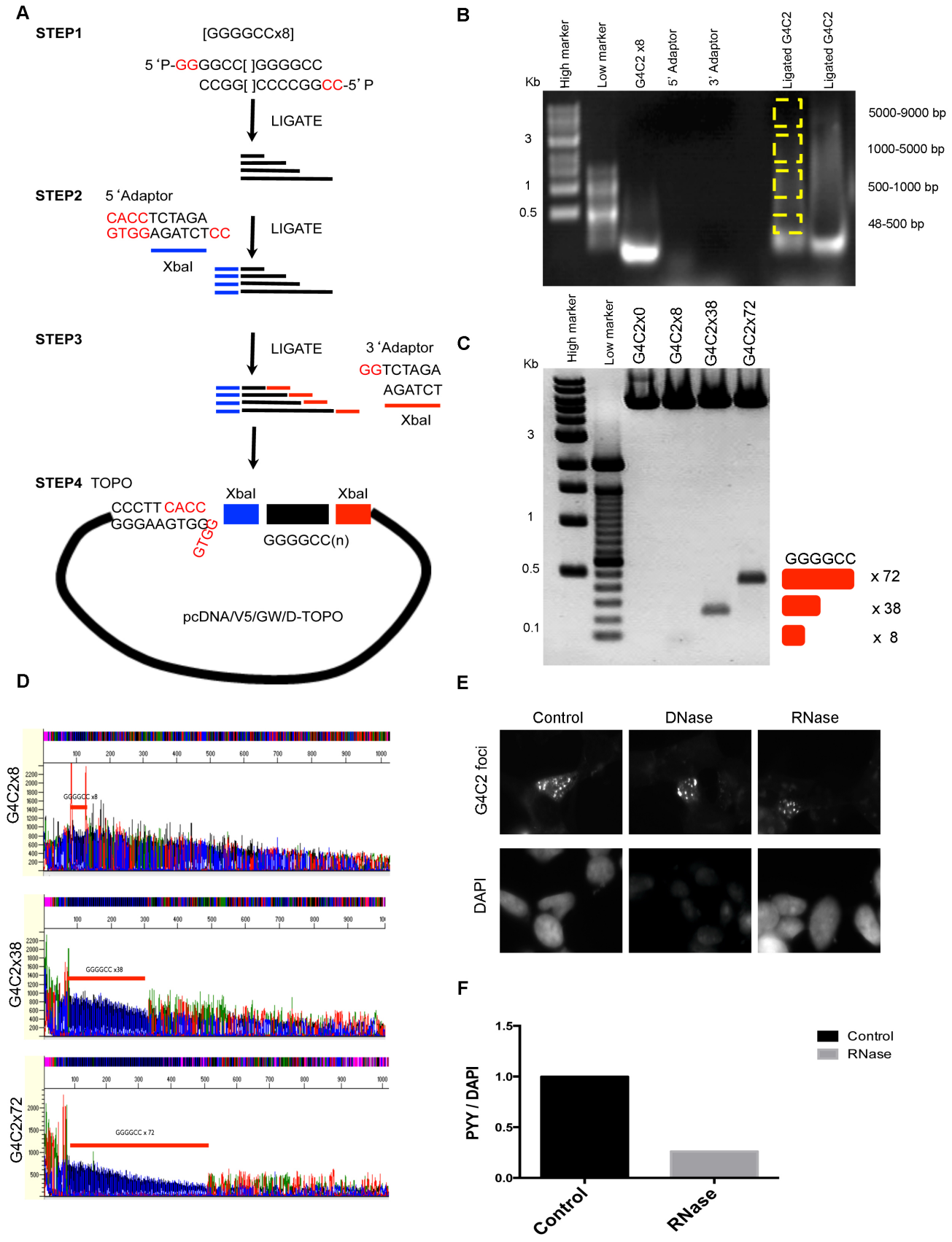
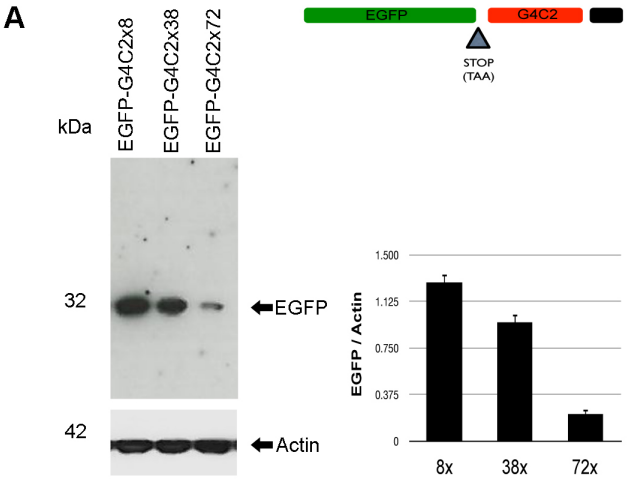
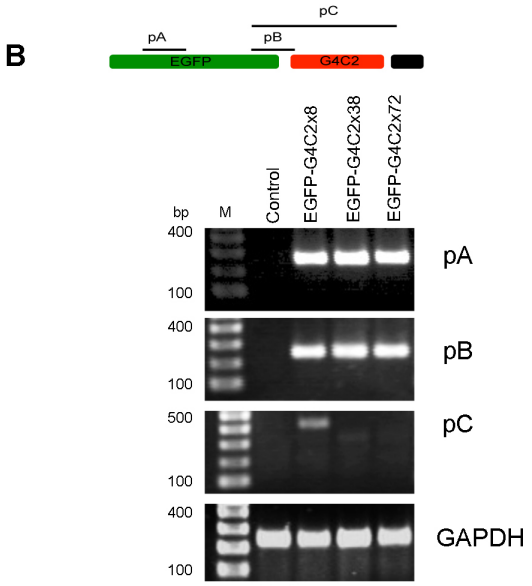

A
STEP1
[GGGGCCx8]
5 'P-GGGGCC[ ]GGGGCC
CCGG[ ]CCCCGGCC-5' P
LIGATE
STEP2
5 'Adaptor
CACCTCTAGA
GTGGAGATCTCC
XbaI
LIGATE
STEP3
LIGATE
3 'Adaptor
GGTCTAGA
AGATCT
XbaI
STEP4 TOPO
XbaI
XbaI
CCCTT CACC
GGGAAGTGG
GTGG
GGGGCC(n)
pcDNA/V5/GW/D-TOPO
B
High marker
Low marker
G4C2 x8
5' Adaptor
3' Adaptor
Ligated G4C2
Ligated G4C2
Kb
3
1
0.5
5000-9000 bp
1000-5000 bp
500-1000 bp
48-500 bp
C
High marker
Low marker
G4C2x0
G4C2x8
G4C2x38
G4C2x72
Kb
3
1
0.5
0.1
GGGGCC
x 72
x 38
x 8
D
G4C2x8
GGGGCC x8
G4C2x38
GGGGCC x38
G4C2x72
GGGGCC x 72
E
Control
DNase
RNase
G4C2 foci
DAPI
F
PYY / DAPI
1.5
1.0
0.5
0.0
Control
RNase
Control
RNase

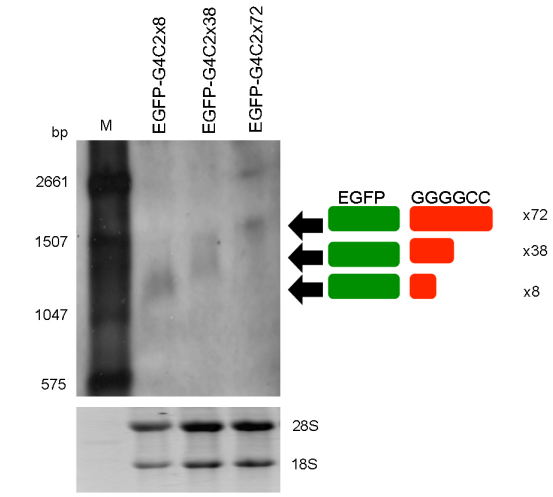
Figure S1. G4C2 repeats expression vector cloning, related to Figure 1

(A) Schematic diagram of G4C2 extension methods. 5' phosphorylated oligo 5'P-GG-(G4C2)x8 and 5'P-CC-(C4G2)x8 form a sticky end with overhang of GG and CC after overnight ligation. (B) Ligation products were separated by gel electrophoresis and dissected according to the size of the G4C2 expansion (yellow box). Gel extracted DNAs were used for Topo cloning and positive colonies were selected by XbaI digestion. (C) Stable E.coli lines containing 8x, 38x, 72x were selected and the presence of the insert confirmed by XbaI digestion as indicated. (D) G4C2 DNA sequence were verified by dGTP secondary structure sequencing. G4C2 repeats were indicated with red bar. (E) SH-SY5Y cells were transfected with the G4C2x72 plasmid and G4C2 foci were detected after treatment with DNase (10 U/ml) or RNase (400 ng/ml) for 10 min at 37 °C. DAPI staining shows decreased intensity in DNase-treated cells. (F) To identify the level of degraded RNA, RNase-treated cells were stained with the RNA-specific dye Pyronin Y (PYY, 30 μg/ml) for 5 min at room temperature and staining intensity was measured using a FLUOstar plate reader (DAPI, 358 nm; PYY, 567 nm) and PYY values normalized to DAPI values.

Figure S2


A
EGFP
G4C2
STOP
(TAA)
kDa
EGFP-G4C2x8
EGFP-G4C2x38
EGFP-G4C2x72
32
42
EGFP
Actin
EGFP / Actin
1.500
1.125
0.750
0.375
0
8x
38x
72x


B
pA
pB
pC
EGFP
G4C2
bp
M
Control
EGFP-G4C2x8
EGFP-G4C2x38
EGFP-G4C2x72
400
100
500
pA
pB
pC
GAPDH


C
bp
M
EGFP-G4C2x8
EGFP-G4C2x38
EGFP-G4C2x72
2661
1507
1047
575
EGFP
GGGGCC
x72
x38
x8
28S
18S

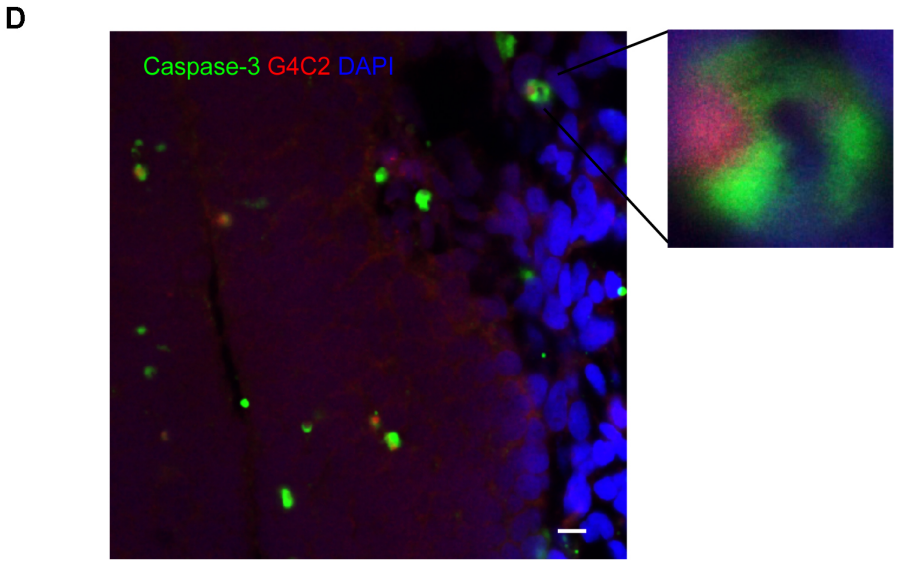

D
Caspase-3 G4C2 DAPI

Figure S2. G4C2 repeat expansion decreased protein translation of EGFP but not mRNA level, related Figure 2

(A) We cloned the EGFP reporter gene into the 5' to the G4C2 construct. Schematic diagram shows EGFP-G4C2 construct with a stop codon in between EGFP and G4C2. Westernblot shows that reduced EGFP translation of longer G4C2 transcripts is entirely consistent with size of 8x > 38x >72x repeats. (B) pA (primer A, amplify internal EGFP), pB (primer B, amplify between 3' end EGFP and G4C2), pC (primer C, amplify full sequence G4C2). Semi-quantitative reverse transcriptase (RT)-PCR showed equal expression of mRNA from internal EGFP primer (pA) for G4C2x8, G4C2x32 and G4C2x72. There was no difference of internal EGFP sequence between EGFP-G4C2 x38 and EGFP-G4C2x72 (pB). In contrast, the amplification of G4C2 region (pC) showed no signal from EGFP-G4C2x38 and 72x except 8x. These data suggest that a hairpin structure of G4C2 may disturb polymerase and prevent further amplification. (C) mRNA extracted from plasmid transfected SH-SY5Y cells were used for Nothernblot, which shows the predicted size of the EGFP-G4C2 constructs. (D) G4C2 foci are detected in active caspase-3 positive cells from zebrafish. G4C2 RNA foci positive cells (white arrow) were observed in zebrafish cells (Green=active caspase-3, Red=G4C2, Blue=DAPI), scale bar = 10 μm.

Figure S3

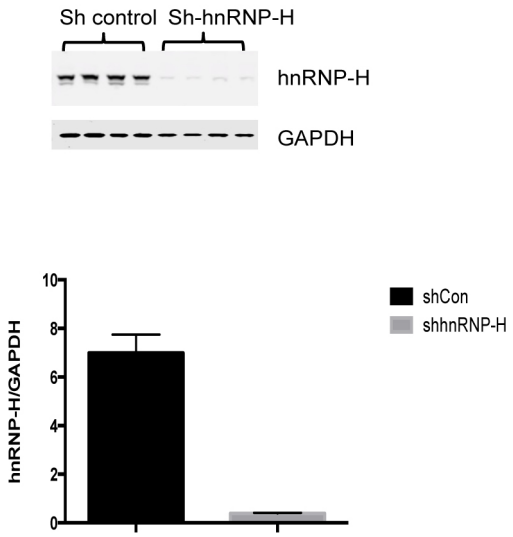


Figure S3. hnRNP-H knockdown efficiency by lentivirally expressed shRNA, related to Figure 3

Anti-hnRNP-H antibody was used for the detection of endogenously expressing hnRNP-H from SH-SY5Y cell lysate (n=4). Westernblot was analysed by Image j and GraphdPad Prism 5 software.

Figure S4

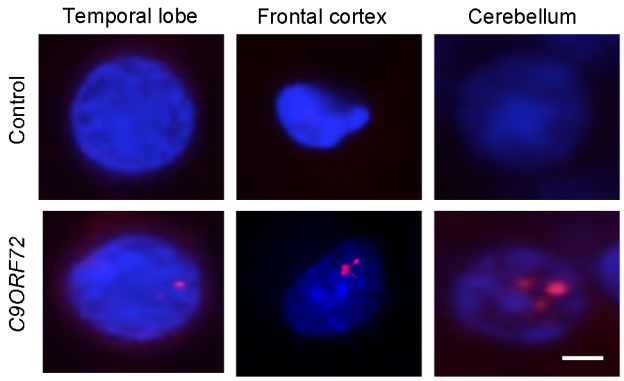


Figure S4. G4C2 foci are found in temporal lobe, frontal cortex and cerebellum in *C9ORF72* patients, related to Figure 4

G4C2 repeat FISH in autopsy brain from three areas (temporal lobe, frontal cortex and cerebellum), Scale bar = 3 μm.

Table S1

A

RNA binding proteins	
GENE	G4C2-Foci
TDP-43	**-**
EWS	**-**
PABP	**-**
PABPN1	**-**
9GB	**-**
TAFIIp68	**-**
TAP	**-**
PUR-a	**-**
MBNL	**-**
CUGBP	**-**
CUGBP2	**-**
RBM4	**-**
RBMX	**-**
U2AF65	**-**
SF2	**+**
SC35	**+**

B

hnRNPs	
GENE	G4C2-Foci
hnRNP-pan	**-**
hnRNP-A1	**-**
hnRNP-A2B1	**-**
hnRNP-A3	**-**
hnRNP-C1/C2	**-**
hnRNP-E1	**-**
hnRNP-F	**-**
hnRNP-H	**+**
hnRNP-K	**-**
hnRNP-L	**-**
hnRNP-M	**-**
hnRNP-Q	**-**
hnRNP-R	**-**
hnRNP-U	**-**

Table S1. Screening result of G4C2 RNA foci binding proteins, related to Figure 3

(A-B) SH-SY5Y cells were transfected with a plasmid expressing 72x repeats and probed 24 hours after transfection for G4C2 by FISH and immunocytochemistry (ICC) (A) RNA binding proteins. (B) Heterogeneous nuclear ribonucleoproteins (hnRNPs).

Table S2

Name	Species	Cat. Number	Dilution	Localization (N-C)
CUGBP1 (3B1)	Mouse	sc-20003	1in300	Nucleus
CUGBP2 (1H2)	Mouse	sc-47731	1in50	Nucleus
EWS (C-19)	Goat	sc-6532	1in100	Cyto+Nucleus
EWS (H-60)	Rabbit	sc-28865	1in100	Cyto+Nucleus
hnRNP A1	Mouse	sc5670	1in100	Nucleus
hnRNP C1/C2 (N-16)	Goat	sc-10037	1in100	Nucleus
hnRNP E1 (E-2)	Mouse	sc-137249	1in200	Cyto+Nucleus
hnRNP E2	Mouse	sc-101136	1in100	Cyto+Nucleus
hnRNP F	Mouse	sc-32309	1in50	Nucleus
hnRNP F/H (H-300)	Rabbit	sc15387	1in100	Cyto+Nucleus
hnRNP H (N-16)	Goat	sc10042	1in300	Cyto+Nucleus
hnRNP L	Mouse	sc-32317	1in300	Nucleus
hnRNP M	Mouse	sc134360	1in50	Cyto+Nucleus
hnRNP K	Rabbit	sc-25373	1in50	Nucleus
hnRNP K	Mouse	sc-28380	1in300	Cyto+Nucleus
hnRNP R (C-16)	Goat	sc-16541	1in50	Cyto+Nucleus
hnRNP pan	Mouse	sc166577	1in100	Nucleus
hnRNP Q (I8E4)	Mouse	sc-56703	1in200	Cytoplasm
hnRNPU (H-94)	Rabbit	sc-25374	1in100	Cytoplasm
MBNL1	Mouse	sc47740	1in50	Cyto+Nucleus
PABP (F-20)	Goat	sc-18611	1in200	Cytoplasm
PAPBN1 (G-17)	Goat	sc-33007	1in100	Cytoplasm
PSF (H-80)	Rabbit	sc-28730	1in50	Nucleus
Purα (80-L)	Mouse	sc-130397	1in50	Cyto+Nucleus
RBM4 (H-100)	Rabbit	sc-98346	1in100	Cyto+Nucleus
RBM4 (E-17)	Goat	sc-82352	1in50	Cytoplasm
RBMX (H-30)	Rabbit	sc-48796	1in50	Cytoplasm
TAF II p68 (TAF15B11)	Mouse	sc-81121	1in50	Nucleus
TAP	Mouse	sc-32319	1in300	Nucleus
U2AF65 (H-300)	Rabbit	sc-48804	1in300	Nucleus
9G8/SRp20 (H-120)	Rabbit	sc-28722	1in100	Cyto+Nucleus
hnRNP H	Rabbit	Abcam ab10374	1in500	Nucleus
SF2	Rabbit	Abcam ab38813	1in500	Nucleus
SC35	Mouse	Sigma s4045	1in500	Nucleus
TDP-43	Rabbit	Proteintech 10782-2AP	1in500	Cyto+Nucleus

Table S2. Antibody details, related to Figure 3

Table S3

A

Patient\Brain region	Temporal lobe	Frontal Cortex	Cerebellum
ALS (*C9ORF72*)	N/A	*	***
ALS/FTLD (*C9ORF72*)	N/A	*	***
ALS/FTLD (*C9ORF72*)	*	*	***
FTLD-U (*C9ORF72*)	N/A	*	***
ALS/FTLD (*C9ORF72*)	*	*	***
Control	N/A	N/A	N/A
Control	N/A	N/A	N/A
Control	N/A	N/A	N/A
Control	N/A	N/A	N/A
Control	N/A	N/A	N/A

B

Case	Age	Sex	PMD (hours)	Age of Onset
ALS (*C9ORF72*)	39	F	70	35
ALS/FTLD (*C9ORF72*)	43	F	69	42
ALS/FTLD (*C9ORF72*)	62	M	74	61
FTLD-U (*C9ORF72*)	79	M	35	68
ALS/FTLD (*C9ORF72*)	53	M	82	52
Control	43	F	43	N/A
Control	67	M	41	N/A
Control	90	F	50	N/A
Control	57	M	26	N/A
Control	74	M	23	N/A

Table S3. Human case details, related to Figure 4

(A) Foci in *C9ORF72* patients and control (number of foci low= *, number of foci high = ***).

(B) Age, Sex, post-mortem delay (PMD) and Age of onset.