Supplemental information on biophysical studies

r(GGGGCC)$_8$ preserves a hairpin structure with periodically repeating 1×1 nucleotide GG internal loops in equilibrium with a G-quadruplex

With the goal of designing small molecule modulators of r(GGGGCC)$_{exp}$, we investigated the structure of r(GGGGCC)$_{exp}$ by completing spectroscopic, chemical, and enzymatic analyses. First, circular dichroism (CD) studies of r(GGGGCC)$_4$, r(GGGGCC)$_6$ and r(GGGGCC)$_8$ were conducted in the presence of monovalent metal cations (100 mM K^+, Na^+, or Li^+) at pH 7.4 (**Fig. 1A**). Parallel G-quadruplexes, which are stabilized by K^+ and Na^+, give a signature negative peak at 242 nm and a large positive peak at 264 nm. Comparison of these CD spectra with r(CGG)$_{12}$, an RNA repeat known to form a hairpin structure (Zumwalt et al., 2007), revealed that both hairpin and quadruplex structures have a positive signal at 264 nm, leaving the negative signal at 242 nm as the only signature to distinguish quadruplexes from hairpins. In the presence of K^+, the formation of a negative signal at 242 nm and increased molar ellipticity at 264 nm indicated the potential formation of G-quadruplexes; however, no such effect was observed when r(GGGGCC)$_n$ was studied in the presence of other cations, in particular Na^+.

Given that G-quadruplexes have signature melting curves (a large hypochromic transition of UV absorbance at 295 nm (Kumari et al., 2007; Marin and Armitage, 2005; Mergny et al., 1998; Mullen et al., 2010)), we probed the structure of r(GGGGCC)$_4$, r(GGGGCC)$_6$, and r(GGGGCC)$_8$ by optical melting in the presence of 100 mM K^+ or 100 mM Na^+. In the presence of Na^+, no hypochromic shift was observed in UV melting profiles at 295 nm for any RNA. Moreover, all three RNAs form intramolecular structures, as their melting temperatures were independent of concentration. In contrast, a hypochromic shift was observed at 295 nm in the presence of K^+, but not until above 85^{o}C, indicating the presence of both hairpin and G-quadruplex structures (**Fig. 1B & Table S1**).

The folding of r(GGGGCC)$_8$ was next examined using enzymatic and chemical mapping

in the presence of Li^+ or K^+, the latter known to stabilize G-quadruplex formation (Ehresmann et al., 1987; Hardin et al., 1992). Enzymatic mapping was performed using S1 (cleaves single stranded and non-canonically paired nucleotides), T1 (cleaves single stranded and non-canonically paired G's), and V1 (cleaves base pairs). If $r(GGGGCC)_8$ forms a quadruplex, G residues should be protected from cleavage by T1 and S1 (Todd and Neidle, 2011). If $r(GGGGCC)_8$ forms a hairpin structure with internal loops in the stem, an alternating pattern of T1/S1 cleavage and V1 cleavage should be observed. Indeed, our mapping studies revealed such a pattern (**Fig. 1C**), suggesting that some population forms a hairpin structure. Enzymatic mapping data were used to construct a model of $r(GGGGCC)_8$'s structure using the program RNAstructure (Mathews et al., 2004), affording a hairpin with GG internal loops (**Fig. 1C**). Our hypothesis that $r(GGGGCC)_8$ forms a hairpin was further investigated by chemically probing the RNA's structure by reaction with dimethyl sulfate (DMS). DMS methylates the N7 position of G's (Ehresmann et al., 1987). The N7 position of G's in quadruplexes are hydrogen bonded and thus protected from methylation (Todd and Neidle, 2011). In these studies, the majority of guanine N7 positions were susceptible to DMS modification and the pattern did not change in the presence of Li^+ or K^+.

We additionally explored the structure of $r(GGGGCC)_8$ by analyzing its 1D ^1H NMR spectra. Guanines in non-canonically paired conformations (internal loops, quadruplexes) typically give rise to imino proton signals from 10 to 12 ppm whereas the resonances from G's in base pairs appear from 12 to 14 ppm (Bugaut et al., 2012). $r(GGGGCC)_8$ prepared in 10 mM Tris HCl and 100 mM KCl was heated at 37°C, 60°C, or 95°C, followed by slow cooling and equilibration at room temperature for 2 h. As shown in **Fig. 1D**, well defined peaks were observed in two regions (10.0-11.8 and 12.0-13.5 ppm) after annealing at 37°C, suggesting formation of a hairpin structure with non-canonically paired Gs. As the annealing temperature increased, however, the peaks in both regions became broad. The signals in the 12-14 ppm range (Watson-Crick paired Gs) were reduced while those in the 10-12 ppm range were

increased, indicating increased population of a G-quadruplex. The existence of both conformations is not surprising, as other studies have suggested RNAs that form quadruplexes can form alternative structures that include hairpins (Bugaut et al., 2012; Fojtik et al., 2004).

Supplemental experimental procedures

Abbreviations. BLI, biolayer interferometry; bp, base pair; CA, chlorambucil; DCM, dichloromethane; DIPEA, *N,N*-diisopropylethylamine; DMF, *N,N*-dimethylformamide; DMS, dimethyl sulfate; DMSO, dimethyl sulfoxide; DNA, deoxyribonucleic acid; DEPC-PBS, DEPC-treated PBS; DPBS, Dulbecco's phosphate buffered saline; EDTA, ethylenediaminetetraacetic acid; EtOAC, ethyl acetate; HBTU, O-(benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate; HEPES, 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid; HOBt, 1-hydroxybenzotriazole; HPLC, high performance liquid chromatography; HRMS, high resolution mass spectrometry; LC-MS, liquid chromatography-mass spectrometry; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; MS, mass spectrometry; NMR, nuclear magnetic resonance; PBS, phosphate buffered saline; qRT-PCR, quantitative real time polymerase chain reaction; RNA, ribonucleic acid; t_R, retention time; TBE, TBST, Tris buffered saline supplemented with 0.05% Tween-20; TFA, trifluoroacetic acid; Tris, tris(hydroxymethyl)aminomethane; UV, ultraviolet; Vis, visible

Reagents and oligonucleotides preparation. All reagents used for chemical synthesis were purchased from commercially available sources and used without further purification unless noted otherwise. NMR solvents were obtained from Cambridge Isotope Labs and used as is.

Instrumentation. Mass spectra were collected using an ABI 4800 MALDI-TOF or Varian 500-MS IT mass spectrometer. Reverse-phase HPLC was completed using a Waters 1525 binary

HPLC pump equipped with a Waters 2487 dual absorbance detector system. Optical melting spectra were acquired using a Beckman Coulter DU800 UV-Vis spectrometer connected to a Peltier heater. Circular dichroism experiments were performed on a Jasco J-815 spectrometer equipped with a Jasco Peltier temperature controller. TO-PRO-1 displacement assays were performed on a PerkinElmer Envision® multilabel reader. Gel images were acquired using a Molecular Dynamics Typhoon 9410 variable mode imager. The concentration of total RNA isolated from cells was determined using a Thermo Scientific Nanodrop 2000C spectrophotometer. ^{1}H NMR spectra of RNA were recorded at 10 ºC using a 700 MHz Bruker Avance TCI spectrometer equipped with a cryogenic TCI ATM probe, water suppression was achieved using excitation sculpting. ^{1}H NMR (400 MHz) and ^{13}C NMR (100 MHz) spectra for compound characterization were recorded at 25 ºC on a 400 MHz Bruker Avance spectrometer. Chemical shifts (δ) are given in ppm relative to tetramethylsilane or the respective NMR solvent; coupling constants (*J*) are in Hertz (Hz). Abbreviations used are s, singlet; bs, broad singlet; d, doublet; dd, doublet of doublets; t, triplet; dt, doublet of triplets; td, triplet of doublets; tt, triplet of triplets; bt, broad triplet; q, quartet; m, multiplet; and bm, broad multiplet. HRMS were obtained at the Scripps Florida Mass Spectrometry and Proteomics Laboratory.

Circular Dichroism (CD). RNA samples (4 μM) were folded in 1× CD Buffer (10 mM Tris HCl, pH 7.4 containing no monovalent cation or 100 mM LiCl, NaCl or KCl) by heating at 95 ^{o}C for 5 min and then slowly cooling to room temperature. Samples were then transferred into a 1 mL quartz cell with a pathlength of 1 mm. CD spectra were recorded at 20 ^{o}C by measuring ellipticity from 220 to 320 nm at a rate of 50 nm/min, a 2 second digital integration time (D.I.T.), 1 nm data pitch, and 1 nm band width. The background was subtracted from each spectrum, which were smoothed and normalized to zero at the starting point (320 nm).

Nuclease Mapping. Nuclease mapping experiments were performed as previously described (Auron et al., 1982). Briefly, 5' end-^{32}P-labeled r(GGGGCC)$_8$ was dissolved in 1× Mapping Buffer (10 mM Tris HCl, pH 7.4, 0.3 mM $MgCl_2$) supplemented with 185 mM KCl. In the case of S1 footprinting, the buffer was also supplemented with 10 mM $ZnCl_2$. The RNA was folded by heating in the corresponding buffer to 95° C for 5 min and slowly cooling to room temperature on the bench top. Enzymatic digestions using T1 (0.01 U/μL) under denaturing conditions (1× RNA Sequencing Buffer; Life Technologies), T1 (1 U/μL), V1 (0.001 U/μL) and S1 (0.1 U/μL) under non-denaturing condition were carried out at room temperature for 15 min and quenched by the addition of 1× Loading Buffer (1 mM Tris HCl, pH 7.5, 10 mM EDTA, and 4 M urea) and incubation at 95 °C for 2 min. Cleavage products were separated on a denaturing 20% polyacrylamide gel and visualized by autoradiography. Sites of cleavage were used as restraints in secondary structure prediction by free energy minimization (RNAstructure, version 5.4) (Bellaousov et al., 2013).

DMS Footprinting. DMS footprinting experiments were completed as previously described (Ziehler and Engelke, 2001). Briefly, 5' end-^{32}P-labeled r(GGGGCC$_)_8$ was folded in 10 mM Tris HCl, pH 7.4, containing 185 mM KCl, NaCl, or LiCl by heating at 95° C and slowly cooling to room temperature. To the samples were added DMS (dissolved in 1:1 EtOH:H_2O) to a final concentration of 3% (v/v), and the samples were incubated for 2 min. Reactions were quenched by ethanol precipitation; the resulting pellets were washed once with 70% ethanol and briefly dried in a vacuum concentrator. The RNA samples were dissolved in 1 M Tris HCl, pH 8 followed by addition of 0.1 M $NaBH_4$ and incubation on ice for 30 min in the dark. The reactions were quenched by ethanol precipitation as described above. Aniline cleavage of the modified RNA was completed by dissolving the RNA in freshly prepared 1 M aniline in 0.3 M NaOAc, pH 4.5 followed by incubation at 60° C for 20 min. The samples were ethanol

precipitated and dissolved in 1× Loading Buffer. Fragments were separated on a denaturing 20% polyacrylamide gel and visualized by autoradiography.

Optical Melting. The RNA of interest (1 – 35 μM) was folded in 10 mM Tris HCl, pH 7.4 and 100 mM NaCl or 100 mM KCl by heating at 95 ^{o}C for 5 min and slowly cooled to room temperature. For experiments completed for r(GGGGCC)$_8$ in the presence of small molecule, 1 μM RNA was folded as described above followed by addition of 3 μM compound and incubation at room temperature for 15 min. Absorbance versus temperature spectra were then acquired at 260 nm and 295 nm at a rate of 1 ^{o}C/min. Melting curves were fit to a self-complementary model using MeltWin (http://www.meltwin.com). The program fits each curve and calculates thermodynamic parameters and melting temperature (T_m) (**Table S1, Table S3, Fig. 1B, Fig. S1C**).

^{1}H NMR spectroscopy. A 600 μM sample of r(GGGGCC)$_8$ was prepared in 10 mM Tris HCl, pH 7.4 and 100 mM KCl and annealed at the appropriate temperature for 5 min. The sample was then slowly cooled to room temperature. After equilibration at room temperature for 2 h, the sample was transferred to a 3 mm Shigemi D_2O NMR tube, and NMR spectra were recorded at 10 ^{o}C.

TO-PRO-1 displacement screening. r(GGGGCC)$_8$ (36 nM) was folded in 8 mM Na_2HPO_4, pH 7.0, 185 mM NaCl, and 1 mM EDTA by heating at 95 ^{o}C for 5 min and slowly cooling to room temperature. TO-PRO-1 and BSA were then added to final concentrations of 10 nM and 40 μg/mL, respectively, and the samples were incubated at room temperature for 15 min. The compound of interest (100 μM) was added, and the samples were incubated for an additional 15

min at room temperature. After incubation, fluorescence intensity was recorded and converted to the percentage of dye-RNA complex using equation 1:

$$y = \frac{I - I_0}{I_{max} - I_0} \times 100\% \ (eq. 1)$$

where I is the observed fluorescence intensity, I_0 is the fluorescence intensity in the absence of RNA, I_{max} is the fluorescence intensity in the absence of compound.

This screen identified 31 compounds (out of 132) that displaced >95% of TO-PRO-1 from the RNA, which were carried forward to additional screening at lower concentrations (10 and 1 μM) (**Fig. S1A, Table S2**). As a control, the fluorescence of hit compounds in the presence of TO-PRO-1 but in the absence of r(GGGGCC)$_8$ was also measured.

Biolayer Interferometry (BLI). Biotinylated RNA (ligand) was loaded onto the surface of streptavidin biosensors (SA) for 660 s. Optimal response levels were between 0.5 and 2 nm, and variability within a row of eight tips did not exceed 0.2 nm. Biosensors were then washed in 1× Kinetics Buffer for 300 s followed by association of the compound (analyte) for 5000 s. Finally, dissociation of the ligand-analyte interaction was analyzed for 5000 s. The resulting curves were corrected by subtracting the response recorded on a sensor loaded with ligand (RNA) but incubated with no analyte (compound). Data analyses and curve fitting were completed using Octet Data Analysis, version 7.0. Experimental data were fitted using the 2:1 heterogeneous ligand (HL) curve fit. Global analysis of all data sets acquired for different analyte concentrations, assuming reversible binding, was completed using nonlinear least squares fitting. K_ds were calculated using steady-state kinetic analysis of the estimated response at equilibrium (R_{eq}) according to equation 2 and 3 (**Fig. S3**).

$$y = R_{max} \frac{[Analyte]}{[Analyte] + K_d} \ (eq. 2)$$

$$R_{max} = R_{eq}\frac{k_{on} \times [Analyte]}{k_{on} \times [Analyte] + k_{off}} \ (eq.3)$$

where $[Analyte]$ is the concentration of compound, R_{eq} is the estimated response at equilibrium, k_{on} is association constant, k_{off} is dissociation constant.

Synthesis of 1a-CA-biotin

2-(5-(aminopentyl)-9-hydroxy-5,11-dimethyl-6H-pyrido[4,3-b]carbazol-2-ium. A sample of 9-hydroxyellipticine (Deane et al., 2011; Plug et al., 1992) (50 mg, 0.19 mmoles) was dissolved in 4 mL of DMF, and tert-butyl (5-bromopentyl)carbamate (Hingorani et al., 2013) (130 mg, 0.49 mmoles in 2 mL of DMF) was added. The mixture was stirred at room temperature overnight. After removing DMF *in vacuo*, TFA dissolved in DCM was added to the residue, and the mixture was stirred at room temperature for 1 h. The mixture was concentrated, and the product was purified by HPLC (20-60% MeOH/H_2O with 0.1% TFA over 60 min) to yield the desired product as a red solid (22 mg, 33% yield). ^{1}H NMR (400 MHz, CD_3OD) δ 9.51 (s, 1H), 8.15 (d, *J* = 7.2 Hz, 1H), 8.05 (d, *J* = 7.2 Hz, 1H), 7.27 (d, *J* = 2.2 Hz, 1H), 7.14 (d, *J* = 8.6 Hz, 1H), 6.89 (dd, *J* = 8.6 Hz, *J* = 2.3 Hz 1H), 4.63 (t, *J* = 7.6 Hz, 2Hs), 3.00 (t, *J* = 7.4 Hz, 2Hs), 2.90 (s, 3Hs), 2.58 (s, 3Hs), 2.13 (m, 2Hs), 1.81 (m, 2Hs), 1.58 (m, 2Hs). ^{13}C NMR (400 MHz, CD_3OD) δ 153.0, 146.2, 146.0, 137.4, 134.1, 133.6, 130.9, 127.3, 124.0, 121.3, 121.2, 118.1, 112.7, 110.9, 110.3, 61.1, 40.4, 31.9, 28.1, 24.4, 15.0, 11.9. HRMS (FAB) calculated for $C_{22}H_{26}N_3O$ (M^+) 348.2070, found 348.2073.

2-(5-(2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-6-((tert-butoxycarbonyl)amino) hexanamido)pentyl)-9-hydroxy-5,11-dimethyl-6H-pyrido[4,3-b]carbazol-2-ium Fmoc-Lys(Boc)-OH (54 mg, 0.11 mmoles), HBTU (130 mg, 0.34 mmoles), HOBT (52 mg, 0.34 mmoles), and DIPEA (88 mg, 0.68 mmoles) were dissolved in 1 mL DMF, and the mixture was stirred at room temperature for 30 min. Then, 2-(5-(aminopentyl)-9-hydroxy-5,11-dimethyl-6H-pyrido[4,3-b]carbazol-2-ium (20 mg, 0.057 mmoles) was added to the mixture, which was stirred

at room temperature overnight. EtOAc was added to the solution, and the organic layer was washed with H_2O and dried over Na_2SO_4. The concentrated residue was filtered by silica gel column and was used for the next reaction without further purification.

2-(5-(6-amino-2-(5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido) hexanamido)pentyl)-9-hydroxy-5,11-dimethyl-6H-pyrido[4,3-b]carbazol-2-ium. A solution of 2-(5-(2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-6-((tert-butoxycarbonyl)amino) hexanamido)pentyl)-9-hydroxy-5,11-dimethyl-6H-pyrido[4,3-b]carbazol-2-ium from the previous synthetic step was dissolved in 5 mL 20% piperidine/ DMF and stirred at room temperature for 2 h, and the solvent was removed *in vacuo*. A mixture of biotin (98 mg, 0.40 mmoles), HBTU (303 mg, 0.80 mmoles), HOBT (122 mg, 0.80 mmoles), and DIPEA (206 mg, 1.6 mmoles) in 2 mL DMF was stirred for 30 min at room temperature. The mixture was then added to the concentrated residue and stirred at room temperature overnight. The solution was concentrated and treated with 10 mL 50% TFA in DCM for 2 h. After removing the solvent, the product was dissolved in 20% MeOH in H_2O and purified by HPLC (20-75% MeOH/H_2O with 0.1% TFA over 60 min) to yield the desired product as an orange solid (11 mg, 28% yield). ^{1}H NMR (400 MHz, CD_3OD) δ 9.74 (s, 1H), 8.28 (s, 2Hs), 7.63 (d, *J* = 2.2 Hz, 1H), 7.38 (d, *J* = 8.6 Hz, 1H), 7.10 (dd, *J* = 8.6 Hz, *J* = 2.3 Hz 1H), 4.67 (t, *J* = 7.4 Hz, 2Hs), 4.50 (dd, *J* = 7.8 Hz, *J* = 4.5 Hz 1H), 4.29 (dd, *J* = 7.8 Hz, *J* = 4.4 Hz 1H), 4.23 (dd, *J* = 9.0 Hz, *J* = 5.3 Hz 1H), 3.28 (m, 2Hs), 3.16 (s, 3Hs), 3.15 (m, 2Hs), 2.91 (m, 3Hs), 2.77 (s, 3Hs), 2.71 (d, *J* = 12.8 Hz, 1H), 2.26 (t, *J* = 7.1 Hz, 2Hs), 2.13 (m, 2Hs), 1.60 (m, 10Hs), 1.47 (m, 4Hs), 1.39 (m, 4Hs). ^{13}C NMR (400 MHz, CD_3OD) δ 174.8, 173.0, 164.8, 159.6, 151.9, 145.1, 136.4, 133.2, 132.5, 126.5, 123.0, 120.1, 117.0, 111.5, 109.8, 109.3, 61.9, 60.2, 60.0, 55.7, 53.5, 39.6, 39.1, 38.4, 34.8, 30.9, 30.6, 29.3, 28.4, 28.0, 26.7, 25.3, 23.1, 22.8, 13.8, 10.6. HRMS (FAB) calculated for $C_{38}H_{52}N_7O_4S$ (M^+) 702.3796, found 702.3802.

2-(5-(6-(4-(4-(bis(2-chloroethyl)amino)phenyl)butanamido)-2-(5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido)hexanamido)pentyl)-9-hydroxy-5,11-dimethyl-6H-

pyrido[4,3-b]carbazol-2-ium Chlorambucil (26 mg, 0.085 mmoles), HBTU (43 mg, 0.11 mmoles), HOBT (17 mg, 0.11 mmoles) and DIPEA (58 mg, 0.44 mmoles) in 1 mL DMF were stirred at room temperature for 40 min. Into the mixture was added 2-(5-(6-amino-2-(5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido)hexanamido)pentyl)-9-hydroxy-5,11-dimethyl-6H-pyrido[4,3-b]carbazol-2-ium (11 mg, 0.016 mmoles). The solution was stirred at room temperature overnight and concentrated *in vacuo*. After removing the solvent, the product was dissolved in 20% acetonitrile in H_2O and purified by HPLC (20-75% acetonitrile/H_2O with 0.1% TFA for 60 min) to yield the desired product as a yellow solid (2.1 mg, 15% yield). HRMS (FAB) calculated for $C_{52}H_{69}Cl_2N_8O_5S$ (M^+) 987.4483, found 987.4488; t_R = 29 min.

Cloning of r(GGGGCC)$_n$ expression vectors. The generation of r(GGGGCC)$_2$, r(GGGGCC)$_{20}$ and r(GGGGCC)$_{66}$ expression vectors was previously reported (Gendron et al., 2013). In brief, genomic DNA from muscle or spleen from a *C9ORF72* expanded repeat carrier was used as a template in a nested PCR strategy using ThermalAce DNA Polymerase (Invitrogen) to amplify the (GGGGCC)$_n$ repeat region, including 113 bp of 5′ and 99 bp of 3′ flanking sequence. The PCR products were cloned into the pAG3 expression vector. These constructs contain 3 upstream stop codons in each reading frame. Clones containing r(GGGGCC)$_2$, r(GGGGCC)$_{20}$ and r(GGGGCC)$_{66}$ were verified by hairpin sequence analysis.

Identification of the RNA targets of 1a, 2 and 3 by qRT-PCR. COS7 cells were grown as monolayers in a 75 cm^2 flask to ~95% confluency and then transfected with r(GGGGCC)$_{66}$ using Lipofectamine 2000 (Invitrogen) per the manufacturer's recommended protocol. Approximately 16 h post-transfection, **1a-CA-Biotin** and the compound of interest, or vehicle, were added to the cells, and the samples were incubated at 37 ^{o}C for 20-24 h. Total RNA was extracted using Trizol reagent (Ambion) according to the manufacturer's protocol. Approximately 100 µg of isolated total RNA in 100 µl of 1× PBS was added to a 2 ml centrifuge tube containing a filter

column (Sigma-Aldrich) with 300 µl of streptavidin beads (Sigma-Aldrich; washed three times with 300 µl of 1× PBS). The RNA and beads were incubated at room temperature for 1 h with gentle agitation (700 rpm). The solution containing unbound RNA was removed by filtration, and the beads were washed with 1× TBST (6 x 200 µl) until RNA was no longer eluted as determined by absorbance at 260 and 280 nm. Bound RNA-**1a-CA-Biotin** adducts were released from beads by heating in 50 µl 1× Elution Buffer (95% formamide, 10 mM EDTA, pH 8.2) at 65 °C for 5 min. The concentration of the bound RNA was quantified by UV absorbance. cDNA was generated from 50 ng of RNA using a qScript cDNA Synthesis Kit (Quanta Biosciences) per the manufacturer's protocol. Power SYBR® Green PCR Master Mix (Applied Biosystems) was used to quantify the amount of r(GGGGCC)$_{66}$ according to the manufacturer's protocol. The amount of the r(GGGGCC)$_{66}$ was normalized relative to 18S rRNA. Primer sequences for r(GGGGCC)$_{66}$ (C9down-F and C9down-R), 18S rRNA (18S-F and 18S-R), and β-actin internal control (hACTB-F and hACTB-R) are provided in the table below.

Sequences of primers used in qRT-PCR analysis.[a]

Primer ID	Sequence
C9down-F	5'-GGG CCC TAT TCT ATA GTG TCA CC
C9down-R	5'-ACA ACA GAT GGC TGG CAA C
18S-F	5'-GTA ACC CGT TGA ACC CCA TT
18S-R	5'-CCA TCC AAT CGG TAG TAG CG
hACTB-F	5'-CCT GGC ACC CAG CAC AAT
hACTB-R	5'-GGG CCG GAC TCG TCA TAC

[a] "F" and "R" indicates forward and reverse primers, respectively.

Western blot analysis of c9RAN proteins. HEK293 cells were cultured in Opti-Mem supplemented with 10% FBS and 1% penicillin/streptomycin. To detect products of

r(GGGGCC)$_n$ RAN translation, 90% confluent cells grown in 6-well plates were transfected with 5 µg of (GGGGCC)$_2$, (GGGGCC)$_{20}$ or (GGGGCC)$_{66}$ vectors using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Twenty-four hours later, cell pellets were collected. To determine the effect of compounds on RAN translation, cells were treated with DMSO (vehicle) or compound (**1a**, **2** or **3**) 4 h after transfection, followed by collection of cell pellets 24 h later. Western blotting was performed as previously described (Gendron et al., 2013). In brief, cell pellets were lysed in Co-IP buffer (50 mm Tris–HCl, pH 7.4, 300 mM NaCl, 1% Triton-X-100, 5 mM EDTA, 2% sodium dodecyl sulfate (SDS), plus phenylmethylsulfonyl fluoride (PMSF) and both a protease and phosphatase inhibitor mixture). After centrifugation at 16,000 × g for 20 min at 4°C, the supernatant was collected and protein concentration determined by BCA assay. Samples were prepared in Laemmli's buffer, heated for 5 min at 95°C, and equal amounts of protein were loaded into Novex® 4–20 % Tris-Glycine gels (Invitrogen). After transfer, blots were blocked with 5 % non-fat dry milk in Tris-buffered saline + 0.1 % Triton X-100 (TBST) for 1 h, and then incubated with rabbit polyclonal anti-GP, anti-GA or anti-GR (1:1,000) overnight at 4°C. Anti-GAPDH (1:10,000, BioDesign) was used to ensure equal loading among wells. Membranes were washed three times for 10 min in TBST and incubated with donkey anti-rabbit or anti-mouse IgG conjugated to horseradish peroxidase (1:5000; Jackson ImmunoResearch) for 1 h. Protein expression was visualized by enhanced chemiluminescence treatment and exposure to film.

Immunoassay analysis of c9RAN proteins. As an alternative means to measure poly(GP) proteins, Meso Scale Discovery (MSD) electrochemiluminescence detection technology was utilized to establish sandwich immunoassays using polyclonal anti-GP as capture and detection antibodies. For validation of poly(GP) assay specificity, synthetic peptides (200 ng/ml) representing each possible protein RAN translated from the sense or antisense transcripts of the expanded *C9ORF72* repeat [(GA)$_8$, (GR)$_8$, (GP)$_8$, (PA)$_8$, (PR)$_8$] were assayed (**Fig. S4A**).

Poly(GP) proteins were also measured in lysates from cultured cells (10-35 µg of protein per well) prepare as described above, or from RIPA-soluble homogenates from frozen frontal cortical tissues (35 µg of protein per well). Brain homogenates were prepared as previously described (Almeida et al., 2013). In brief, tissue was lysed in cold RIPA buffer and sonicated on ice. Lysates were cleared by centrifugation at 100,000 g for 30 min at 4°C. The supernatant was collected and protein concentration was determined by BCA assay. Poly(GP) protein expression was similarly evaluated in CSF (90 µl per well, in duplicate or triplicate wells) from 5 healthy controls, 25 ALS patients without the *C9ORF72* repeat expansion, and 14 ALS patients with the expansion (see **Table S4** for patient information and the section on Human Samples below for additional information on CSF collection).

A second MSD sandwich immunoassay was developed for the detection of poly(PR) proteins using polyclonal anti-PR as capture and detection antibodies. To validate specificity of the poly(PR) assay, lysates from cells transfected to express each possible protein RAN translated from the sense or antisense transcripts of the expanded *C9ORF72* repeat [GFP-$(GA)_5$, GFP-$(GR)_5$, GFP-$(GP)_5$, GFP-$(PA)_5$, GFP-$(PR)_5$] were assayed, as were lysates from cells expressing $(CCCCGG)_{66}$ (**Fig. S2F**).

RNA fluorescence *in situ* hybridization (FISH) of $(GGGGCC)_n$-expressing cells. HEK293T cells grown on glass coverslips in 24-well plates were transfected with 0.5 µg r$(GGGGCC)_2$, r$(GGGGCC)_{20}$ or r$(GGGGCC)_{66}$ vectors. After 24 h, cells were fixed in 4% paraformaldehyde for 20 min, permeabilized in ice-cold methanol for 10 min, and washed 3 times with DEPC-treated PBS (DEPC-PBS). Cells were hybridized with denatured Cy3-conjugated $(GGCCCC)_4$ probe (2 ng/µl) in hybridization buffer (50% formamide, 10% dextran sulfate, 0.1 mg/mL yeast tRNA, 2xSSC, 50 mM sodium phosphate) overnight at 37°C. Cells were then washed once with 40% formamide/1xSSC for 30 min at 37°C and twice with DEPC-PBS at room temperature for 5 min,

followed by counterstaining with Hoechst 33258 (1 μg/ml, Invitrogen). Immunostained cells were visualized using a Zeiss Axiovert Fluorescence Microscope with apotome module.

To evaluate the effect of compounds on foci formation, HEK293 cells grown on glass coverslips in 24-well plates were transfected with 0.6 μg of r(GGGGCC)$_{66}$ vector. Four hours after transfection, cells were treated with DMSO or compound (**1a**, **2** or **3**) for 24 h, and then subjected to FISH as described above. To quantify foci-bearing cells, coverslips mounted on slides were scanned by Aperio ScanScope. Ten fields were randomly selected under 20× magnification. For each field, the number of foci-positive nuclei and the total number of nuclei were counted using MetaMorph software. These counts were used to determine the average percentage of foci-positive cells for each condition.

To determine whether r(GGGGCC)-binding compounds impair binding of the RNA-probe to r(GGGGCC), non-treated (GGGGCC)$_{66}$-expressing cells were fixed with 4% PFA, permeabilized with 0.2% Triton X-100 in DEPC-PBS, and washed twice with DEPC-PBS. RNA FISH was then performed using hybridization buffer containing the Cy3-(GGGGCC)$_4$ RNA probe and either DMSO or **1a** in excess of 20 times the molar concentration of the probe.

Western blot analysis and RNA-FISH of cells expressing antisense (CCCCGG) repeats. To examine the effect of **1a** on RAN translation and foci formation in cells expressing antisense (CCCCGG) repeats, we utilized a previously described (CCCCGG)$_{66}$ expression vector (Gendron et al., 2013). Transfection, treatment, Western blotting, and RNA-FISH using a 5'Cy3-(GGGGCC)$_4$-3' probe from IDT, were conducted as described above for (GGGGCC)$_{66}$-expressing cells.

Western blot analysis of RAN translation in a FXTAS cell model. Studies were completed using a plasmid in which r(CGG)$_{88}$ is embedded in the 5' UTR of an open reading frame encoding GFP (Todd et al., 2013). Therefore, RAN products are fused to GFP and can be

detected using an anti-GFP antibody (Todd et al., 2013). COS7 cells were grown as monolayers in 96-well plates in growth medium (1X DMEM, 10% FBS, and 1X GlutaMax (Invitrogen)). After the cells reached 90−95% confluency, they were transfected with 200 ng of plasmid using Lipofectamine 2000 (Invitrogen) per the manufacturer's standard protocol. Compound **1a** was added to the transfection cocktail, which was then applied to the cells. The transfection cocktail was replaced with growth medium containing **1a** approximately 5 h post transfection, and the cells were incubated at 37 °C for 18 h. Cells were lysed in the plate using 100 µl/well of MPER Mammalian Protein Extraction Reagent (Pierce Biotechnology) containing 1 µl of Halt Protease Inhibitor cocktail (Thermo Scientific). Cellular proteins were separated by SDS-PAGE (10% polyacrylamide) and then transferred to a PVDF membrane by wet transfer method. Protein content was analyzed by Western blotting by using anti-GFP (Santa Cruz) or anti-β-actin (Sigma Aldrich) as primary antibodies and anti-IgG-horseradish peroxidase conjugate as the secondary antibody. Chemiluminescent signal was generated by SuperSignal West Pico Chemiluminescent substrate (Thermo Scientific), and the blot was imaged with X-ray film (Phenix Research).

Human Samples. Frozen frontal cortex tissue used for biochemical analysis included samples from 6 FTD/ALS cases with the *C9ORF72* expansion, and 4 FTD/ALS cases without the expansion.

Fibroblasts were derived from skin sampled by punch biopsy on the anterior aspect of the forearm. Skin biopsies were obtained from six individuals, which included three control participants (control 1: female diagnosed with sixth nerve palsy, 61 years of age at the time of biopsy; control 2: healthy female, 64 years of age at the time of biopsy; control 3: healthy female, 38 years of age at the time of biopsy) and three repeat expansion carriers (carrier 1: 28 year old female at the time of biopsy; carrier 2: female diagnosed with ALS at 49 years of age, 50 years of age at the time of biopsy; carrier 3: male diagnosed with ALS/FTD at 41 years of age, 43

years of age at the time of biopsy). Fibroblasts were generated by ReGen Theranostics Inc (Rochester, MN).

CSF was obtained from healthy controls or ALS patients seen at the ALS Center at Mayo Clinic Florida, the National Institutes of Health (NIH), the IRCCS Istituto Auxologico Italiano (Milan, Italy), the University of Massachusetts Medical School, and Massachusetts General Hospital (**Table S4**). CSF was collected via standard lumbar puncture, aliquoted and stored at -80°C. ALS patients had El Escorial clinically definite, probable, laboratory supported probable or possible ALS of <5 years' duration. Patients received lumbar puncture generally in the diagnostic early phase of the disease. Patients receiving tracheostomy ventilation or non-invasive mechanical ventilation for >23 h/day were excluded. Also excluded were patients with a history of conditions which could potentially alter the blood–CSF barrier (i.e., spinal surgery). The presence or absence of the *C9ORF72* repeat expansion was determined by repeat-primed polymerase chain reaction (PCR) method as previously described (DeJesus-Hernandez et al., 2011) supported by amplicon-length analysis and, in select cases, by Southern blotting (Akimoto et al., 2014) or using commercial PCR (Athena diagnostics). These studies received Institutional Review Board approval; all subjects provided written informed consent.

Differentiation and treatment of iNeurons for immunohistochemistry and Western blotting. Fibroblasts were maintained in Dulbecco's modified Eagle's medium (Lonza) supplemented with 10% heat-inactivated fetal bovine serum (Sigma-Aldrich), 100 units/ml penicillin, and 100 µg/ml streptomycin (Gibco) at 37°C, in an atmosphere containing 5% CO_2 and 95% air. Lentiviral shRNA against human PTBP1 (shPTB) cloned into pLKO.1 was a kind gift from Dr. Fu (University of California, San Diego). Both shPTB and non-silencing shRNA in the pLKO.1 vector (Sigma-Aldrich) were packaged in HEK293FT cells using Virapower (Invitrogen) packaging mix. Viral particles were collected 48 and 72 h after transfection.

To generate iNeurons, fibroblasts were seeded on a poly-D-lysine-coated surface and were transduced with pLKO.1 coding for shPTB1 or non-silencing control shRNA for 12-18 h in the presence of 5 µg/ml polybrene. Two days post-infection, cells were selected with 1.5 µg/ml puromycin for 48 h. At day 5, 10 ng/ml basic fibroblast growth factor (bFGF, GenScript) was added to the medium for two days. Cells were then maintained in DMEM/F12 medium containing 2% FBS, 25 mg/ml insulin (Sigma-Aldrich), 100 nM putrescine (Sigma-Aldrich), 50 mg/ml transferrin (Sigma-Aldrich), 30 nM sodium selenite (Sigma-Aldrich) and 15 ng/ml bFGF. After six days, the medium was enriched with B27 supplement (Gibco) and a cocktail of neurotrophic factors, including 10 ng/ml each of BDNF, GDNF (R&D Systems), NT3 (Peprotech), and CNTF (Sigma). Immunocytochemical analysis was performed 2-6 days later.

For immunocytochemistry, cells were fixed with 4% paraformaldehyde, permeabilized with 0.5% Triton X-100/PBS and blocked in 5% skim milk/TBS-T. The following antibodies were used in 5% skim milk/TBS-T: mouse anti-MAP2 (Sigma, 1:2,000), mouse anti-Tuj1 (Cell Signaling Technology, 1:2,000), mouse anti-Neurofilament H (Smi-32; Millipore, 1:2,000), rabbit anti-Synapsin 1 (Syn1; Millipore, 1:500), rabbit anti-PSD95 (Abcam, 1:250), rabbit anti-Drebrin (Abcam, 1:500), goat anti-PTB1 (Abcam, 1:200 ICC, 1:1000 WB), rabbit anti-GP (1:1000) and rabbit anti-PR (1:1000). Secondary fluorescent antibodies (Invitrogen) were used at 1:1000 in 5% skim milk/TBS-T. Confocal microscopy was performed using Zeiss LSM 510 microscope.

For Western blot analysis, fibroblasts were transduced with shPTB1 or non-silencing control shRNA. Five days later, cell lysates were prepared and analysed by Western blot using an antibody to PTB1.

For treatment of iNeurons, fibroblasts were converted to iNeurons in 96-well plates and treated with compound **1a** (2 or 4 µM) or DMSO for four days to analyze their effect on the accumulation of poly(GP) or poly(PR) protein inclusions. Serial pictures were generated using the BD Pathway Bioimager. For each condition, the percentage of cells containing poly(GP) or poly(PR) inclusions was calculated from 3-6 wells for each of 3 independent experiments.

RNA Fluorescent in situ hybridization (RNA-FISH) in fibroblasts and iNeurons. RNA FISH of fibroblasts and iNeurons treated with DMSO or compound **1a** (2 μM) for four days was performed as previously described (Lagier-Tourenne et al., 2013) with some modifications. Briefly, plated cells were fixed in 4% PFA/DEPC-PBS, permeabilized with 0.2% Triton X-100/DEPC-PBS, washed twice with DEPC-PBS, dehydrated through 70% and 100% ethanol, and air dried. In some cases, iNeurons were treated with RNase A (2.5 μM; Qiagen) for 15 min at 37°C, or with DNase I (3 U/ml; Invitrogen) for 30 minutes at room temperature, prior to dehydration. Cells were incubated in hybridization buffer (10% dextran sulfate, 50% formamide, 50 mM sodium phosphate buffer (pH 7), 2XSSC) at 66°C for 20-60 min. Prior to use, the locked nucleic acid probe (5′TYE563-CCCCGGCCCCGGCCCC-3', Batch #612968, Exiqon) was denatured at 80°C for 75 s and diluted to 40 nM with hybridization buffer. Cells were hybridized with probe in a sealed, light-protected chamber for 16h-24h at 66°C. The coverslips were subsequently washed with 0.1% Tween-20/2XSSC for 5 min followed by three 10 min stringency washes in 0.1xSSC at 66°C. The cells were stained with Hoechst (Invitrogen), rinsed with DEPC-treated water, dehydrated through 70% and 100% ethanol and air dried. Coverslips were mounted with Prolong Gold antifade reagent (Life Technologies). RNA foci in iNeurons were visualized and quantified using a Zeiss Axiovert Fluorescence Microscope with apotome module. For each of 3 cell lines, 3 fields were randomly selected per condition. For each field, the number of foci-positive nuclei and the total number of nuclei were counted to determine the average percentage of foci-positive cells.

RNA extraction and quantitative PCR (qRT-PCR) of *C9ORF72*. iNeurons (three different cell lines, in triplicate) were harvested in 1ml of Trizol after treatment with DMSO or **1a** (4 μM, 4 d). RNA was extracted using the Direct-Zol RNA kit combined with in-column DNase I digestion, as per the manufacturer's instructions (Zymo Research, Irvine, CA, USA). RNA integrity was obtained using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

cDNA was obtained after reverse transcription polymerase chain reactions (RT-PCR) using approximately 500 ng of RNA with random primers and the High Capacity cDNA Transcription Kit (Applied Biosystems, Foster City, CA, USA) as per the manufacturer's instructions. Following standard protocols, qRT-PCR was conducted in triplicates for all samples using inventoried TaqMan gene expression assays for total *C9ORF72* [transcript variants 1 (NM_145005.5), 2 (NM_018325.3), 3 (NM_001256054.1) (Hs00376619)], the long form of *C9ORF72* [variants 2, 3 (Hs00945132)], and GAPDH (Hs00266705) (Applied Biosystems) on an ABI Prism 7900HT Fast Real-Time PCR System (Applied Biosystems). Relative quantification of *C9ORF72* variants was determined using the $\Delta\Delta$Ct method and normalized to GAPDH.

Supplemental references

Akimoto, C., Volk, A.E., van Blitterswijk, M., Van den Broeck, M., Leblond, C.S., Lumbroso, S., Camu, W., Neitzel, B., Onodera, O., van Rheenen, W.*, et al.* (2014). A blinded international study on the reliability of genetic testing for GGGGCC-repeat expansions in C9orf72 reveals marked differences in results among 14 laboratories. J Med Genet *51*, 419-424.

Almeida, S., Gascon, E., Tran, H., Chou, H.J., Gendron, T.F., Degroot, S., Tapper, A.R., Sellier, C., Charlet-Berguerand, N., Karydas, A.*, et al.* (2013). Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol *126*, 385-399.

Auron, P.E., Weber, L.D., and Rich, A. (1982). Comparison of transfer ribonucleic acid structures using cobra venom and S1 endonucleases. Biochemistry *21*, 4700-4706.

Bellaousov, S., Reuter, J.S., Seetin, M.G., and Mathews, D.H. (2013). RNAstructure: web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res *41*, W471-474.

Bugaut, A., Murat, P., and Balasubramanian, S. (2012). An RNA hairpin to G-quadruplex conformational transition. J Am Chem Soc *134*, 19953-19956.

Deane, F.M., Miller, C.M., Maguire, A.R., and McCarthy, F.O. (2011). Modifications to the Vilsmeier-Haack Formylation of 1,4-Dimethylcarbazole and Its Application to the Synthesis of Ellipticines. J Heterocyclic Chem *48*, 814-823.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J.*, et al.* (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron *72*, 245-256.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.P., and Ehresmann, B. (1987). Probing the structure of RNAs in solution. Nucleic Acids Res *15*, 9109-9128.

Fojtik, P., Kejnovska, I., and Vorlickova, M. (2004). The guanine-rich fragile X chromosome repeats are reluctant to form tetraplexes. Nucleic Acids Res *32*, 298-306.

Gendron, T.F., Bieniek, K.F., Zhang, Y.J., Jansen-West, K., Ash, P.E., Caulfield, T., Daughrity, L., Dunmore, J.H., Castanedes-Casey, M., Chew, J.*, et al.* (2013). Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol *126*, 829-844.

Hardin, C.C., Watson, T., Corregan, M., and Bailey, C. (1992). Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). Biochemistry *31*, 833-841.

Hingorani, D.V., Randtke, E.A., and Pagel, M.D. (2013). A CatalyCEST MRI Contrast Agent That Detects the Enzyme-Catalyzed Creation of a Covalent Bond. Journal of the American Chemical Society *135*, 6396-6398.

Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. Nat Chem Biol *3*, 218-221.

Lagier-Tourenne, C., Baughn, M., Rigo, F., Sun, S., Liu, P., Li, H.R., Jiang, J., Watt, A.T., Chun, S., Katz, M.*, et al.* (2013). Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc Natl Acad Sci U S A *110*, E4530-4539.

Marin, V.L., and Armitage, B.A. (2005). RNA guanine quadruplex invasion by complementary and homologous PNA probes. J Am Chem Soc *127*, 8032-8033.

Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A *101*, 7287-7292.

Mergny, J.L., Phan, A.T., and Lacroix, L. (1998). Following G-quartet formation by UV-spectroscopy. FEBS Lett *435*, 74-78.

Mullen, M.A., Olson, K.J., Dallaire, P., Major, F., Assmann, S.M., and Bevilacqua, P.C. (2010). RNA G-Quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles. Nucleic Acids Res *38*, 8149-8163.

Plug, J.P.M., Koomen, G.J., and Pandit, U.K. (1992). An Expedient Synthesis Of 9-Hydroxyellipticine. Synthesis-Stuttgart, 1221-1222.

Todd, A.K., and Neidle, S. (2011). Mapping the sequences of potential guanine quadruplex motifs. Nucleic Acids Res. *39*, 4917-4927.

Todd, Peter K., Oh, Seok Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, Abigail J., Chen, K.-c., Scaglione, K.M., Basrur, V.*, et al.* (2013). CGG Repeat-Associated Translation Mediates Neurodegeneration in Fragile X Tremor Ataxia Syndrome. Neuron *78*, 440-455.

Ziehler, W.A., and Engelke, D.R. (2001). Probing RNA structure with chemical reagents and enzymes. Curr Protoc Nucleic Acid Chem *Chapter 6*, Unit 6 1.

Zumwalt, M., Ludwig, A., Hagerman, P.J., and Dieckmann, T. (2007). Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. RNA biology *4*, 93-100.